Analysis and Implementation of Threat Agents Profiles in Semi-Automated manner for a Network Traffic in Real-Time Information Environment.

Gaurav
Ph.D. Researcher
University of Hertfordshire
g.gaurav@herts.ac.uk

Dr. Stilianos Vidalis
Principal Lecturer
University of Hertfordshire
s.vidalis@herts.ac.uk

Dr. Catherine Menon
Senior Lecturer
University of Hertfordshire
c.menon@herts.ac.uk

Dr. Niharika Anand
Assistant Professor
IIIT, Lucknow, India.
niharika@iiitl.ac.in

Abstract- The threat assessment is the continuous process of monitoring the threats identified in the network of the real-time informational environment of an organization and the business of the companies. The sagacity and security assurance for the system of an organization and business of companies is more seemed to need that information security exercise to be unambiguously and ineffective manner for handling the threat agent's attacks. How this unambiguous and effective manner in the present-day state of information security practice working? Given the prevalence of threats in the modern information environment, it is essential to guarantee the security of national information infrastructure. As the existing models and methodology are not addressing the attributes of threats like motivation, opportunity, and capability (C, M, & O) and the critical threat intelligence (CTI) feed to the threat agents during the penetration process ineffective manner due to which security assurance arises for an organization and the business of companies. we propose a semi-automatic model of information security, which can deal with situational awareness data, strategies prevailing information security activities, and protocols monitoring of specific types of the network next to the real-time information environment. In this paper, we analyze and implement the threat assessment of network traffic in a specific real-time informational environment. To achieve this, we are determining various unique attributes of threat agents from the Packet Capture Application Programming Interface (Pcap files/DataStream) collected from the network between 2012 to 2019. We used hypothetical and real-world examples of a threat agent to evaluate the three different factors of threat agents i.e., Motivation, Opportunity, and Capability (M, O, C). Based on this, we also designed and determine the threat profiles, critical threat intelligence (CTI), and complexity of threat agents that are not covered in the existing threat agent taxonomies models and methodologies.

Keywords – Threat Agents; Motivation; Opportunity; Capability; User Profiling; Implicit Modeling; Real-Time User Monitoring; Complexity Threat Agent; Threat Assessment.

I – Introduction

Identifying the potential cybersecurity threat capability in real-time is a crucial activity, given that useful information about the threat in a network helps cybersecurity practitioners to take suitable action to mitigate the risk in a network [1]. Elaborating all the information about the potential cybersecurity threats of an organization is a typical achieved manually by the existing models and methodology as discussed in section 2. But it can be implemented in an automated manner with help of machine learning tools, techniques, and various real-time models [2]. The behaviors of threat agents are erratic, and the goals of threat agents change with time or the purpose of the task based on the motivation, opportunity, and capability [3][4]. Profiling is a process that generates a profile for the threat agents based on the historical information extracted from the Packet Capture Application Programming Interface

(Pcap) files captured in a network with the help of penetration testing phases. The profile can be populated by having suitable, ample, and precise information about the threat agent like behavior and other useful information such as source IP address, destination IP address, number of open ports, number of packets generated, location of the threat agent, and time spend on the network with minimal user intervention [5]. The minimal user intervention is because the footprints captured by the capturing data tool during threat assessment in the form of Packet Capture Application Programming Interface (Pcap) files cannot be altered by the potential threat agent while traversing the network of an organization. The alteration cannot be done by the threat agent because once they generate the packets in the network then they are unable to erase the footprint of generating the packets because of the accessing property of the network. This research attempts to recognize the aspects and deliver solutions that why do we require to implement profiling of threat agents? Threat profiling is the most important aspect to perform threat assessment for an organization. If we have the threat profile for the historically identified threat agents from the network of an organization then we can use these profiles as references, while executing the threat assessment for the situational awareness data captured from the network can be used effectively and in an optimized manner in order to address the recent threat agent identified from the network.

It has been accepted that continuous threat assessments do mitigate the risks [6]. In the modern socially driven, virtual computing era, threat assessments are hindered by a lack of resources, complexity, and size of data [7]**.** Information Environments are large heterogeneous infrastructures, hosting a large amount of data, collected from different types of sensors and platforms [8]. To cope with a large amount of data, decision aid tools should provide their understanding of situation awareness and threat assessments. University Computer Emergency Response Team (CMU-CERT) groups determined that there are three key groups of threat agents i.e., the technology of organization sabotage, compromising with intellectual property, and data stream fraud [9]. As the number of growing cases highlighted by internet media in recent years revealed that both business organizations and government organizations suffered a similar experience, whereas the priority information has been filtrated by internal users of the organization and shared with the threat agents [10]. The threat agents require serious attention from both users and organizations.

Referencing to covid-19 Nowadays, the organization and business mostly sharing their file and documents with each other with the help of the internet to run their business. It is now common practice for users of the organization to have admittance to large repository documents which are electronically warehoused on distributed file servers. Many organizations offer company laptops and desktops to users for working while using e-mail to organize and scheduling/rescheduling meetings. Amenities such as video conferencing are repeatedly used for holding meetings throughout the world, and users of an organization are continuously connected to the internet. The electronic nature of the files and records of an organization on the internet makes it easier for the threat agents to attack it. While on the advantage side for threat assessment practitioners of an organization can easily capture the activity logs of the internal threat agent while analyzing their captured packets [11]. However, practically analyzing such activity logs is infeasible due to the high volume of activities performed by the user every day.

In this research, we present an efficient model for threat detection and analysis based on the conception of anomaly detection. Given a large variety of the data stream in the form of Packet Capture Application Programming Interface (PCAP) files (between 2012-2019), the model implements the threat agent profiles from the Packet Capture Application

Programming Interface (PCAP) files and determine the cyber threat intelligence (CTI) based on the evaluation of motivation, opportunity, and capability of threats. With the help of these profiles, comparisons can be populated that how the current observations fluctuated from the previous observation. To assess the performance of the tactic, we extracted the useful information from the Packet Capture Application Programming Interface (PCAP) files in a semi-automated manner, and output generated in the form of an excel sheet which consists of various attributes of threat agents identified in the next to the real-world information environment. It was found that the system executed expressively sound for detecting the attacks, and the visualization of reports enabled us to identify which attributes help to determine M, O, C factors for the threat agents. This paper however illustrates all the threats identified in a network captured during the penetration testing against the ESXi server of the University of Hertfordshire. The rest of this paper is as follows. Section II discusses the related work. Section III labels the necessities of analysis and implementation of the system. Section IV presents the proposed system, describing in detail how to evaluate motivation, capability, and opportunity of threat agents. Section V presents the process of effective experimentation of the system, and Section VI concludes this paper.

II – Related Work.

The field of threat agents profiling and analysis of cyber threat intelligence (CTI) has freshly received ample attention in the literature. Researchers have proposed a multiplicity of different models and methodology that are designed to detect or prevent the occurrence of attacks (e.g., [12] and [13]). Likewise, in Vidalis et al. [8] “Assessing Cyber-Threats in the Information Environment”, the author briefly addresses the TAME (Threat Assessments Model For EPS) methodology for threat assessments in real-time informational environments and provides a high-level overview of its phases and process while performing threat assessments. Here, they compare the TAME (Threat Assessments Model For EPS) methodology with other existing methodology because of the number of parameters as sting, effectiveness, and understanding of information security from the threat. TAME (Threat Assessments Model For EPS) was the upgrading version of METEORE 2000 for the micropayment system (MPS). In the initial phases, the author analyses the number of methodologies like Alberts 1999, 2001, Baker 1998, Bayne 2002, Blyth 2003, Dimitrakos 2001, Forte 2000, Hancock 1998, Jones 2002, Nichols 2001 etc. and they found that all are working on waterfall model principle, but such approach is not suitable for the Micro Payment System (MPS). So, they developed a new methodology i.e., TAME (Threat Assessments Model For EPS) which has ability to resolve the issues related to Micro Payment System (MPS). TAME (Threat Assessments Model For EPS) is working simultaneously in four phases named as: -

- Scope of Assessments.
- Threat Agent and Vulnerability Analysis.
- Scenario Construction and System Modelling.
- Evaluation.

According to these phases, TAME (Threat Assessments Model For EPS) determined that how much security is required for a particular organizational system. As all four phases are working simultaneously and one input from a phase becomes the output of another phase and in the same manner vice-versa of inputs and outputs are generating from the TAME (Threat Assessments Model For EPS), it depends on the requirements of threat assessments. The author concludes the TAME (Threat Assessments Model For EPS) by using the assessor as an asset for better understanding and analyzing the systems of an organization.

Morakis et. al. [14] Measuring Vulnerabilities and Their Exploitation Cycle tools used COPS, NESSUS, SYSTEM SCANNER, RETINA, NET RECON, WHISKER, and CYBER COB. In this author's address, a problem faced by a large amount of data in the informational environment is cyber-attacks and the author proposed a vulnerability tree analysis to address such problems faced by several organizations for a long time. They believe that constructing knowledge information concerning a specific domain in an object-oriented hierarchy tree and construct a formal model to analyze them concerning possible scenarios of attacks faced by the computer systems. The main purpose of this is to provide a depth classification of vulnerabilities, as to why such attacks happened on a particular data/asset, analysis of footprints, and scenario of threat agents to exploit vulnerabilities. The main aim of such vulnerability tree analysis is to identify the attacks in early stages and address them or handled them before they do some severe damage to the real-world informational systems. Here, the author illustrates the various tools which are capable to analyse the vulnerability of complex organizational environments, such tools are: - COPS, NESSUS, SYSTEM SCANNER, RETINA, NET RECON, WHISKER, and CYBER COB, etc. But they are not adequate or sufficient in today's modern electronic era of cyber-crime. Because they are not able to address the hazards like- fault-tree analysis, checklists, event-tree analysis, and cause-consequences analysis, etc. To cope with such hazards author combines these tools of vulnerabilities tree analysis with object-oriented trees (OO) and adequately addresses such hazards concerning Boolean Mathematics.

Gerald L et. al. [15] Threat Agents: what Infosec officers need to know? These authors briefly explain about threat agents that how they can do unauthorized access to the computer systems of real-world informational environments and from where they got the motivation, capability, and opportunity to perform such damage in the networks systems. Here, they also illustrate the threat agents and their attributes, function, and impact on a network of informational systems. The author also analyses the digital attacks occur in 2002 in several various countries. They identify that the threat agents of real-world informational environments consist of: -

- Threat agent catalog.
- Historical data.
- Technical report enterprises.
- Reports of business environments.
- Reports of physical environments.
- Recent knowledge/information.
- Recent knowledge of stakeholders.
- Recent knowledge of the staff.
- List of stakeholders.

The authors evaluate the capabilities, motivation, opportunities, and impact with the help of 3-dimension matrix mathematics. They evaluate each factor with the help of metrics and ESA (Empowered Small Agents) threat agents. They identify that because of threat agents in 2002 European union worldwide economic damage is $35millions. So, as the cost of damage is quite more, the system security officer needs to require all knowledge and information about the threat agents or risk management. So that they can secure the system from damage done by cyber-attacks in informational environments.

Adetorera Sogbesan et. al. [16] Collusion Threat Profile Analysis tool used CERT (Computer Emergency Response Team)/USSS (the United States Secret Service) they developed a model to identify the MERIT (Management & Education of Risk of Insider Threat) based on the study of insider threat concerning institute of CERT/USSS. This MERIT provides the facility to mitigate the insider threat of an organization and the key finding is to do the case study of individual threat agents i.e., collision threat. MERIT model the case studies on the insider threat for an organization and based on that threat assessments has been conducting for determining the impact of threat on the business. They also show some figures for losses based on studies done by USSS/ CERT. They just categories the insider attack based on the Ex-employee or the financial gain of any important position hold by an employee in an organization. Based on the study of the number of organizations, last year, 69% of companies measured stated events of data theft (which were not external attacks) these threats were originated from inside the organization. While a massive 91% of companies testified not having operative detection systems for recognizing an insider threat. MERIT model has a limitation/shortcoming to analyze compressive pattern analysis based on motivation factors and behavioral characteristics. The motivation factor of collusion attack not able to address by the MERIT model. They are not able to explain the capability of an insider threat.

These related works draw a strong observation, that access to a real-world data stream is enormously challenging, and thus, researchers synthesize data into several groups based on the threat agents identified in a network. The existing model and methodology doing threat assessment manually due to which their complexity is high. In this research, we predominantly want to epitomize the volume and variety of data that would be analyzed in a modern real-world information environment and display how this could be pooled to form an overall threat assessment for each PCAP file. We also want to exhibit a wide range of threat scenarios as epitomized by our data collected from a real-world in a specific environment and show how our profiling and CTI system of threat agents would be capable of detecting the different attacks based on the patterns identified.

III - Analysis, and Implementation of the System.

The work described in this research has been carried out as part of a wider interdisciplinary project that includes computer security researchers, and cyberpsychology experts. As the research question for the “Near Real-Time Semi-Automated Threat Assessment for Information Environment” is CTI (Cyber Threat Intelligence) data-driven threat agent profiling can be used for determining the motivation, opportunity, and capabilities attributes of threat agent under the context of a continuous threat assessment [17]. The threat remains to be of budding apprehension to governments and businesses organization, it becomes an acute necessity for practical tools to help mitigate the threat that is posed. The modern risk assessment methods or models recognize that there is a need to perform several threat assessments in order to identify/analyze various threats in the modern information environment. As if we do iteratively threat assessment for the network then-new type of threat agents identified in situational awareness data will be addressed easily with help of profiling which the practitioners preparing every time while performing the threat assessments. Security concern, the continuous threat assessments may help in generating the paradox of warning to the cyber operations performed in the information environment. This paper identifies the research gap in the semi-automated information environments, which consists of large heterogeneous infrastructures, hosting a large amount of data, collected from different types of platforms [18]. The different types of platforms mean, the type of environment, and the conditions used by the threat agent to attack the particular network. To

analyze or identify the solution for such an issue of a large amount of data, decision aid tools should provide their understanding toward situation awareness and critical intelligence feeds of the threats in real-time information environments.
In the modern knowledge-based socially driven, virtually computing era, threat assessments are hindered by lack of resources, complexity, and size of data. Information Environments are large heterogeneous infrastructures, hosting a large amount of data, collected from different types of platforms with the help of a number of tools. In this project, the state of art on threat assessment models and methodologies will be considered, while procedural and technology issues will be resolved by applying cyber analytics principles [19].
The purpose of the research paper is to introduce a novel approach that will enable us to take advantage of the vast amount of data collected by the large number of platforms designed in order to identify suspicious traffic, malicious intention, and network attacks in an automated manner.

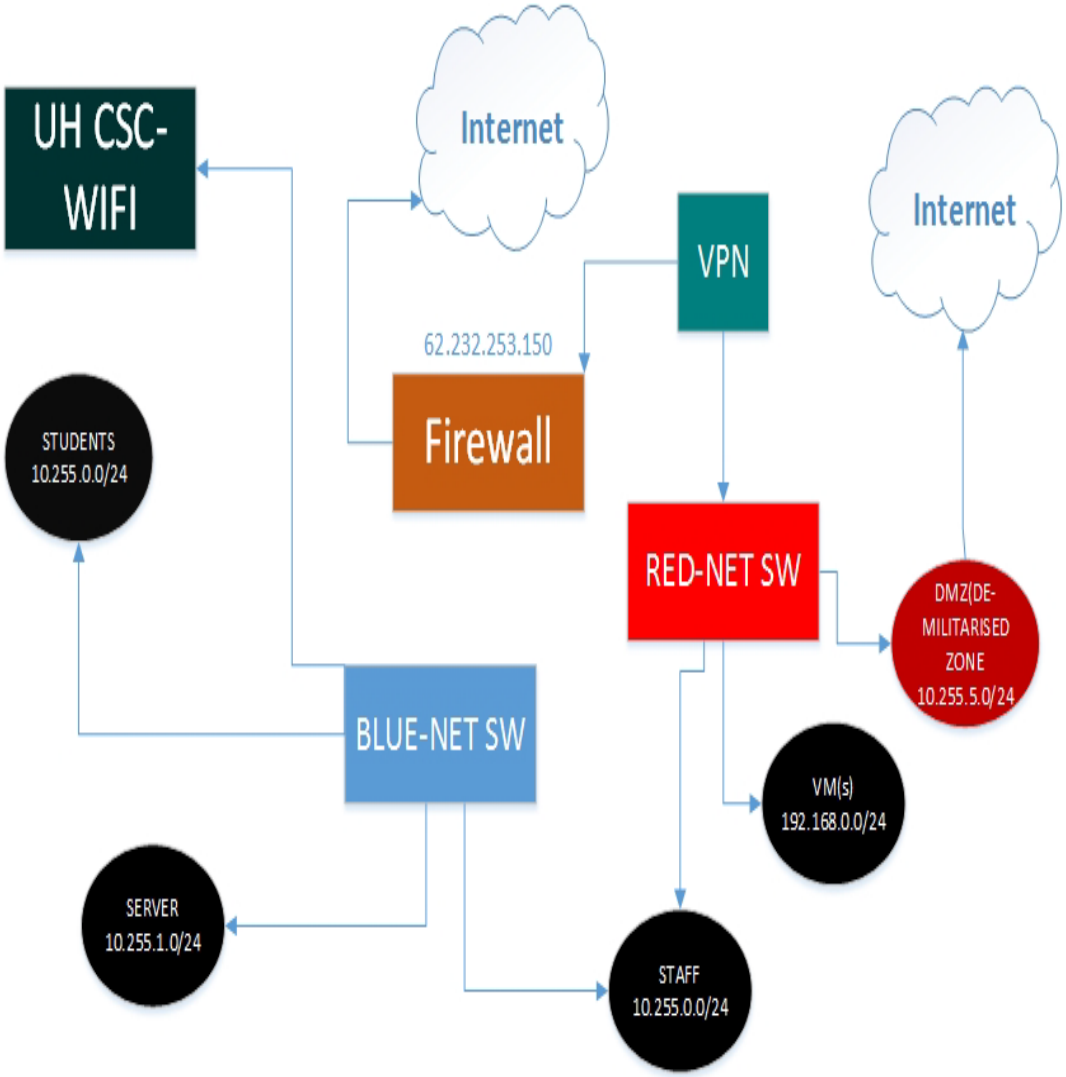

Figure 1 Penetrating Testing Setup at Cybersecurity Laboratory.

1- Experimental Environment of the System.

This is the environment through which we execute the penetration testing against the specific condition of the platform. The number of VPNs used to connect with the REDNET network and also connect through the firewall to save the data from unauthorized access. Further, REDNET connects to DMZ (Demilitarized zone), the number of VM's, and public IP of staff

through which activities can be controlled. BLUENET connects to the user's VM's IPs, ESXi server, UH CSC WIFI (University of Hertfordshire Wi-Fi), and public IP of staff. In this environment, the DataStream or PCAP files were collected from the server with the help of the Wireshark tool [20]. As there are other tools also available like SolarWinds Deep Packet Inspection and Analysis, Paessler Packet Capture, ManageEngine NetFlow Analyzer, Omnipeek Network Protocol Analyzer, TCPdump, and WinDum, etc. As compared to other tools Wireshark is more efficient to extract useful information from Pcap files and also provides the advantage to save the information in CSV format. Figure 1 basic data flow diagram shows that the source of the attack IP address and the destination of the attack IP address through which penetration going on in the network. The role of DMZ is to stop the hacker at the threshold point that after that not allowed anyone to do access [21]. The BLUENET refers to the internal security team that defends against real-world attackers and Red Teams and REDNET are internal/external entities dedicated to testing the effectiveness of a security program by emulating the tools and techniques of likely attackers in the most realistic way possible.

2- The Architecture of System.

The above architecture shows that the ESXi server consists of RED, BLUE, and BLACK NET HP-DL380 ESXi VM WARE CD, DNS, DHCP. Which is further connected to Blue ESXi security zone and DMZ (Demilitarised security zone) and Black ESXi connected to 27x juniper srx240 and srx340 firewalls via 27x lab system multiple images of the environment and dedicated interface in red, blue, and black networks. In this server, all the data and information of university Hertfordshire is available as well as a dedicated environment available for the attackers installed on VM's. DMZ's role is to stop the hacker at the threshold point so that further damage can be controlled by the attacker groups.

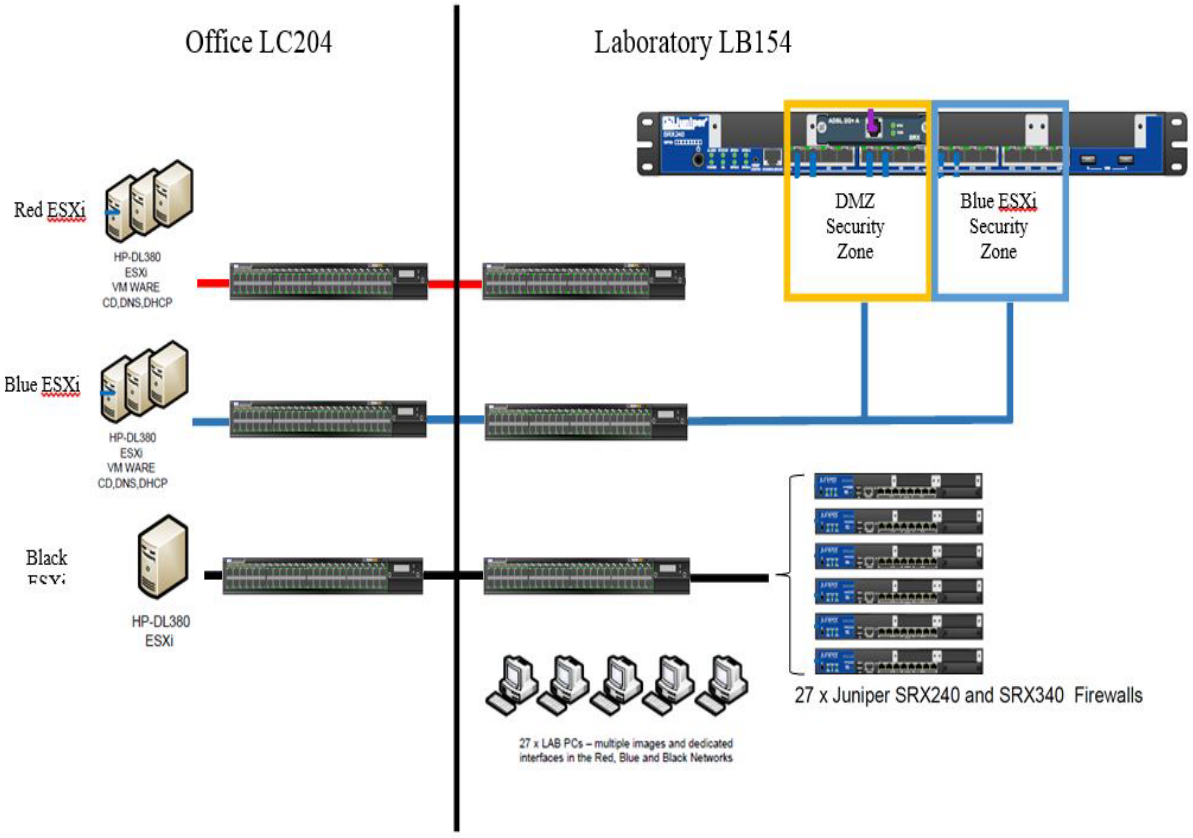

Figure 2 Architecture of System

The main purpose of the Figure 2 architecture is to understand how these attacker groups are generating the traffic in the network, increase a delay time to upload the web page, and extract useful information from the server such as user credentials, webpages, and accessing the files from the databases.

IV - Evaluation of Motivation, Capability, and Opportunity.

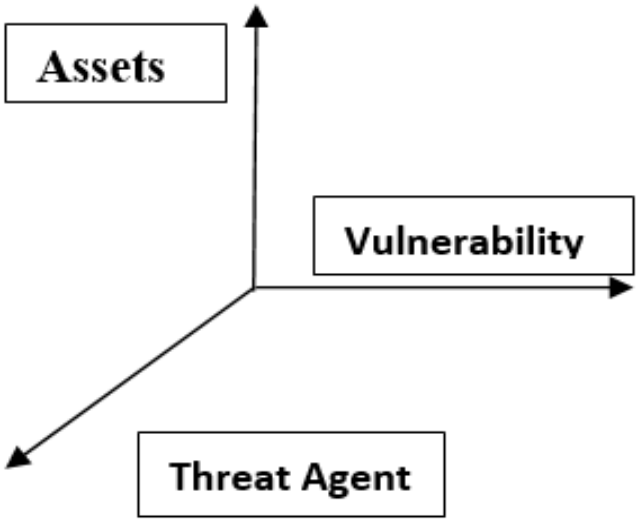

Figure 3 3-D Representation of Threat Assessment

The threat assessment of a model is a continuous process for the DataStream/ Pcap files collected from the network in an information environment. While evaluating the impact of threat agent's groups on the organization or the business, determining the value of assets, vulnerability identification, and threat agent's footprint attributes play the main role for calculation [22]. In Figure 3 the representation of main attributes in a 3-dimensional matrix has been shown. Which needs to be addressed by the model while performing threat assessments of the real-time network.

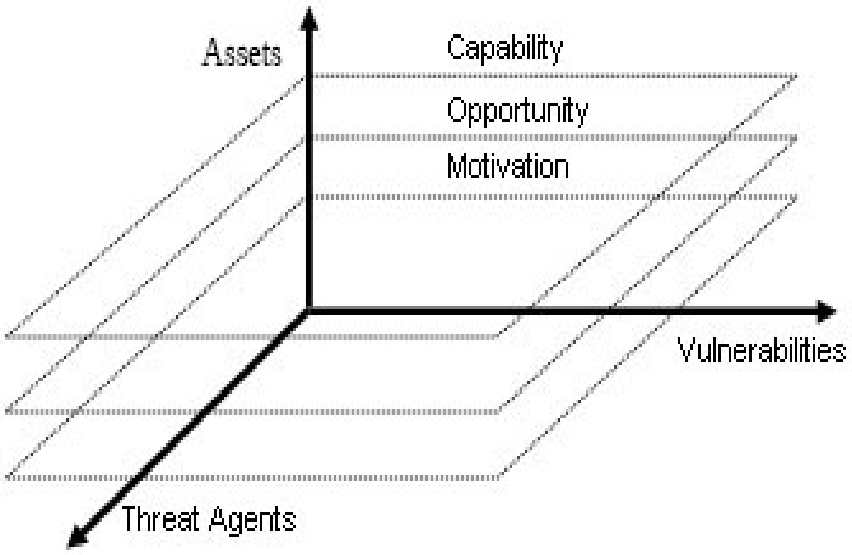

Figure 4 3-Dimensional Matrix

"A threat assessment is a statement of threats that are related to vulnerabilities of company assets and threat agents, and also a statement of believed capabilities that those threat agents

possess". The function threat can be calculated with the help of the threat agent's motivation, capability, opportunity, and the impact of the successful attacks on an organization of the nation.

$$Threat = "Function\ (Motivation, Capability, Opportunity, and\ Impact)" \quad (1)$$

1. Motivation

The evaluation of motivation for threats is the difficult part and could be determined with the help of, analysis of hacktivism branded attacks by groups of assessment models and the vulnerability of the network in next to real-time semi-automated information environments. Motivations of attackers are constantly changing, as it could be noticed by growing the rate of hacktivism attacks by different groups of peoples. It can also see in the differences in unique motivations based on each group or the organization or vertical market, some common motivations include [23]:

- Profit (direct or indirect)
- Hacktivism
- Direct grudge
- Fun / Reputation
- Further access to partner/connected systems.

$$F\ (X)\ = f\ \left(Cap,\ Opp,\ Mto,\ V\left(VIA\right)\right)\ Y\ + f\ \left(Vulnerability\right) Asset\ +\ Impact + T \quad (2)$$

Where Cap stands for capabilities, Opp is an opportunity of the threat agent, Mto is motivation, V (VIA) stands for the value of intangible assets and Y is for threat assessments, and T stands for time complexity.

2. Capability

The capability of threats could be determined with the analysis of risk assessment models and the vulnerability of the network in a next to real-time semi-automated information environment [15].

$$Risk = (Threat) + (Vulnerability) + (Consequences) \quad (3)$$

$$Threat = Intent\ \times Capability \quad (4)$$

Further investigation is achieved with help of several kali Linux tools such as NESSUS, SAINTS, WHISKER, and SARA, etc. The initial phase of the Automatic version of the Threat assessment model is to collect the data from the server. Which has been achieved by the administration of the server between 2012 to 2019. This data mainly consists of PCAP files, which are going to be extracted in a semi-automatic manner with help of a machine learning python tool library available on Tensorflow. The information extracted from these PCAP files having some unique attributes such as: - Time (in min), Highest Protocol, TCP protocol, Source IP Address, Destination IP Address, Source port, Destination port, Total Packet Length, City, Region, Country, Latitude, Longitude, and Internet Service Provider. While executing the extraction process of the large number of PCAP files collected from the server will be converted into a large number of excel sheets based on the unique attributes. These excel sheets consist of all the useful information available about the threat in the Pcap

files such as time spent on the network, location of their IPs, and environment used by them while penetrating the server, etc.

A large amount of information about the threats can be profiled based on their activities performed on the network or specific environment or protocol used to achieve their goal/task. Now, we use all this information to extract all critical threat intelligence (CTI) from these groups of threat can be used to determine the capability, opportunity, and motivation of the threats. This CTI can also be used to identify the new threat in-network and extracted all information by taking previously identified CTI as a reference.
As in Figure 4 motivation of these threat agent groups is going to be calculated based on the environment used by them, or the extraction of type of data executing during the process, factors are responsible for digging into the server like financial gain, breaching the security and socially responsible.

3. Opportunity

Similarly, opportunity can be calculated by checking which ports are open, protocols have open access, and what other factors help a hacker to do the unauthorized access to the server. All this information will be led to the evaluation of the opportunity of the threat agent groups. The same way capability of a threat agent is going to calculated when we identified all information about the threat agents that what type of environment they are using, which protocol they are targeting, how much time they spend on the network, and what type of knowledge they have about the penetrating the network. With the help of all these attributes, we can determine the capability of the threat agent's groups.

V – Results.

1. State-of-the-Art Algorithms

Many different models are used to perform threat assessment for a network in an informational environment on specialized datasets, where some of the datasets are discussed in Sect. IV. Here, we illustrate all the threats identified in a network captured during the penetration testing against the ESXi server of the University of Hertfordshire. To provide an overview of the current state-of-the-art ML approaches used to perform the threat assessment, we group all the identified threats from a network based on their profile maintenance concerning the Python program run against the DataStream/Pcap files captured during the experiment. Similarly, the critical threat intelligence (CTI) [24] feed is identified from this group of threat agents based on their footprints extracted during the analysis phase of the experiment. This overview is further divided into two main categories i.e., Traditional extraction of information from the Pcap files and machine learning techniques applied on the information extracted from the Pcap files to generating the footprints used by the threat agents during traversing in a network of the server.

The first python program provides the accuracy and the unique attributes of the threat agents for precision, false-positive rate (FPR), Anomaly detection rate (ADR), and Fault-measure as originally reported [25]. Secondly, we calculated the performance of the threat agent followed by our proposed 3- dimensional metrics i.e., motivation, opportunity, and capability. Figure 5 shows that the input is a large number of heterogeneous Pcap files used, which have been captured during the experiment.

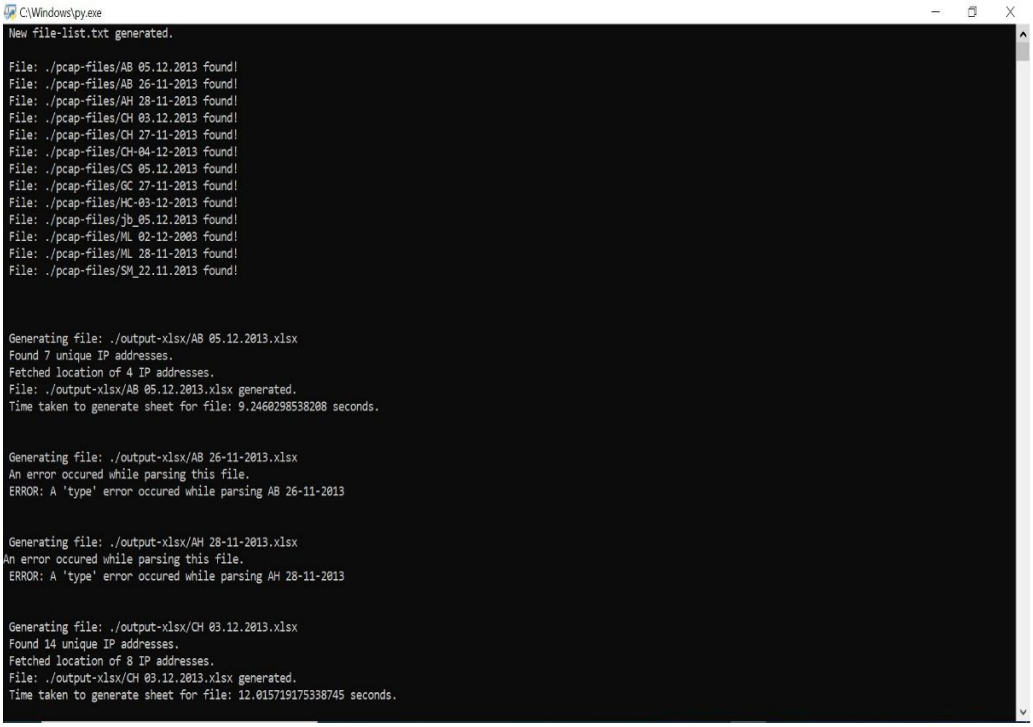

Figure 5. Proposed workflow for raw Pcap file traffic-based feature extraction and experimental results for Unique IP addresses with Time complexity.

The potential output generated with analysis of Pcap files is the unique number of Excel sheets which consist of information about the threat agents such as Time (in min), Highest Protocol, TCP protocol, Source IP Address, Destination IP Address, Source port, Destination port, Total Packet Length, City, Region, Country, Latitude, Longitude, and Internet Service Provider. The specific attributes for each experiment run against the Pcap files can be retrieved from: - https://github.com/Gauravsbin/Excell-sheets-of-pcap-files. Furthermore, With the help of these unique attributes, we can determine the capability and opportunity of the threat agents [26]. Based on the footprints followed by the threat agents during the analysis we can determine the motivation factor for attackers.

Figure 6. Workflow for raw Pcap file and experimental results for Unique IP addresses with Time complexity.

The Pcap files which have been captured during the experiment with the help of the Wireshark tool, few of the files have been corrupted as well while testing with the python program list of crashed files generated during the experiment is also shown in Figure 6**.** The further confirmation about these files we checked it manually as well with help of Wireshark and other analysis tools for PCAP files, we found the same result that no information can be extracted from it. There may be some capture issue or might be the connection lost at the hacker end during the establishment of the network. The time complexity can also be evaluated with help of the addition of all time taken by each Pcap file to generate the unique IPs with attributes of information about it. This is the unique feature of this model as compared to the existing model and methodologies. This could be happened because of using the semi-automatic approaches for threat assessment of networks next to the real-time informational environment.

2. Workflow and Comparative Experiments

Now, in the previous section, the output is generated in form of excel sheets with the unique attribute of threat agents in a semi-automatic manner. So, to determine the motivation, opportunity, and capability of threat agent groups we applied machine learning techniques on the output of the previous phase in such a manner to provide a semi-automatic feature to the model [27].

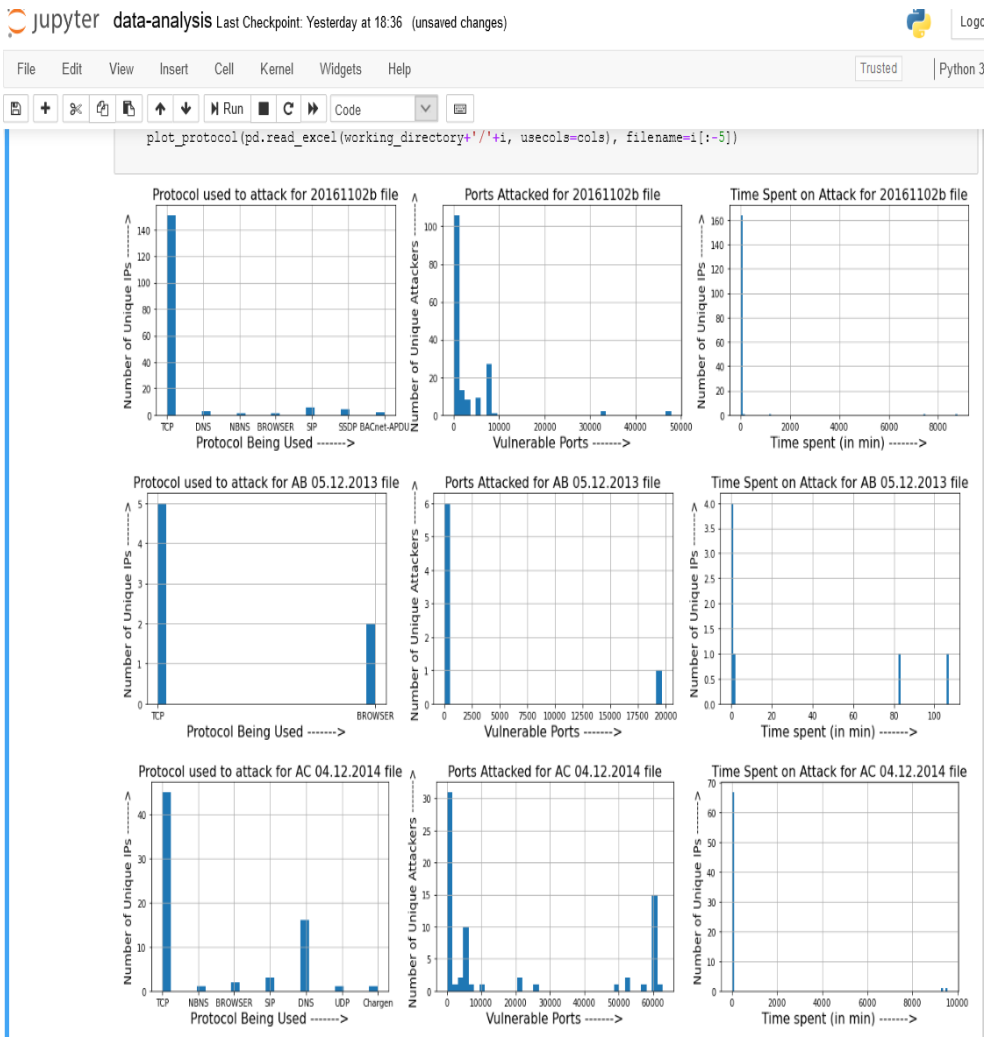

Figure 7 Experimental Results for Each Pcap file, Feature Extraction ML Strategy, and Network.

This novel approach helps us to optimize the complexity of the threat assessment of a network. This paper also shows the process of using machine learning, libraries of python on Tensorflow and deep learning techniques to identify the unique tuples of DataStream/Pcap files will be examined. This approach mainly depends on the chronological order of packets in Pcap files.

Here, we first make groups of all the unique IPs extracted from raw Pcap files captured from the network with help of Wireshark. The grouping of all unique IPs based on their attributes and characteristic features identified during the analysis and implementation of DataStream. Similarly, the potential output generated in the previous phase will be used as potential input for the second phase of analysis and implementation. Such a process is known as the profiling of threat agents. As in the previous phase, we generated the Excel sheet for each captured Pcap file consist of useful information like ports open, on which layer they are operating, time spends on the network, and location of the threat agent, etc.

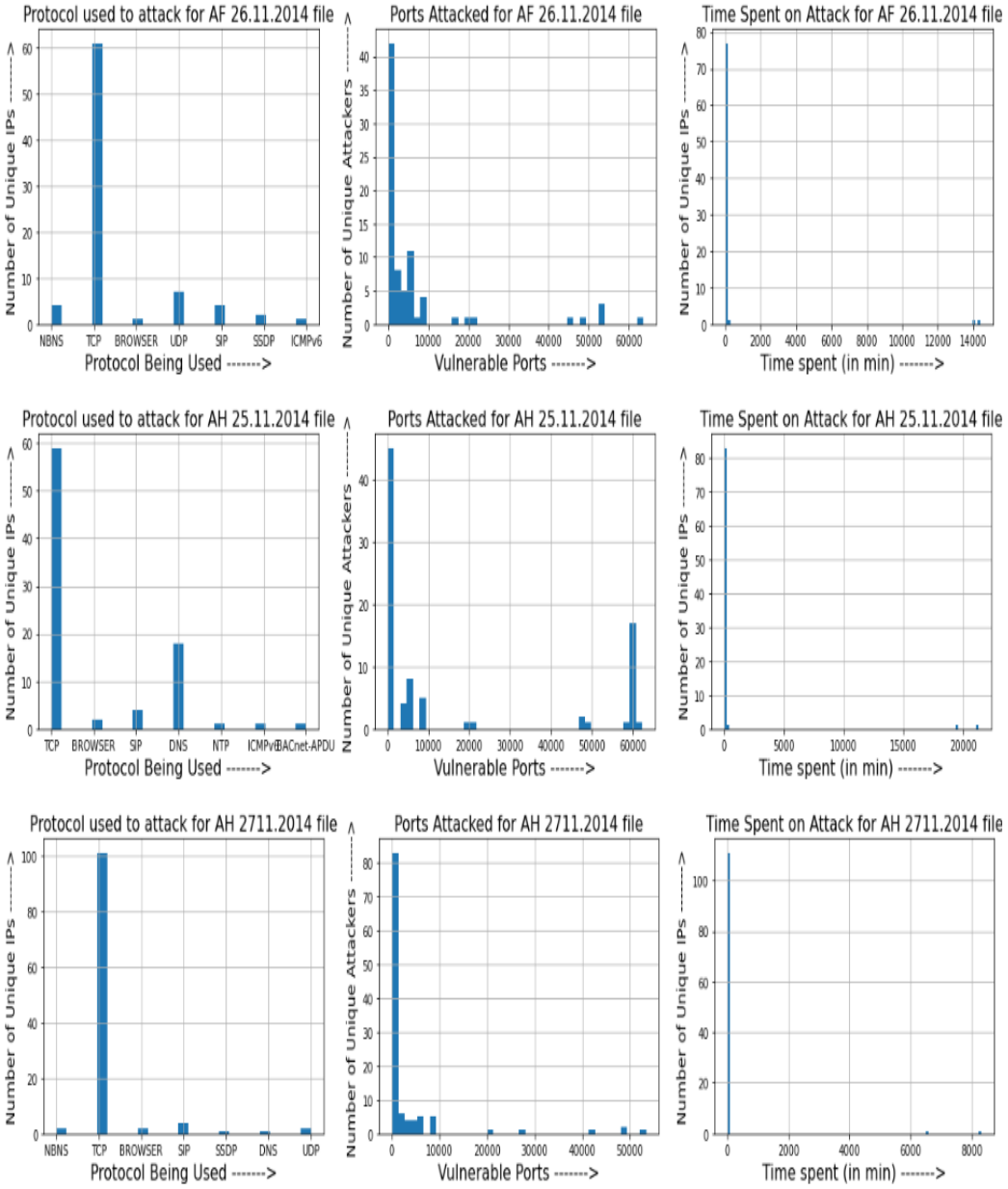

Figure 8 Histogram for each input based on Protocol, Ports, and Time.

Based on this analysis, now we make one more IPYNB file means Interactive Python Notebook which is now known as Jupyter notebook. “Jupyter is a free, open-source, interactive web tool known as a computational notebook, which researchers can use to combine software code, computational output, explanatory text, and multimedia resources in a single document. A Jupyter Notebook document is a JSON document, following a versioned schema, containing an ordered list of input/output cells which can contain code, text (using Markdown), mathematics, plots, and rich media, usually ending with the IPYNB extension[28][29][30]”. This file consists of an algorithm that is performing data clustering of Unique IPs found in the excel sheet of the previous phase. The data clusters of Ips formed based on the number of IPs facing a particular type of attack.

This particular type of attack is determined based on the number of factors identified during the analysis. The IPYNB file collecting all the unique IPs as input and extracting the information like on which layer, they are operating, what type of ports and protocols were compromised when they are attacking the source IPs of end-users, and what information they extracted from the particular environment of the VM’s, etc. Based on analysis they group all the threat agents into the particular category concerning their attacking behaviors identified during the analysis.

Above Figure 7 shows the Histogram of the bar chart with the help of the IPYNB algorithm for each Excel sheet generated during the first phase. The first bar chart shows protocols used in attacking that is on the y-axis the number of Unique IPs and on the x-axis the number of protocols being assessed for them. The same second bar chart shows vulnerability ports i.e., on the y-axis the number of Unique IPs, and on the x-axis the number of ports being assessed for them. Similarly, the third bar chart shows time spend on the network for an attack that is on the y-axis the number of Unique IPs and on the x-axis is time spend on the network in minutes.

The above Histograms for the protocols, ports, and time spent on the network will help to evaluate the three main attributes for the threat agents i.e., motivation, opportunity, and capability. Once we identified the port open during the access of the network, we can determine the opportunity for the groups of threat agents used during the penetration of the network. In the same way, the above Histograms will help us to identify the protocols accessed by the threat agents will lead to evaluate the potential capability of the hacker.
Similarly, the above Figure 8 shows the Histogram of bar charts and analysis of different inputs. The first bar chart shows protocols used in attacking that is on the y-axis the number of Unique IPs and on the x-axis the number of protocols being assessed for them. The same second bar chart shows vulnerability ports i.e., on the y-axis the number of Unique IPs, and on the x-axis the number of ports being assessed for them. Similarly, the third bar chart shows time spend on the network for an attack that is on the y-axis the number of Unique IPs and on the x-axis is time spend on the network in minutes.

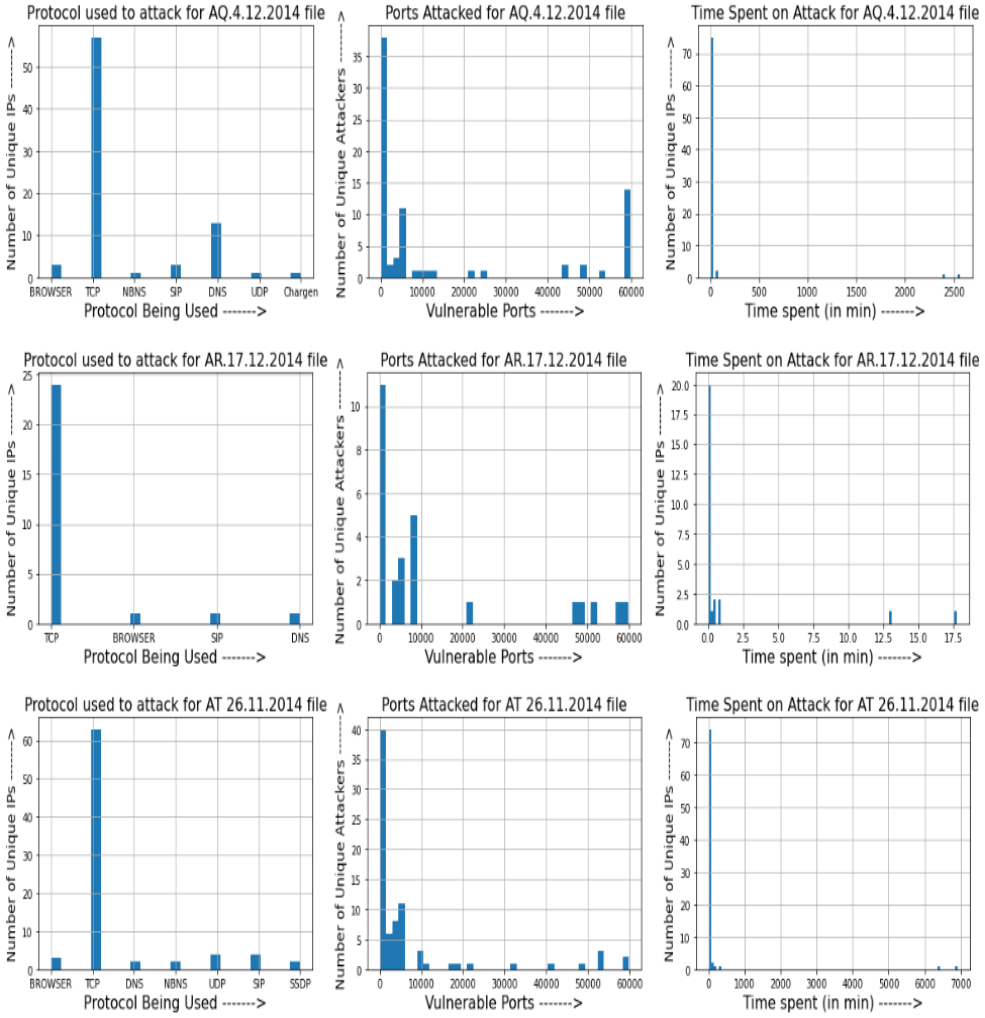

Figure 9 Histogram for each input based on Protocol, Ports, and Time.

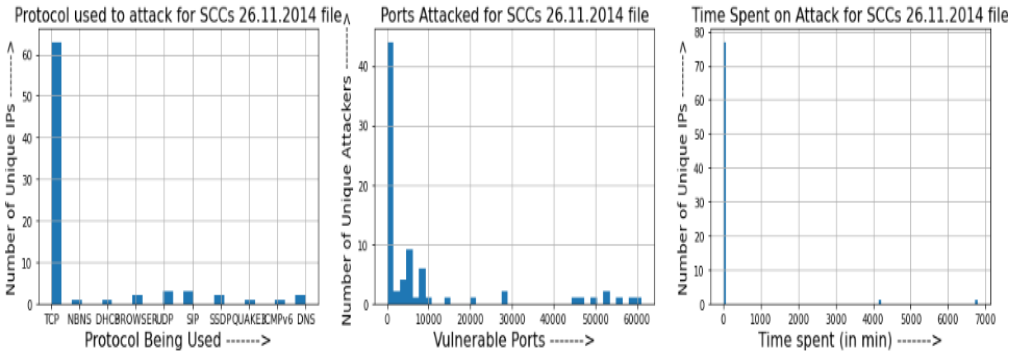

Figure 10 Histogram for each input based on Protocol, Ports, and Time.

In the above Figure 9 and Figure 10, there are two parts to the outputs generated by the. IPYNB file. In the first part, three Histograms are generated for every file in the output-excel sheet, and the second part generates the Histograms on the cumulative data of all the files in the folder.

1. For every file in the output-excel sheet, three Histograms have been generated and all these three Histograms consist of common data at y-axis contains a common data, i.e., 'Number of unique IPs and on the x-axis as follows: -

1.1. The 1st one is a Histogram between ‘Protocols being used’ and ‘Number of unique IPs’ that are using these protocols. This Histogram shows us the protocols being used by the attackers.
1.2. The 2nd Histogram is between the ‘Ports on the host’ which have been targeted and ‘Number of unique IPs’ that targeted them. This Histogram highlights the vulnerable ports.
1.3. The 3rd Histogram is between ‘Time spent’ and ‘Number of unique IPs’. This Histogram highlights how much time an attacker will usually spend to attack a host.

These results will help to identify that the particular groups of threat agents accessing a specific protocol for penetration of the network. Which leads to determining the category of the threat agent. For example in the above Figure 10 TCP protocol used by most of the IPs and mainly targetting the network layers. So, we can conclude that in this analysis the threat agents have mostly distributed denial of services (DDOS).
4 Histograms are generated by using data from all the files in the folder.

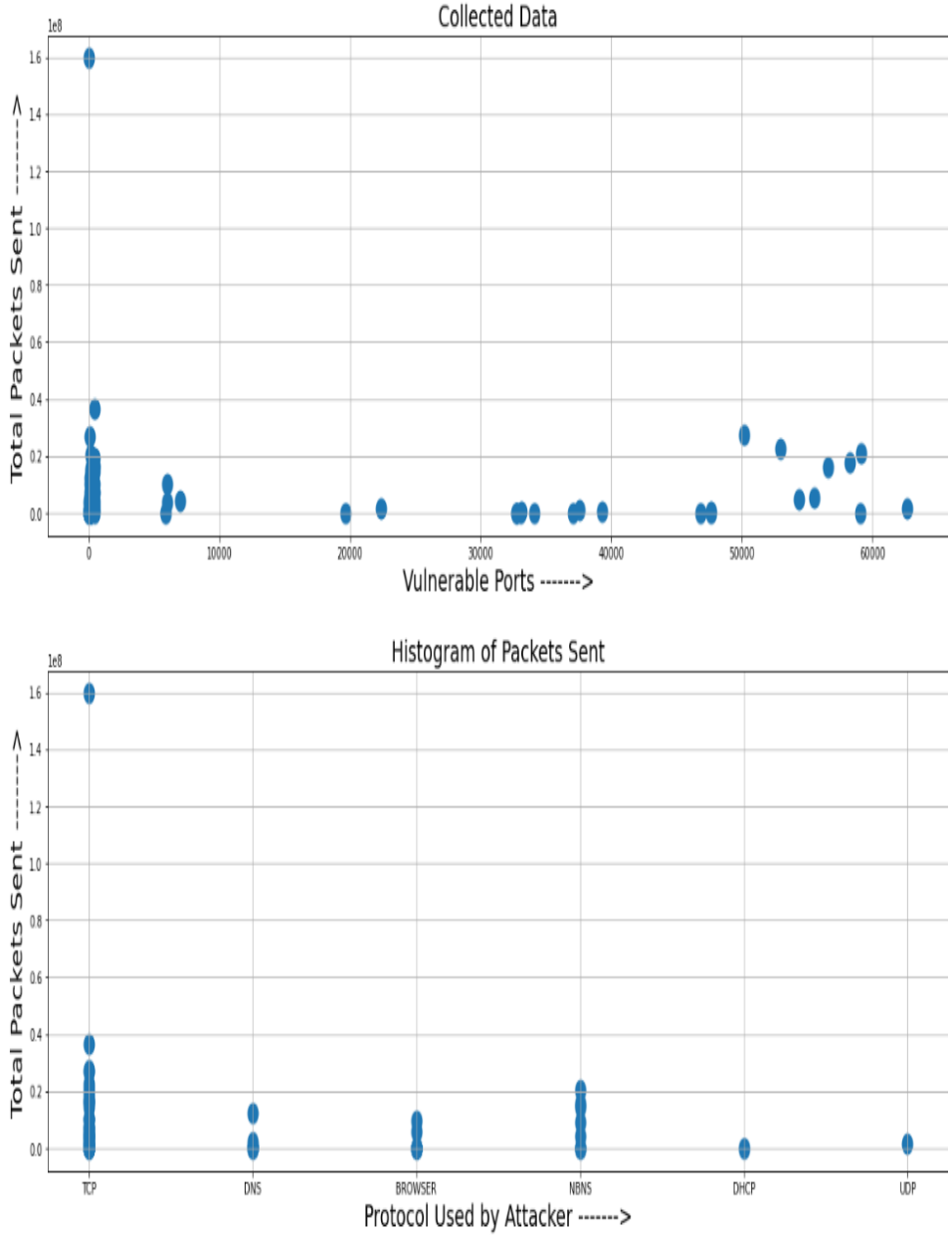

Figure 11 Histogram for Vulnerable Ports, Protocol, and Total Packets

In above Figure 11, These Histograms are based on the accumulated data in the potential output produced in the excel sheets. They used to represent the number of packets generated for traffic during penetration testing, protocols, or layers being used by threat agents and targeting vulnerable ports for achieving the goal.

- The 1st Histogram is between 'Vulnerable Ports' and 'Total Packets Sent'. This data shows how many packets were sent to which port on the host machine.
- The 2nd Histogram is between 'Protocol Used by Attacker' and 'Total Packets Sent'. This data shows the volume of packets for every protocol used to attack the host.

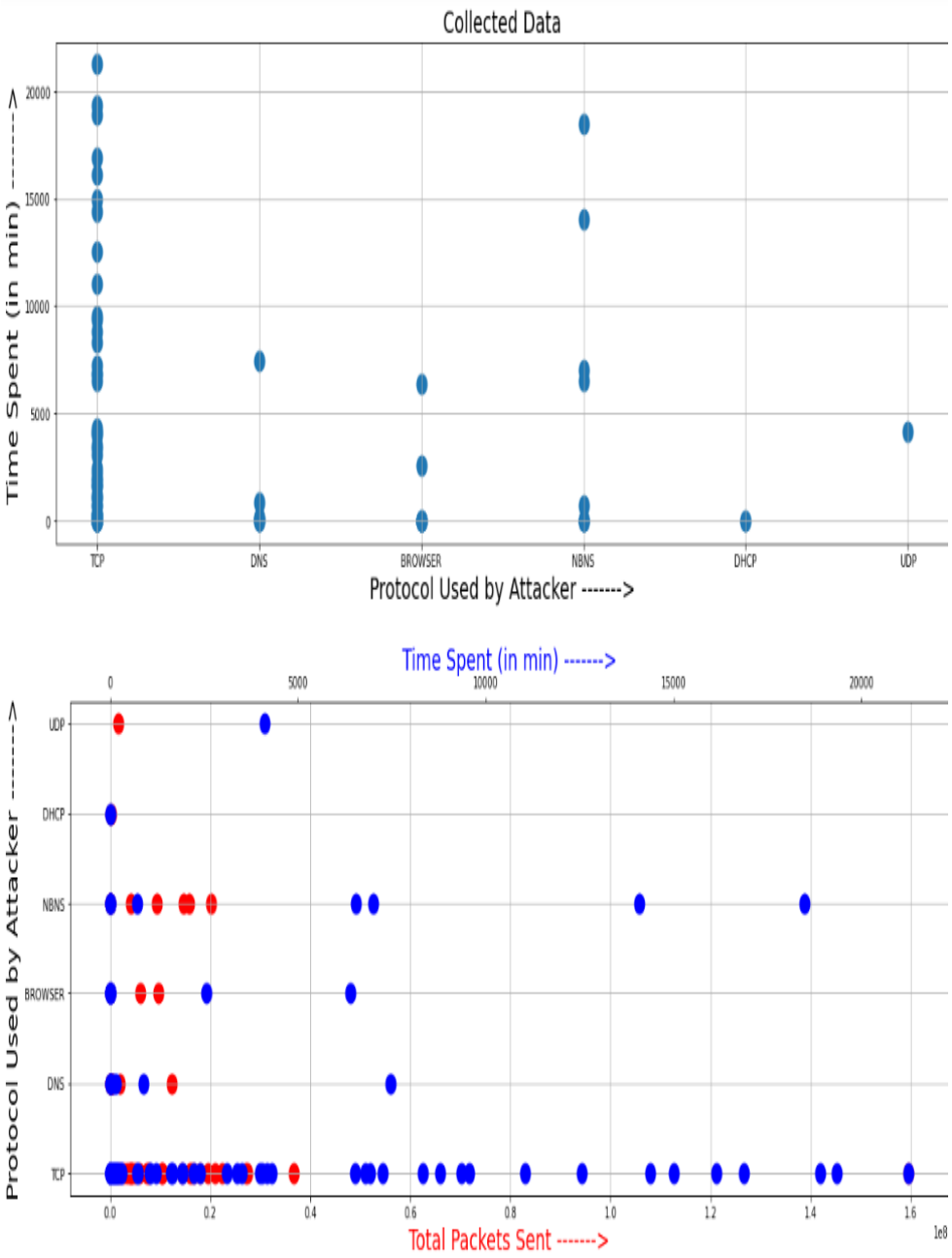

Figure 12 Histogram between Total Packets, Time, Protocol and Collected Data

In above Figure 12, we represent the histogram between the total data collected from each unique IPs, total time spent on the network, and the protocols used for attacking the network.

- The first Histogram is between ‘Protocol Used by Attacker’ and ‘Time spent’. This data highlights the amount spent by the attacker for every protocol used to attack the host.
- The second Histogram has ‘Protocol Used by Attacker’ in the y-axis with both ‘Total Packets Sent’ and ‘Time spent’ on the x-axis. The data points for ‘Time Spent’ are highlighted in blue, whereas the data points for ‘Total Packets Sent’ are highlighted in red. Even though these have different units, it gives us a statistical relative visual of how the time spent by the attacker varies concerning the number of packets sent for the same protocols used.

VI – Conclusion and Future Work.

Threats and threat agent’s risks are emerging in threat assessment of a network for an organization and business of the companies, and security risk management practitioners enable a mechanism to explore these risks and enforce their countermeasures based on the threat agent profiling and determining the Critical Threat Intelligence (CTI) feed to them. This paper presents a semi-automatic model based on the threat assessment of the PCAP files captured by the semi-automatic featured tools during the penetration testing run against the ESXi server of the University of Hertfordshire. The framework captured the data between 2012-2019, illustrates the value of assets stored on the server, motivation, opportunity, and capability of the threat agents while accessing the network. We evaluate the situational awareness data as well by this semi-automatic model of threat assessment by exploring the threat profiles for the historically captured data with the aid tools. Also, we provide the threat agent practitioners an idea of using an automatic model for threat assessment of a network. This research’s findings would support decision-makers management and software developer’s practitioners regarding the role of the building of threat agent profiling for the historical data and determine of Critical Threat Intelligence (CTI) feeds for the threat agent’s groups might be helpful for evaluation of new threats found in the network. In the end, we propose the future research directions work for the threat assessment models and methodology.

In our future work, we aim to build an automatic machine learning-based vulnerability tree analysis security reference model as a security risk management tool to evaluate the security needs of PCAP files or DataStream with sequential requirements of the next to the real-time informational environment. The (CVE) Common Vulnerabilities and Exposures list available on the (NIST) National Institute of Standards and Technology database can be extracted based on the analysis and implementation of Packet Capture Application Programming Interface (PCAP) files captured during the penetration testing against the network. These CVE lists will further be extracted based on their information or footprints captured by the design aid tool led to generate an output as a vulnerability tree for the analysis of threat agents identified in the situational awareness data of a network. According to our analysis and implementation of threat assessment and study of various models and methodology. We suggest that, if the future model able to evaluate both threat assessment and vulnerability assessment for the Packet Capture Application Programming Interface (PCAP) files in an automatic manner with the help of machine learning tools then the complexity will be more effective as compared to the existing model and methodology.

References -

[1] J. A. Iglesias, P. Angelov, A. Ledezma, and A. Sanchis, “Modelling evolving user behaviors,” in 2009 IEEE Workshop on Evolving and Self-Developing Intelligent Systems, 2009, pp. 16–23.

[2] M. Xue, C. Yuan, H. Wu, Y. Zhang, and W. Liu, “Machine learning security: Threats, countermeasures, and evaluations,” IEEE Access, vol. 8, pp. 74720–74742, 2020.

[3] A. Jones, “Identification of a Method for the Calculation of the Capability of Threat Agents in an Information Environment,” Sch. Comput. Pontypridd, Univ. Glamorgan 0-134, 2002.

[4] V. Mavroeidis and S. Bromander, “Cyber threat intelligence model: an evaluation of taxonomies, sharing standards, and ontologies within cyber threat intelligence,” in 2017 European Intelligence and Security Informatics Conference (EISIC), 2017, pp. 91–98.

[5] B. S. Atote, T. S. Saini, M. Bedekar, and S. Zahoor, “Inferring emotional state of a user by user profiling,” in 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), 2016, pp. 530–535.

[6] H. Asgari, S. Haines, and O. Rysavy, “Identification of threats and security risk assessments for recursive Internet architecture,” IEEE Syst. J., vol. 12, no. 3, pp. 2437–2448, 2017.

[7] A. Azaria, A. Richardson, S. Kraus, and V. S. Subrahmanian, “Behavioral analysis of insider threat: A survey and bootstrapped prediction in imbalanced data,” IEEE Trans. Comput. Soc. Syst., vol. 1, no. 2, pp. 135–155, 2014.

[8] S. Vidalis, A. Jones, and A. Blyth, “Assessing cyber-threats in the information environment,” Netw. Secure., vol. 2004, no. 11, pp. 10–16, 2004.

[9] D. M. Cappelli, A. P. Moore, and R. F. Trzeciak, The CERT guide to insider threats: how to prevent, detect, and respond to information technology crimes (Theft, Sabotage, Fraud). Addison-Wesley, 2012.

[10] V. Susukailo, I. Opirskyy, and S. Vasylyshyn, “Analysis of the attack vectors used by threat actors during the pandemic,” in 2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT), 2020, vol. 2, pp. 261–264.

[11] S. Wold, K. Esbensen, and P. Geladi, “Principal component analysis,” Chemom. Intell. Lab. Syst., vol. 2, no. 1–3, pp. 37–52, 1987.

[12] P. A. Legg et al., “Towards a conceptual model and reasoning structure for insider threat detection,” J. Wirel. Mob. Networks, Ubiquitous Comput. Dependable Appl., vol. 4, no. 4, pp. 20–37, 2013.

[13] M. Bishop et al., “Insider threat identification by process analysis,” in 2014 IEEE Security and Privacy Workshops, 2014, pp. 251–264.

[14] E. Morakis, S. Vidalis, and A. Blyth, “Measuring vulnerabilities and their exploitation cycle,” Inf. Secure. Tech. Rep., vol. 8, no. 4, pp. 45–55, 2003.

[15] S. Vidalis and A. Jones, “Threat Agents: what InfoSec officers need to know,” Mediterr. J. Comput. Secure., 2006.

[16] A. Sogbesan, A. Ibidapo, P. Zavarsky, R. Ruhl, and D. Lindskog, “Collusion threat profile analysis: Review and analysis of MERIT model,” in World Congress on Internet Security (WorldCIS-2012), 2012, pp. 212–217.

[17] A. Erola, I. Agrafiotis, J. Happa, M. Goldsmith, S. Creese, and P. A. Legg, “RicherPicture: Semi-automated cyber defense using context-aware data analytics,” in 2017 International Conference On Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA), 2017, pp. 1–8.

[18] U. D. Deore and V. Waghmare, “Cybersecurity automation for controlling distributed data,” in 2016 International Conference on Information Communication and Embedded Systems (ICICES), 2016, pp. 1–4.
[19] P. A. Legg, O. Buckley, M. Goldsmith, and S. Creese, “Automated insider threat detection system using user and role-based profile assessment,” IEEE Syst. J., vol. 11, no. 2, pp. 503–512, 2015.
[20] G. Pogrebna and M. Skilton, “The Twelve Principles of Safe Places,” in Navigating New Cyber Risks, Springer, 2019, pp. 171–197.
[21] A. Iskandar, E. Virma, and A. S. Ahmar, “Implementing DMZ in improving network security of web testing in STMIK AKBA,” arXiv Prepr. arXiv1901.04081, 2019.
[22] S. Vidalis and A. Jones, “Analyzing Threat Agents and Their Attributes.,” in ECIW, 2005, pp. 369–380.
[23] R. Rubini, A. Porta, G. Baselli, S. Cerutti, and M. Paro, “Power spectrum analysis of cardiovascular variability monitored by telemetry in conscious unrestrained rats,” J. Auton. Nerv. Syst., vol. 45, no. 3, pp. 181–190, 1993.
[24] B. Shin and P. B. Lowry, “A review and theoretical explanation of the ‘Cyberthreat-Intelligence (CTI) capability that needs to be fostered in information security practitioners and how this can be accomplished,” Comput. Secure., vol. 92, p. 101761, 2020.
[25] R.-C. Chen, K.-F. Cheng, and C.-F. Hsieh, “Using rough set and support vector machine for network intrusion detection,” arXiv Prepr. arXiv1004.0567, 2010.
[26] A. Rynes and T. Bjornard, “Intent, capability, and opportunity: A holistic approach to addressing proliferation as a risk management issue,” Idaho National Laboratory (INL), 2011.
[27] J. E. Y. Rossebo, F. Fransen, and E. Luiijf, “Including threat actor capability and motivation in risk assessment for Smart GRIDs,” in 2016 Joint Workshop on Cyber-Physical Security and Resilience in Smart Grids (CPSR-SG), 2016, pp. 1–7.
[28] G. Saygili, E. M. Rathje, Y. Wang, and M. El-Kishky, “Cloud-based tools for the probabilistic assessment of the seismic performance of slopes,” in Geotechnical Earthquake Engineering and Soil Dynamics V: Slope Stability and Landslides, Laboratory Testing, and In Situ Testing, American Society of Civil Engineers Reston, VA, 2018, pp. 19–26.
[29] H. J. Van Veen, N. Saul, D. Eargle, and S. W. Mangham, “Kepler Mapper: A flexible Python implementation of the Mapper algorithm.,” J. Open Source Softw., vol. 4, no. 42, p. 1315, 2019.
[30] S. Narkar, B. L. Thomson, and P. A. Fox, “Designing for 2030: The Impact and Potential of Virtual Laboratories,” in AGU Fall Meeting Abstracts, 2020, vol. 2020, pp. IN003-03.

Biography-

Gaurav received his bachelor's degree in computer science from Gautama Buddha University, Lucknow, India in 2010, and his master's degree in VLSI and CAD systems from Thapar University Punjab, India in 2012. He worked as Research Assistant in a Modular threat assessment project by Innovative United Kingdom (IUK). He is currently working as Visiting Lecturer and perusing a Ph.D. in the cybersecurity research group at the University of Hertfordshire Hatfield, UK- School of computer science. he is focusing on the Real-Time Semi-Automated Threat Assessments in Informational Environment. His areas of research include Cloud Computing, Internet of Things, Cybersecurity, Cryptography, and Security in Wireless Sensor Networks.

Dr. Stilianos Vidalis 's involvement in the Information Operations arena began in 2001. He has participated in high-profile, high-value projects for large international organizations and Governments. He has collected and analyzed information for prestigious European financial institutions, applying international standards under the context of risk and threat assessment. He was training British Armed Forces personnel in penetration testing and digital forensics. He has developed and published in peer-reviewed scientific journals his own threat assessment methodology and other aspects of his work on threat agent classification, vulnerability assessment, early warning systems, deception in CNO, id theft, and computer criminal profiling. He is currently responsible for developing training courses in cybersecurity, cyber intelligence, and digital forensics at the University of Hertfordshire, England.

Dr. Catherine Menon is a senior lecturer at the University of Hertfordshire. She has led the Soc-Cred project, supported by the Assuring Autonomy International Programme and Lloyd's Register. She has been involved in several industry-funded projects, including the UK MOD SSEI consortium. She is a member of the BSI AMT/10 (Robotics) and AMT/10/1 (Ethics for Robots and Autonomous Systems) Committees, as well as the IET Committee for the Safety and Security Code of Practice and the ISO SC7 working group. She is also a member of the IEEE P7000 Working Group, developing a standard for the fail-safe design of autonomous and semi-autonomous systems.

Dr. Niharika Anand received a Ph.D. degree from the Indian Institute of Information Technology, Allahabad, India. She is currently working as an Assistant Professor with the Department of Information Technology Indian Institute of Information Technology, Lucknow, India. Her areas of research include Cloud Computing, Internet of Things, Cyber Forensics 3-D Wireless Sensor Network, Wireless Sensor Network Localization, Wireless Sensor Network Topology Control and Maintenance.