

Project Review – I

Analysis of Phishing attack PCAP file using scapy library in python

Jishnu Saurav Mittapalli, Shaumaya Ojha

School of Computing Science and Engineering, Vellore Institute of Technology,
Chennai Chennai, Tamilnadu, India.

ABSTRACT

In today's world information security has become very important as all sorts of information about everybody in the world is available. Also information has become of great business importance in recent times. So we do not know who want our information and why. So, we have to keep our information very safely. Even the data of large companies is prone to attacks and are in danger of losing their data. There are so many attacks happening in today's world like Man-in-the-middle, phishing attack, Drive-by attack, Password attack, SQL injection attack. In this project we are mostly concentrating on the phishing attack. We are studying and analyzing the PCAP wireshark file of the attack and presenting the visualized results using scapy. Also presenting the different methods to prevent these phishing attacks.

Keywords: Phishing attack, scapy

1. INTRODUCTION

Information security's main aim is protecting information that is sensitive from attacks like phishing attacks and laundering of money. The Phishing attack is the method of manipulating people who do not know much about such attacks. All huge organizations have issues of security that greatly worry their users, contractors and other people who have their important data with the organization[1]. The phishing attack is the network in which the mugger creates the duplicate of a website to make fool of users online and steal their information that is private. Phishing attack is the an attack that is done by combining the techniques of convincing the user along with social engineering methods to fake the user and fool him. This is mostly done by fake messaging and fake emails. Mostly information regarding the social media accounts, financial accounts, password etc are stolen using this attack.

The Phishing attack is a very serious one and has no limit in the internet, it is an attack from which everyone must learn how to protect themselves. There are so many methods in which people easily give up their valuable private information. In this the entire attack starts from the sending of a fake mail or a fake message. The mail or message appears to be from a serious source that can be trusted. Mostly the victim is provoked to enter his details into a fake website that looks just like the original. Sometimes malware also enters into the victims system in this method. Attackers give some exciting gain options mostly financial to the victim to which the victim gets tempted and does as said. On being asked the victim fills in his credit card/debit card/bank account details and other personal data. Also login details and other details are in some cases stolen from this attacks in order to gain something large. With data the attackers get from social media and other sources they choose good victims and create reliable emails like wishing them for their birthday or anniversary etc., so that they can fool them easily. Good phishing mails generally appear to be from good and famous websites and look just like the original. For this purpose logos, and other detailed information about the company is taken by the attacker from the

company's website. There are also URLs created in these emails that immediately download some kind of malware into the user/victim machine putting them into even higher danger and loss. The various steps are shown in Fig-1. Such attacks also take place through injection of malicious files through other methods like USB drives and other input devices making it very difficult to get to know where we got the malware from. So basically in this paper we want to show the results we obtained by analyzing the pcap file generated during a phishing attack and visualize the details to get a better perspective to solve such problems and make better solutions to prevent these kinds of phishing attacks which are really increasing in today's world. This paper in detail describes what are the packet changes happening during the attack and visualizes all the data hence gained [2][3]. Scapy runs natively on Linux, and on most Unixes with libpcap and its python wrappers. The same code base now runs natively on both Python 2 and Python 3. Scapy is a packet manipulation tool for computer networks, originally written in Python by Philippe Biondi. It can forge or decode packets, send them on the wire, capture them, and match requests and replies. It can also handle tasks like scanning, tracerouting, probing, unit tests, attacks, and network discovery.

2. RELATED WORK

method to filter spam mails through their work, they have used complicated methods of word processing and natural processing to find an effective and useful solution after they trained it from the CSDMC2010 SPAM training data set and they have achieved a great accuracy.[4]

Thomas J. and their team have given in detail shown the methods of classification of emails i.e. the bad emails are removed or separated. In addition, estimation and classification. Different selection methods for features are comparatively presented. A few really good experiments have resulted, comparative selection e-mail by body and header has email classification to be effective in [5].

In the study of Panigrahi P. K., the machine learning techniques like Bayes algorithms, tree algorithms, support vector machines, artificial neural network and other few ones that have been held by the UCI machines Learning Storage were compared and they have done good work to compare and examine the different methods and have given their comparative results.[6]

AlRashid H. and his team have worked to find out a solution that helps us to reduce the number of false positives in email classification. They have done a really good job in doing so. On the basis of the problem that spam classification is done based on different reasons in different situations they have built their mode[7].

Leslie F. Sikos, in his paper "Packet analysis for network forensics: A comprehensive survey" has compared the different ways in which this analysis of packets can help us in determining the attacks and give ways to solve or prevent the attacks data breaches, unauthorized website access, malware infection, and intrusion attempts, and to reconstruct image files, documents, email attachments, etc. sent over the network and these are analyzed from the packets[8].

Leendert van Duijn had a goal to detect and prevent these malwares and the traffic due to this and they have helped by showing that analyzing these PCAP files can be of great help to prevent and stop such kind of attacks in his paper "Beacon detection in PCAP files"[9].

Biju Issac and their team in their paper "Analysis of Phishing Attacks and Countermeasures" have given the various measures to be taken to prevent phishing attacks[10].

Muhammet Baykara and their team in their paper “Detection of phishing attack “ have listed out the various ways in which we can detect this phishing attack[11].

JulianJang-Jaccard, in their paper “ A survey of emerging threats in cybersecurity” have suggested the different threats that are there or coming up in this field of cybersecurity[12].

A. Das and their team in their paper "SoK: A Comprehensive Reexamination of Phishing Research From the Security Perspective," in *IEEE Communications Surveys & Tutorials* have given a very different perspective on this phishing attack[13].

A. El Aassal, and their team in their paper "An In-Depth Benchmarking and Evaluation of Phishing Detection Research for Security Needs" have given the various thresholds and have given the level that is required for security needs [14].

C. Phamand team in their paper, "Phishing-Aware: A Neuro-Fuzzy Approach for Anti-Phishing on Fog Networks," have given a new method i.e using fuzzy logic they have devised a method to detect phishing attacks and prevent them[15].

J. Li and S. Wang in their paper "PhishBox: An Approach for Phishing Validation and Detection," have described their project PhishBox, and have helped us and have created an application or tool that detects if there is a phishing attack and validates it[16].

3. PROPOSED ARCHITECTURE

In this project we intend to take the PCAP file captured by wireshark during a phisher attack. i.e, while the phishing attack was happening the PCAP file was captured. PCAP is an application programming interface for capturing network traffic. The . pcap file extension is mainly associated with Wireshark; a program used for analyzing networks. . pcap files are data files created using the program and they contain the packet data of a network. These files are mainly used in analyzing the network characteristics of a certain data. We then analyse it using the PCAP file using Scapy library which is a packet manipulation library in python for computer networks analysis. We use that to identify the packets.

Scapy is a powerful Python-based interactive packet manipulation program and library. It is able to forge or decode packets of a wide number of protocols, send them on the wire, capture them, store or read them using pcap files, match requests and replies, and much more. It is designed to allow fast packet prototyping by using default values that work. It can easily handle most classical tasks like scanning, tracerouting, probing, unit tests, attacks or network discovery (it can replace hping, 85% of nmap, arp spoof, arp-sk, arping, tcpdump, wireshark, p0f, etc.). It also performs very well at a lot of other specific tasks that most other tools can't handle, like sending invalid frames, injecting your own 802.11 frames, combining techniques (VLAN hopping+ARP cache poisoning, VoIP decoding on WEP protected channel), etc. Scapy supports Python 2.7 and Python 3 (3.4 to 3.7). It's intended to be cross platform, and runs on many different platforms (Linux, OSX, *BSD, and Windows).

We try to find out the different IP, TCP and UDP packets that were being transferred during the attack. We then analyse the packets, the different IP addresses and the ports being used for the attack. After analyzing the attack we plan to visualize the data using various other libraries in python like numpy, pandas and matplotlib etc so that we can find better results and find out which ip address is the one that is doing the attack.

The following figure shows the various steps involved in the phishing attack as shown below.

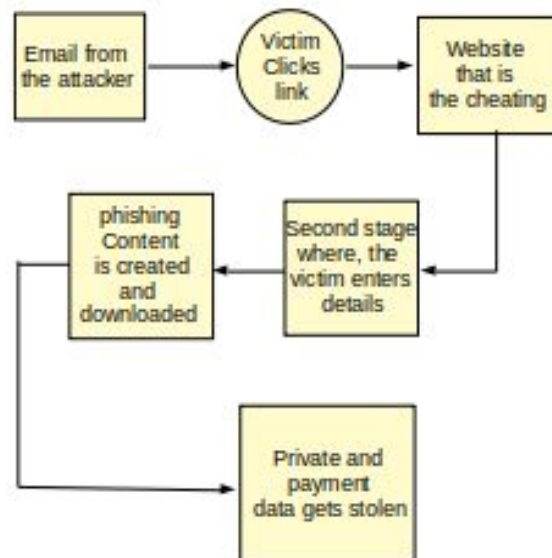


Fig. 1 - Steps involved in Phishing attack

We then intend to find methods to prevent the phishing attacks as the prevention of the phishing attack is our main goal. Using the steps as given in the following diagram.

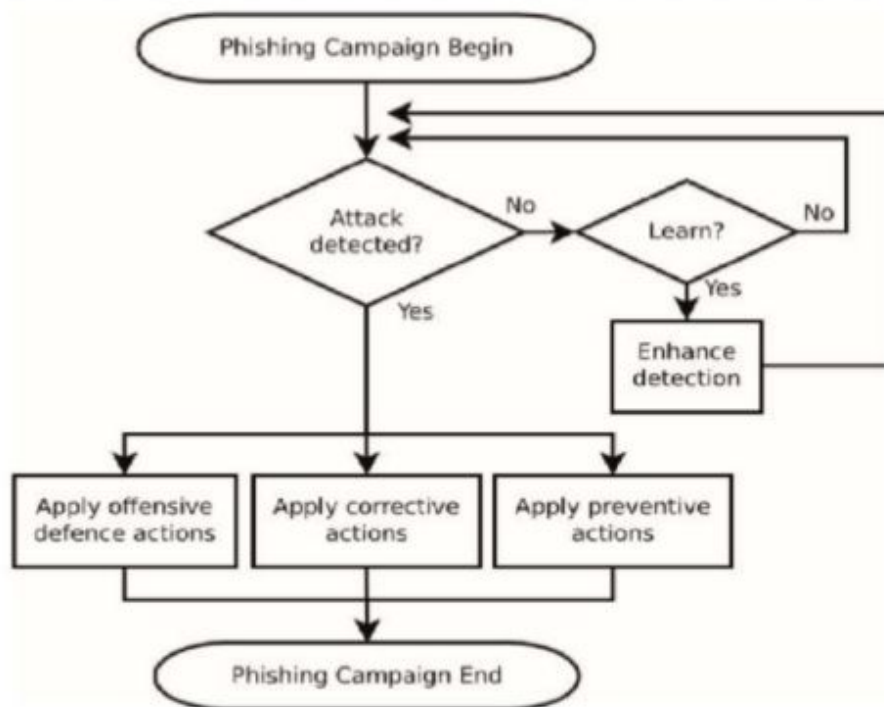


Fig.2 - Steps involved in preventing Phishing attack

4. TESTBED - ENVIRONMENT DETAILS

We are using the virtual machine of Kali linux, on virtualbox to conduct our experiment and to do the analysis. We have downloaded the iso file of kali linux and installed it on a virtual machine on Virtual box. As in the images below:

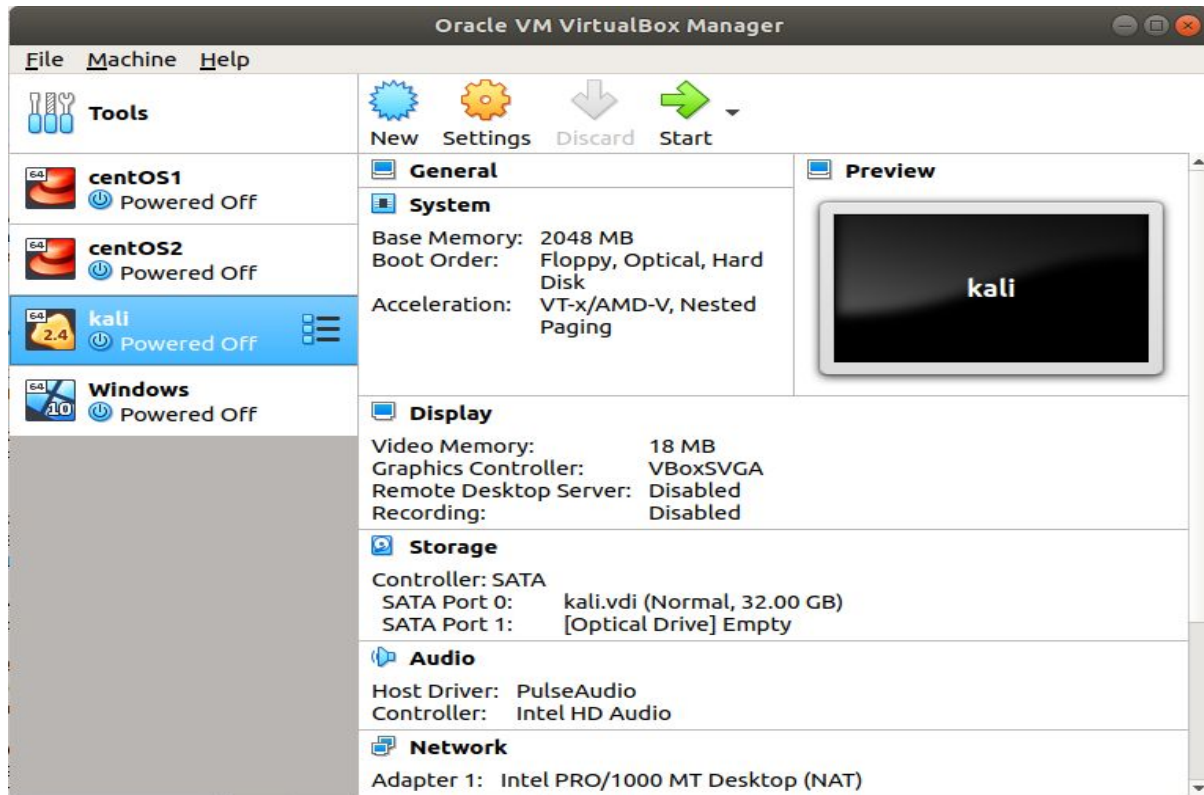


Fig. 3 - Kali linux installation on virtual box

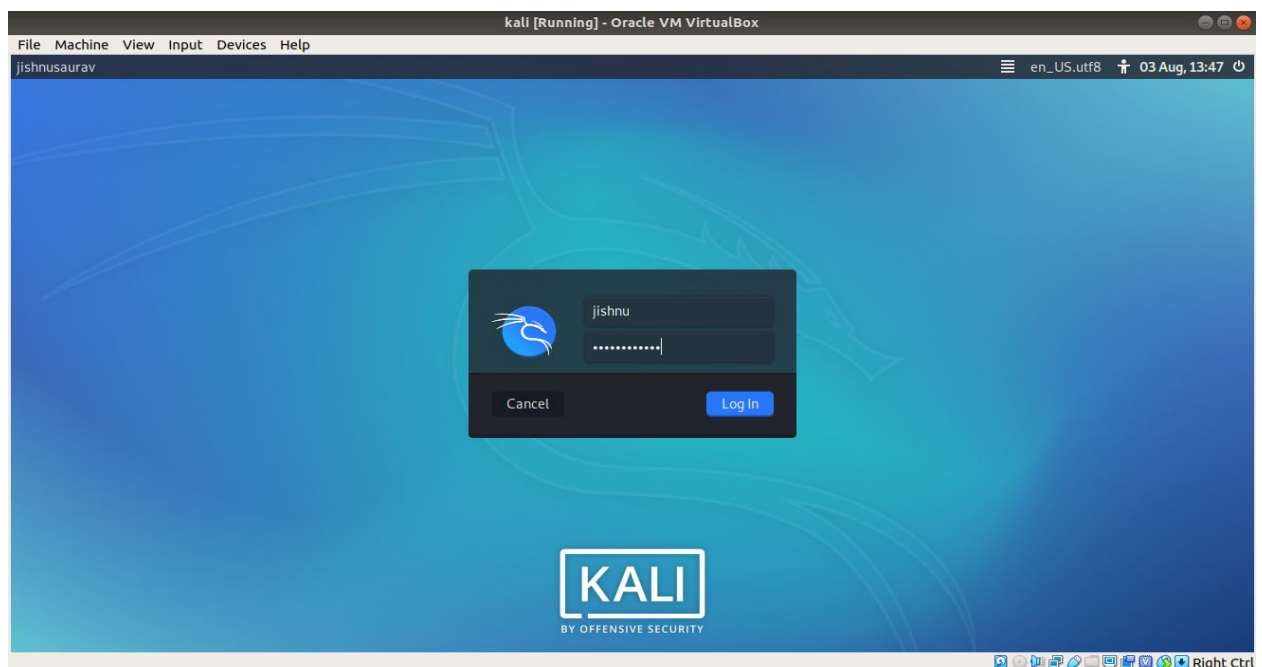


Fig. 4 - Kali linux running on virtual box

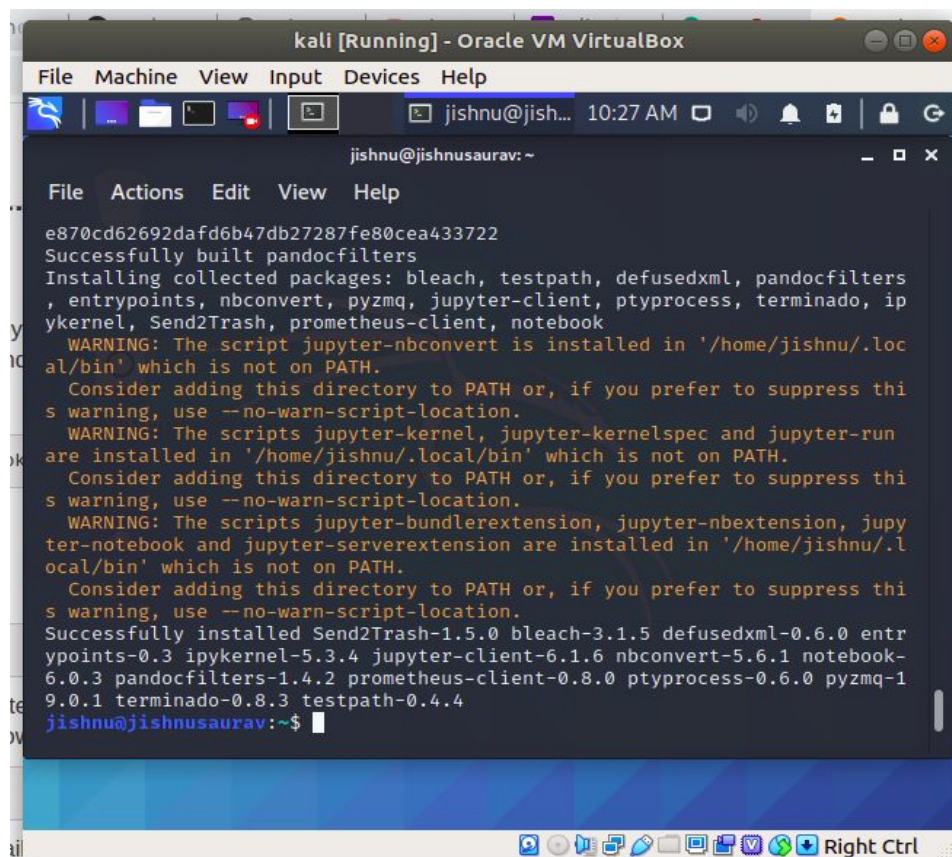
We are using scapy library for the analysis of the pcap file which comes along with python on kali linux.

```
jishnu@jishnusaaurav:~$ pip3 install scapy
Requirement already satisfied: scapy in /usr/lib/python3/dist-packages (2.4.3)
jishnu@jishnusaaurav:~$
jishnu@jishnusaaurav:~$
```

Fig. 5 - Scapy - installation

We are using jupyter notebook for the analysis of the pcap file on python .

The *Jupyter Notebook* is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. We have installed it and it is up and running on the virtual machine.



```
kali [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
jishnu@jish... 10:27 AM
jishnu@jishnusaaurav: ~
File Actions Edit View Help
e870cd62692dafd6b47db27287fe80cea433722
Successfully built pandocfilters
Installing collected packages: bleach, testpath, defusedxml, pandocfilters,
entrypoints, nbconvert, pyzmq, jupyter-client, ptprocess, terminado, ip
ykernel, Send2Trash, prometheus-client, notebook
WARNING: The script jupyter-nbconvert is installed in '/home/jishnu/.loc
al/bin' which is not on PATH.
Consider adding this directory to PATH or, if you prefer to suppress thi
s warning, use --no-warn-script-location.
WARNING: The scripts jupyter-kernel, jupyter-kernelspec and jupyter-run
are installed in '/home/jishnu/.local/bin' which is not on PATH.
Consider adding this directory to PATH or, if you prefer to suppress thi
s warning, use --no-warn-script-location.
WARNING: The scripts jupyter-bundlerextension, jupyter-nbextension, jup
yter-notebook and jupyter-serverextension are installed in '/home/jishnu/.l
ocal/bin' which is not on PATH.
Consider adding this directory to PATH or, if you prefer to suppress thi
s warning, use --no-warn-script-location.
Successfully installed Send2Trash-1.5.0 bleach-3.1.5 defusedxml-0.6.0 entr
ypoints-0.3 ipykernel-5.3.4 jupyter-client-6.1.6 nbconvert-5.6.1 notebook-
6.0.3 pandocfilters-1.4.2 prometheus-client-0.8.0 ptprocess-0.6.0 pyzmq-1
9.0.1 terminado-0.8.3 testpath-0.4.4
jishnu@jishnusaaurav:~$
```

Fig. 6 - Installation of jupyter notebook

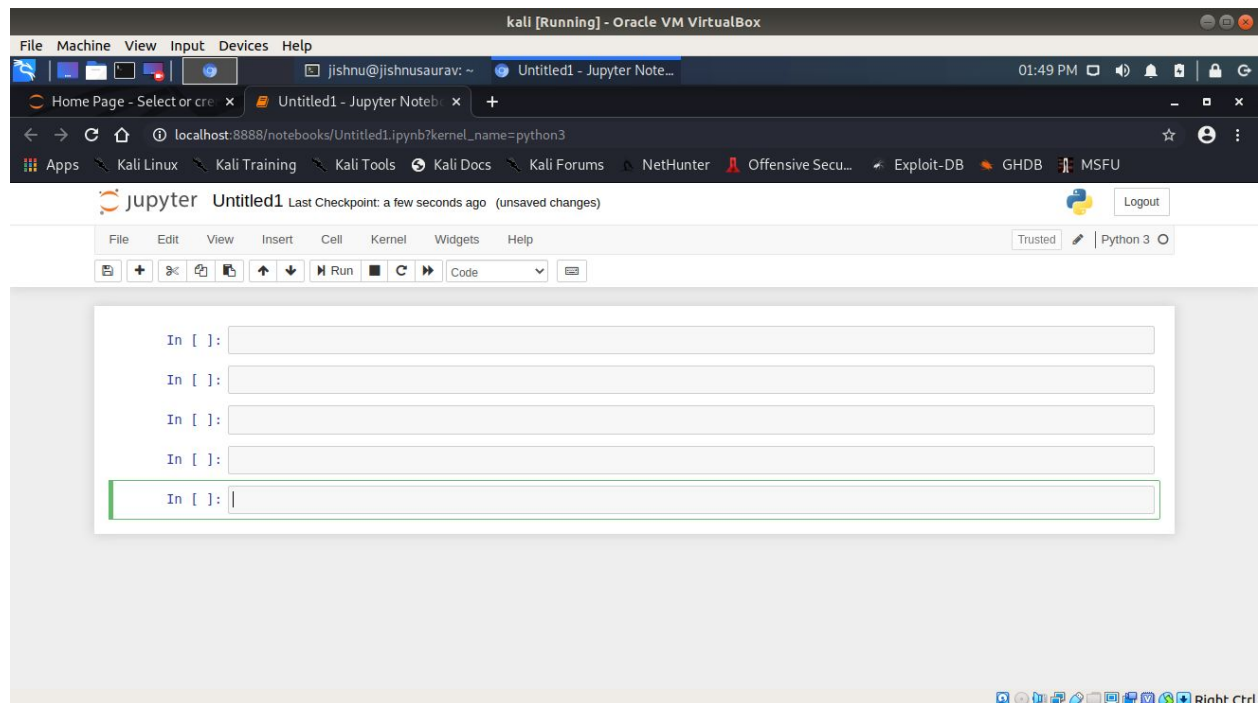


Fig. 7 - Jupyter notebook running

REFERENCES

- [1]. Mouton, M. Malan, L. Leenen and H.S. Venter, "Social Engineer Attack Framework," IEEEConference on Information Security for South Africa , 2014, pp. 1 - 9.
- [2]S. Gupta, A. Singhal and A. Kapoor, "A literature survey on social engineering attacks: Phishing attack," 2016 International Conference on Computing, Communication and Automation (ICCCA), Noida, 2016, pp. 537-540, doi: 10.1109/CCAA.2016.7813778.
- [3]M. Baykara and Z. Z. Gürel, "Detection of phishing attacks," 2018 6th International Symposium on Digital Forensic and Security (ISDFS), Antalya, 2018, pp. 1-5, doi: 10.1109/ISDFS.2018.8355389.
- [4]P. Liu and T. S. Moh, "Content Based Spam E-mail Filtering," 2016 International Conference on Collaboration Technologies and Systems (CTS), Orlando, FL, pp. 218-224, 2016.

[5]J. Thomas, N. S. Raj and P. Vinod, "Towards filtering spam mails using dimensionality reduction methods," 2014 5th International Conference - Confluence The Next Generation Information Technology Summit (Confluence), Noida, pp. 163-168, 2014.

[6]P. K. Panigrahi, "A Comparative Study of Supervised Machine Learning Techniques for Spam E-mail Filtering," 2012 Fourth International Conference on Computational Intelligence and Communication Networks, Mathura, pp. 506-512, 2012.

[7]H. AlRashid, R. AlZahrani and E. ElQawasmeh, "Reverse of e-mail spam filtering algorithms to maintain e-mail deliverability," 2014 Fourth International Conference on Digital Information and Communication Technology and its Applications (DICTAP), Bangkok, pp. 297-300, 2014.

[8]Leslie F.Sikos,"Packet analysis for network forensics: A comprehensive survey"
doi.org/10.1016/j.fsidi.2019.200892

[9]Leendert van Duijn,"Beacon detection in PCAP files, August 2014.

[10]Biju Issac, Raymond Chiong, Seibu Mary Jacob "Analysis of Phishing Attacks and Countermeasures "

[11]M. Baykara and Z. Z. Gürel, "Detection of phishing attacks," 2018 6th International Symposium on Digital Forensic and Security (ISDFS), Antalya, 2018, pp. 1-5, doi: 10.1109/ISDFS.2018.8355389.

[12] JulianJang-Jaccard, Surya nepal "A survey of emerging threats in cybersecurity"

[13]A. Das, S. Baki, A. El Aassal, R. Verma and A. Dunbar, "SoK: A Comprehensive Reexamination of Phishing Research From the Security Perspective," in *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 671-708, Firstquarter 2020, doi: 10.1109/COMST.2019.2957750.

[14]A. El Aassal, S. Baki, A. Das and R. M. Verma, "An In-Depth Benchmarking and Evaluation of Phishing Detection Research for Security Needs," in *IEEE Access*, vol. 8, pp. 22170-22192, 2020, doi: 10.1109/ACCESS.2020.2969780

[15]C. Pham, L. A. T. Nguyen, N. H. Tran, E. Huh and C. S. Hong, "Phishing-Aware: A Neuro-Fuzzy Approach for Anti-Phishing on Fog Networks," in *IEEE Transactions on Network and Service Management*, vol. 15, no. 3, pp. 1076-1089, Sept. 2018, doi: 10.1109/TNSM.2018.2831197.

[16]J. Li and S. Wang, "PhishBox: An Approach for Phishing Validation and Detection," 2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech), Orlando, FL, 2017, pp. 557-564, doi: 10.1109/DASC-PiCom-DataCom-CyberSciTec.2017.101.