electronics

MDPI

Article

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

Gaurav Sharma [1,*], Stilianos Vidalis [1], Catherine Menon [1], Niharika Anand [2] and Somesh Kumar [3]

[1] School of Computer Science & Engineering, University of Hertfordshire, Hatfield Al10 9AB, UK; s.vidalis@herts.ac.uk (S.V.); c.menon@herts.ac.uk (C.M.)
[2] Indian Institute of Information Technology Lucknow (IIITL), Lucknow 226002, India; niharika@iiitl.ac.in
[3] ABV-Indian Institute of Information Technology & Management, Gwalior 474015, India; somesh@iiitm.ac.in
* Correspondence: g.gaurav@herts.ac.uk

Abstract: Threat assessment is the continuous process of monitoring the threats identified in the network of the real-time informational environment of an organisation and the business of the companies. The sagacity and security assurance for the system of an organisation and company's business seem to need that information security exercise to unambiguously and effectively handle the threat agent's attacks. How is this unambiguous and effective way 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. However, 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 is ineffective, due to which security assurance arises for an organisation and the business of companies. This paper proposes a semi-automatic information security model, which can deal with situational awareness data, strategies prevailing information security activities, and protocols monitoring specific types of the network next to the real-time information environment. This paper looks over analyses and implements the threat assessment of network traffic in one particular real-time informational environment. To achieve this, we determined various unique attributes of threat agents from the Packet Capture Application Programming Interface (PCAP files/DataStream) collected from the network between the years 2012 and 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 determined the threat profiles, critical threat intelligence (CTI), and complexity of threat agents that are not addressed or 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

Citation: Sharma, G.; Vidalis, S.; Menon, C.; Anand, N.; Kumar, S. Analysis and Implementation of Threat Agents Profiles in Semi-Automated Manner for a Network Traffic in Real-Time Information Environment. *Electronics* **2021**, *10*, 1849. https://doi.org/10.3390/electronics10151849

Academic Editor: Krzysztof Szczypiorski

Received: 2 July 2021
Accepted: 28 July 2021
Published: 31 July 2021

1. Introduction

Identifying the potential cybersecurity threat capability in real-time is a crucial activity. It helps provide practical information about the threat in a network that allows 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 organisation is typically achieved manually by the existing models and methodology. Threat assessment is implemented in an automated manner with the help of machine learning techniques and various real-time models [2]. The behaviours of threat agents are erratic, and the goals of threat agents change with time. Threat agent groups change their behaviour to penetrate a network based on motivation, opportunity, and capability [3,4]. The motivation of the threat agent constantly changes with time depends on the financial gain, revenge

from an organisation, etc., and the type of environment targeted. 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 behaviour, source I.P. address, destination I.P. address, number of open ports, number of packets generated, location of the threat agent, and time spent on the network with minimal user intervention [5]. The user has minimal intervention because of the footprints captured by the capturing data tool like LibPcap, WinPcap, PCAPng, NPcap, etc., during threat assessment in the form of PCAP files that cannot be altered by the potential threat agent while traversing an organisation's network. The threat agent cannot alter because once they generate the packets in the network, they cannot erase the footprint of generating the packets because of the accessing property of the network. This research attempts to recognise the aspects of profiling and deliver solutions by implementing the profiling of threat agents. Threat profiling is an essential aspect of performing threat assessment for an organisation. Suppose we have the threat profile for the historically identified threat agents from the network of an organisation. In that case, we can use these profiles as references while executing the threat assessment for the situational awareness data captured from the network. The model can address the recent threat agent effectively identified from a network with optimised complexity.

It has been accepted that continuous threat assessments practice mitigate the risks for any organisation and business [6]. However, in the modern, socially driven, virtual computing era, threat assessments are hindered by a lack of resources, complexity, and data size [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 understand the situational awareness property of data and threat assessments required for an organisation. University computer emergency response team (CMU-CERT) groups determined three critical groups of threat agents, i.e., the technology of organisation sabotage, compromising with intellectual property, and data stream fraud [9]. The number of growing cases highlighted by internet media in recent years revealed that both business organisations and government organisations suffered a similar experience. In contrast, the priority information has been filtrated by the organisation's internal users and shared with the threat agents [10]. The threat agents require serious attention from both users and organisations.

Referencing to the COVID-19 pandemic nowadays, organisations and businesses share their file and documents frequently with the help of the internet to run their business. It is now standard practice for users of the organisation to have admittance to large repository documents which are electronically warehoused on distributed file servers. Many organisations offer company laptops and desktops to the users for work while using e-mail to organise and schedule/rescheduling meetings. Amenities such as video conferencing are repeatedly used for holding meetings throughout the world, and users of an organisation are continuously connected to the internet. The electronic nature of the files and records of an organisation on the internet makes it easier for the threat agents to attack the organisation. On the advantageous side of continuous threat assessment, an organisation can easily capture the activity logs of the internal threat agent while analysing their captured packets [11]. However, practically analysing such activity logs is infeasible due to the high volume of activities performed by the user every day.

In this work, we present an efficient model for threat detection and analysis based on the conception of anomaly detection. The proposed model implements the threat agent profiles from the PCAP files and determines the cyber threat intelligence based on evaluating motivation, opportunity, and capability of threats. With the help of these profiles, comparisons can be populated that show the current observations fluctuate from the previous observations. To assess the performance of the tactic, we extracted the valuable information from the PCAP files in a semi-automated manner, and output has

been 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. The system executed expressively soundly for detecting the attacks, and the visualisation of reports enabled us to remember which attributes help determine M, O, C factors for the threat agents. This paper illustrates all the threats identified in a network captured during the penetration testing against the ESXi server of the University of Hertfordshire, UK.

The rest of this paper is as follows. Section 2 discusses the related work. Section 3 labels the necessities of analysis, the experimental set-up of the proposed system, and describes how to evaluate motivation, capability, and opportunity of threat agents. Section 4 presents the actual results from practical experimentation of the system, and Section 5 concludes this paper.

2. Related Work

The field of threat agents profiling and analysis of cyber threat intelligence has recently received ample attention. Researchers have proposed an assortment of different models and methodologies designed to detect or prevent attacks [12,13]. Likewise, Vidalis et al. [8] 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. They compare the TAME (Threat Assessments Model For EPS) methodology with other existing methods based on the number of parameters as sting, effectiveness, and understanding of information security from the threat. TAME is the upgraded version of METEORE 2000 for the micropayment system (MPS). In the initial phases, the authors analyse 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 the 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:

(a) Scope of Assessments.
(b) Threat Agent and Vulnerability Analysis.
(c) Scenario Construction and System Modelling.
(d) Evaluation.

According to these phases, TAME determined how much security is required for a particular organisation and business of the system. All four stages are working simultaneously, and one input from a phase becomes the output of another degree. Similarly, the vice-versa of inputs and outputs are generated from the TAME, and it depends on the requirements of threat assessments. The authors conclude the TAME by using the assessor as an asset for better understanding and analysing an organisation's systems.

Morakis et al. [14] measure vulnerabilities and their exploitation cycle by various tools such as COPS, NESSUS, SYSTEM SCANNER, RETINA, NET RECON, WHISKER, and CYBER COB. In this work, the authors address a problem faced by a large amount of data in the informative environment is cyber-attacks. The authors propose a vulnerability tree analysis to address such issues faced by several organisations for a long time. They believe in constructing knowledge information concerning a specific domain in an object-oriented hierarchy tree and building a formal model to analyse them concerning possible scenarios of attacks faced by the computer systems. The primary purpose of this is to provide a depth classification of vulnerabilities, find why such attacks happened on a particular data/asset, and analyse footprints and scenarios of threat agents to exploit vulnerabilities. The main aim of the vulnerability tree analysis is to identify the attacks in the early stages and address them before severe damage to real-world informational systems. Here, the authors illustrate the various tools capable of analysing the vulnerability of complex organisational environments; such tools are COPS, NESSUS, SYSTEM SCANNER, RETINA, NET RECON, WHISKER, and CYBER COB, etc. However, these are not adequate in today's modern

electronic era of cyber-crime because they cannot address hazards like fault-tree analysis, checklists, event-tree analysis, cause-consequences analysis, etc. To cope with such hazards, the authors combine these tools of vulnerabilities tree analysis with object-oriented trees (O.O.) and adequately address such hazards concerning Boolean Mathematics.

Gerald L. et al. [15] briefly explain about threat agents regarding how they can have unauthorised 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 authors also analyse the digital attacks that occured in 2002 in several countries. They identify that the threat agents of real-world informational environments consist of:

(a) Threat agent catalogue.
(b) Historical data.
(c) Technical report enterprises.
(d) Reports of business environments.
(e) Reports of physical environments.
(f) Recent knowledge/information.
(g) Current knowledge of stakeholders.
(h) Current knowledge of the staff.
(i) List of stakeholders.

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

Adetorera Sogbesan et al. [16] developed a model to identify the MERIT (Management & Education of Risk of Insider Threat) based on the study of insider threat concerning the institute of CERT/USSS. This MERIT provides the facility to mitigate the insider threat of an organisation, and the key finding is to make the case study of individual threat agents, i.e., collision threat. MERIT models the case studies on the insider threat for an organisation, and based on that, threat assessments have been conducted to determine the impact of danger on the business. They also show some figures for losses based on studies done by USSS/CERT. They categorise the insider attack based on the ex-employee, or the financial gain of any vital position held by an employee in an organisation. Based on the number of organisations, 69% of companies measured stated data theft events (not external attacks). These threats were originated from inside the organisation. At the same time, a massive 91% of companies testified not having operative detection systems for recognising an insider threat. The MERIT model has a limitation/shortcoming in analysing compressive pattern analysis based on motivation and behavioural characteristics. The motivation factor of collusion attack is not able to be addressed by the MERIT model. This model is not able to explain the capability of an insider threat.

Casillo, M. et al.'s study [17] "Embedded Intrusion Detection System for Detecting Attacks over CAN-BUS" designs a model based on AIC (availability, integrity, and confidentiality). The authors address the issues related to cyber-attacks on the automotive vehicle system. They introduce the automotive IDS embedded method for the CAN (controlled area network) BUS. Referencing the Bayesian network approaches, identifying malicious messages to the connected devices to the vehicles is accomplished. In this paper, the authors identify the snag for the IoT devices connected to automotive vehicles and their attacks while using automation. They suggest machine learning approaches, particularly the Bayesian network approach to cope with the cyber-attacks on the CAB bus. The authors used the CARLA simulator to provides the solution. The PYTHON library and

several APIs were cast off for clustering the data and FPGA techniques for developing the model's architecture.

Lombardi, M. et al.'s research [18] "EIDS: Embedded intrusion detection system using machine learning to detect attack over the can-bus" introduced an IDS approach to identify the threats in the automated vehicles, particularly CAN (controlled area network) bus. The authors cast off the development of an IDS approach with the help of machine learning techniques through the Bayesian network approach to detect possible attacks on the CAN bus. The main benefit of developing an IDS approach was using the embedded framework for designing and determining the non-linear messages flow. The castigate faced by the connected IoT devices and the intelligent device for self-driving vehicles was identified with the help of an introduce IDS approach in the research.

These related works draw an intense observation that access to a real-world data stream is enormously challenging. Thus, researchers synthesise data into several groups based on the threat agents identified in a network. The existing model and methodology did threat assessment manually, due to which their complexity is exorbitant. This research predominantly wants to epitomise the volume and variety of data analysed 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 epitomised by our data collected from a real-world in a specific environment and show how our profiling and CTI system of threat agents would detect the different attacks based on the patterns identified.

3. Experiment Set-Up and Evaluation of MCO Attributes

The work described in this research has been carried out as part of a more comprehensive interdisciplinary project that includes computer security researchers and cyberpsychology experts. CTI data–driven threat agent profiling can be used for determining the motivation, opportunity, and capabilities attributes of threat agents under the context of a continuous threat assessment [19]. The threat remains of budding apprehension to governments and businesses organisation, and it becomes an acute necessity for practical tools to help mitigate the threat posed. The modern risk assessment models recognise a need to perform several threat assessments to identify and analyse various threats in the contemporary information environment. If we conduct iterative threat assessment for the network, then with the help of designing the profiling prepared by practitioners, a new type of threat agents identified in situational awareness data will be addressed quickly. The continuous threat assessments help generate the paradox of warning to the cyber operations performed in the information environment. This paper identifies the research gap in semi-automated information environments, which consists of large heterogeneous infrastructures, hosting a large amount of data collected from different types of platforms or environments [20]. The different types of platforms mean different kinds of environment and the conditions used by the threat agent to attack the particular network. To identify the solution for such a large amount of data, decision aid tools should understand situational awareness and critical intelligence feeds of the threats in real-time information environments.

In the modern knowledge-based, socially driven, virtual computing era, threat assessments are hindered by lack of resources, complexity, and data size. Information environments are large heterogeneous infrastructures, hosting a large amount of data collected from different platforms with the help of many tools. 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 to identify suspicious traffic, malicious intentions, and network attacks in an automated manner. State of the art on threat assessment models and methodologies will be considered in this project, while procedural and technology issues will be resolved by applying cyber analytics principles [21].

3.1. Experimental Environment of the System

Figure 1 shows the testing set-up through which we execute the penetration testing against the specific condition of the platform or environment. The number of VPNs used to connect with the REDNET network and connect through the firewall saves the data from unauthorised access. Further, REDNET connects to DMZ (Demilitarized Zone), the number of V.M.s, and public I.P. of staff to control the activities. BLUENET connects to the user's V.M.'s I.P.s, ESXi server, UH CSC WIFI (University of Hertfordshire Wi-Fi), and public I.P. of staff. In this environment, the PCAP files are collected from the server with the help of the Wireshark tool [22]. Other tools like SolarWinds Deep Packet Inspection and Analysis, Paessler Packet Capture, ManageEngine NetFlow Analyzer, Omnipeek Network Protocol Analyzer, TCPdump, and WinDum, etc. are also available. Still, Wireshark is more efficient in extracting useful information from PCAP files and provides the advantage of saving the information in CSV formats. Figure 1 shows the source of the attack I.P. address and the destination of the attack I.P. address through which penetration is executing on the network. The role of DMZ is to stop the hacker at the threshold point, and henceforth, no one is allowed to do access excluded the administrator of the server [23]. The BLUENET refers to the internal security team that defends against real-world attackers. Red Teams of 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.

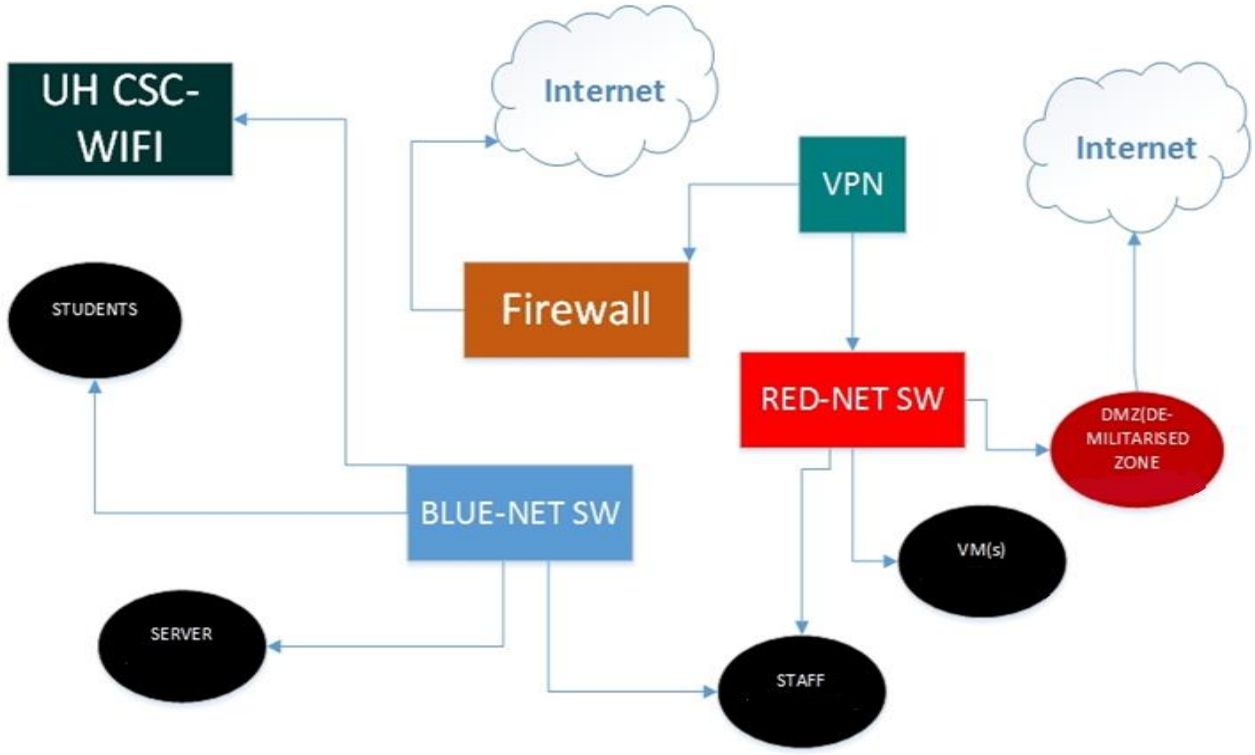

Figure 1. Penetrating Testing Setup at Cybersecurity Laboratory.

3.2. The Architecture of System

The primary purpose of Figure 2 is to understand how the attacker groups generate 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 I.P. addresses, and accessing the files from the databases. The architecture in Figure 2 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 the Blue ESXi security zone, and DMZ (Demilitarised Security Zone). In this server, all the data and information of the University of Hertfordshire are available,

and a dedicated environment installed on V.M.s is available for the attackers. Black ESXi connected to 27 x juniper srx240 and srx340 firewalls via 27 x lab system multiple images of the environment and dedicated interface in red, blue, and black networks. DMZ's role is to stop the hacker at the threshold point to control further damage by the attacker groups.

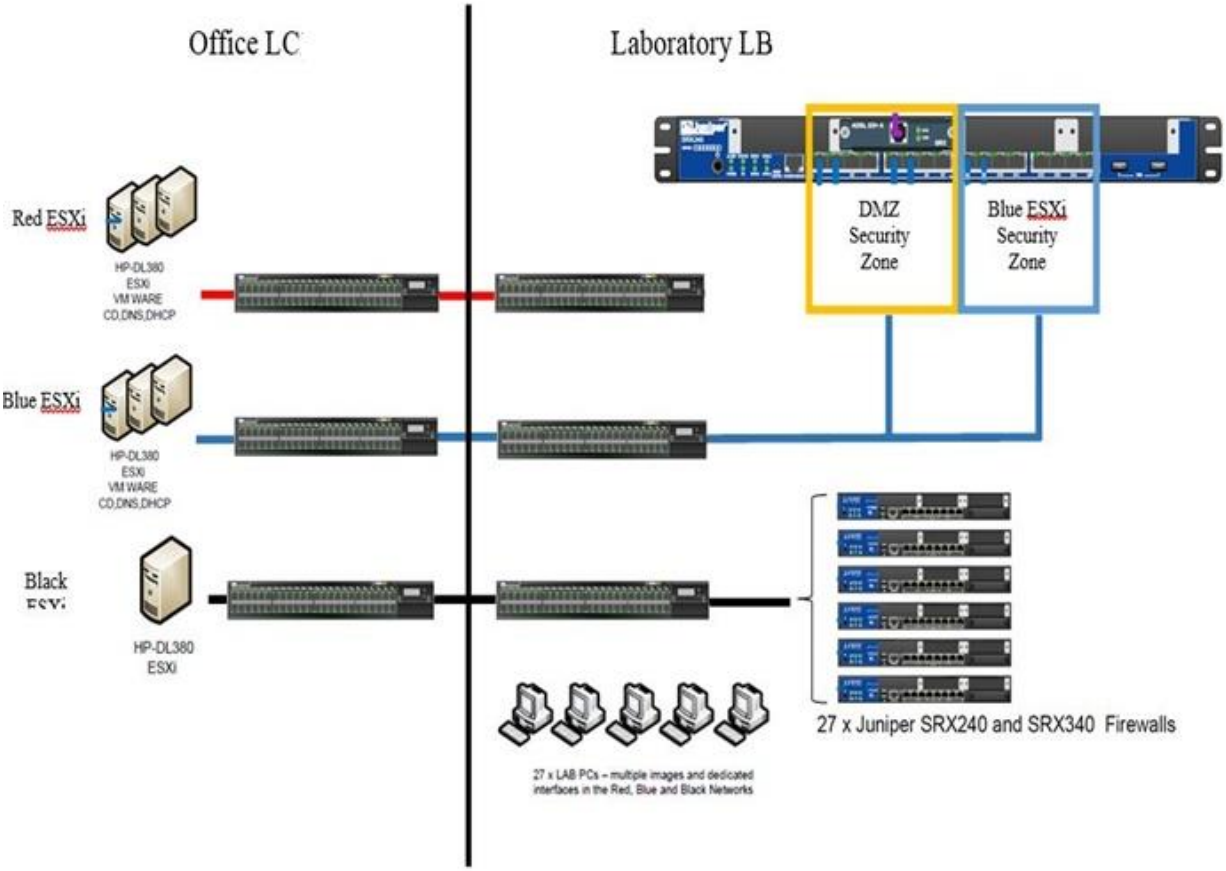

Figure 2. Architecture of System.

3.3. Evaluation of Motivation, Capability, and Opportunity

The threat assessment is a continuous process to collect the PCAP files from the network in an informative environment. The evaluation of the impact of threat agent groups on the organisation or the business, determining the value of assets, vulnerability identification, and threat agent's footprint attributes play a prominent role in the calculation [24]. In Figure 3, the representation of main characteristics in a 3-dimensional matrix is shown, which needs to be addressed by the model while performing threat assessments of the real-time network.

A threat assessment is a statement of threats related to vulnerabilities of company assets and threat agents and a message of believed capabilities that those threat agents possess. In Equation (1), 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 organisation of the nation.

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

The threat can be evaluated in the above Equation (1) when the extracted attribute from the PCAP files is analysed. Then, based on the analysis of characteristics, motivation evaluation can be achieved. Similarly, when the model identifies the open port and the vulnerable ports from the extracted attributes, opportunity can be evaluated. In the same way, the model amalgamating all the information of motivation and opportunity leads to

assess the capability and impact on the assets by the threats. So, the function (f) can be evaluated using motivation, opportunity, capability, and impact of acquisitions.

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

From Equation (2), the function $F(X)$ represents the threat assessment of the model for all the captured files, *Cap* stands for capabilities, *Opp* is an opportunity of the threat agent, *Mto* is motivation, $V(VIA)$ stands for the value of intangible assets, Y is for threat assessments, and T stands for time complexity.

The threat assessment can be evaluated by amalgamating all results determined by the function for the motivation, opportunity, capability of threat agents, and value of intangible assets of environments. Similarly, vulnerability exploitation of assets concerning the CVE list available on the Nation Institute of standard and technology (NIST) database, the impact of threat agents on an organisation's assets, and the time complexity to evaluate all the parameters of the threat agents can be assessed.

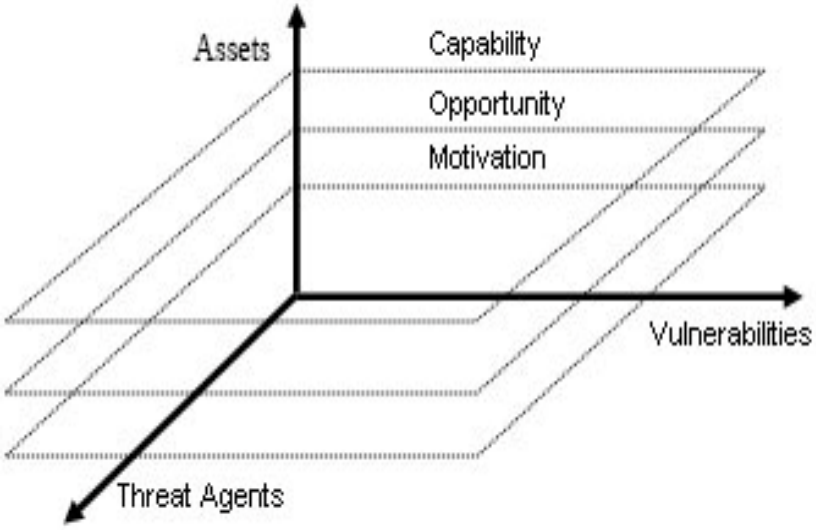

Figure 3. Three-Dimensional Matrix and 3D Representation of Threat Assessment.

3.3.1. Motivation

The evaluation of motivation for threats is the problematic part. It could be determined with the help of analysis of hacktivism branded attacks by groups of assessment models and the network's vulnerability in next to real-time semi-automated information environments. The motivations of attackers are constantly changing, and it could be noticed by the growing rate of hacktivism attacks by different groups of people. It can also see differences in unique motivations based on each group. Motivation is the degree to which a threat agent is prepared to implement a threat. The motivational factors are the elements that drive a threat agent to consider attacking a computer system. Some common motivations for threats include [25]:

a. Profit (direct or indirect).
b. Direct grudge.
c. Fun / Reputation.
d. Further access to partner/connected systems.
e. Political.
f. Secular.
g. Personal gain.
h. Religious.
i. Revenge.
j. Power, terrorism.
k. Curiosity.

3.3.2. Capability

The capability of threats is determined by analysing risk assessment models and the network vulnerability in a next to real-time semi-automated information environment [15].

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

In Equation (3), the risk of the threat agent can be evaluated by the combination of threats, the vulnerability identified for the threat concerning the CVE list of the NIST database and identified consequences of the threat agents.

$$\text{Threat} = \text{Intent} \times \text{Capability} \quad (4)$$

Similarly, in Equation (4), the capability of the threat will be evaluated by the multiplication of intention of the threat agents determined by the model and the overall capability of the threat agent. Further, vulnerability exploitation is achieved with the help of several kali Linux tools such as NESSUS, SAINTS, WHISKER, SARA, etc. The initial phase of the automatic version of the threat assessment model is collecting the DataStream/ PCAP files 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 will be extracted in a semi-automatic manner with the 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 I.P. Address, Destination I.P. Address, Source port, Destination port, Total Packet Length, City, Region, Country, Latitude, Longitude, and Internet Service Provider. 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 valuable information available about the threat in the PCAP files, such as time spent on the network, location of their I.P.s, and environment used by them while penetrating the server.

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. We use all this information to extract all critical threat intelligence (CTI) from these threats to determine the threats' capability, opportunity, and motivation. 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 shown in Figure 3, the motivation of these threat agent groups can be calculated based on the environment used by them, the type of activities executing during the process, factors responsible for digging information, and data from the server.

In the first phase of the model, an algorithm was executed against the PCAP files captured from the ESXi server and extracted the unique attributes from the PCAP files I.P. addresses, such as time (in min), Highest Protocol, TCP protocol, Source I.P. Address, Destination I.P. Address, Source port, Destination port, Total Packet Length, City, Region, Country, Latitude, Longitude, and Internet Service Provider. When the model has all this information about the attacker, the next phase model extracts the location of the threat agents from where they generate the traffic in the network. The model considers only those threat agents for location identification who have generated more than 1000 packets in the network. The model considers the threshold point based on the level of skill or knowledge the threat agent showing while traversing the network. Likewise, if considered less than 1000 packets generated I.P. address of threat agent, then the exploitation of vulnerable port is significantly less or can be ignorable. It is the primary reason for a semi-automatic model to provide the optimised time complexity for threat assessment of an organisation.

3.3.3. Opportunity

Similarly, opportunity can be calculated by identifying the number of open ports, the number of protocols that have unrestricted access and would be vulnerable, and what other factors help a hacker do unauthorised access to the server. The model will evaluate

all this information by initialising the PCAP file captured by the model. The model will determine the open ports with the help of various tools like NMAP, NS-LOOKUP, DIPScan, etc. The combination of all the information about such attributes led to the evaluation of the opportunity of the threat agent groups.

4. Results and Discussion

4.1. State-of-the-Art Algorithms

Many different models are used to perform threat assessment for a network in an informational environment on specialised datasets, where some of the datasets are discussed in the previous section. 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 in the experiment. Similarly, the critical threat intelligence [26] feed is evaluated from the 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 generate the footprints used by the threat agents during traversing network of the server.

The PYTHON script 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 initially reported [27]. Secondly, we calculated the performance of the threat agent followed by our proposed three-dimensional metrics, i.e., motivation, opportunity, and capability. Figure 4 shows that the input is an enormous number of heterogeneous PCAP files captured during the experiment. 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 I.P. Address, Destination I.P. 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-and-results-of-Threat-Assessment-analysis (accessed on 8 May 2021) [28]. Furthermore, with the help of these unique attributes, we can determine the capability and opportunity of the threat agents [29]. Based on the footprints followed by the threat agents during the analysis, we can determine the motivation factor for attackers.

Some of the captured PCAP files were corrupted during the experiment, and the PYTHON program list of crashed files generated during the investigation can be fetched as shown in Figure 5. We also checked all these crashed files manually and with other analysis tools. We found the same result that no information can be extracted from these files. There may be some capture issue or the connection lost on the hacker's end during the network establishment. The time complexity to generate the unique I.P.s with information attributes can also be evaluated from this experiment. This is the unique feature of this model as compared to the existing model and methodologies. This could happen because of the use of semi-automatic approaches for threat assessment of networks next to the real-time informational environment.

4.2. Workflow and Comparative Experiments

As per the previous discussion, the output is generated in the 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 previous phase's output to provide a semi-automatic feature to the model [30]. This novel approach helps us optimise the threat assessment's complexity against the network of influential organisations. This paper also shows the process of

using ML libraries of PYTHON on TensorFlow and automatic techniques of the JUPYTER notebook to identify the unique tuples of DataStream/PCAP files. This approach mainly depends on the chronological order of packets in PCAP files. Here, we first make groups of all the unique I.P.s extracted from raw PCAP files captured from the network with the help of Wireshark. The grouping of all unique I.P.s based on their attributes and characteristic features was identified during the analysis and implementation of DataStream.

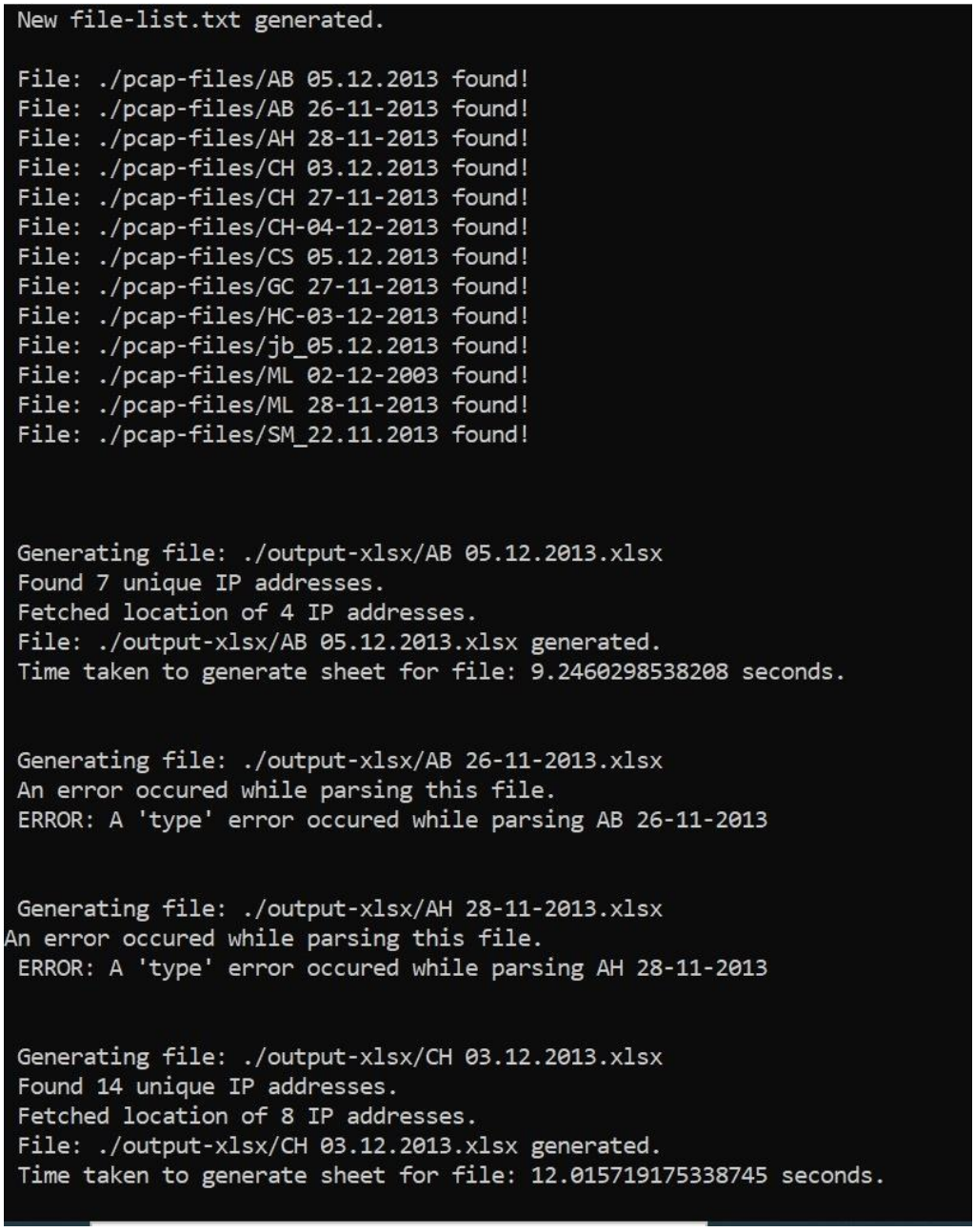

Figure 4. Workflow for raw PCAP file traffic-based feature extraction and experimental results for Unique I.P. addresses with Time complexity.

Similarly, the potential output generated in the previous phase is 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 stage, we generated the Excel sheet for each captured PCAP file consist of helpful information like ports open. They are operating on that layer: time spent on the network, location of the threat agent, etc.

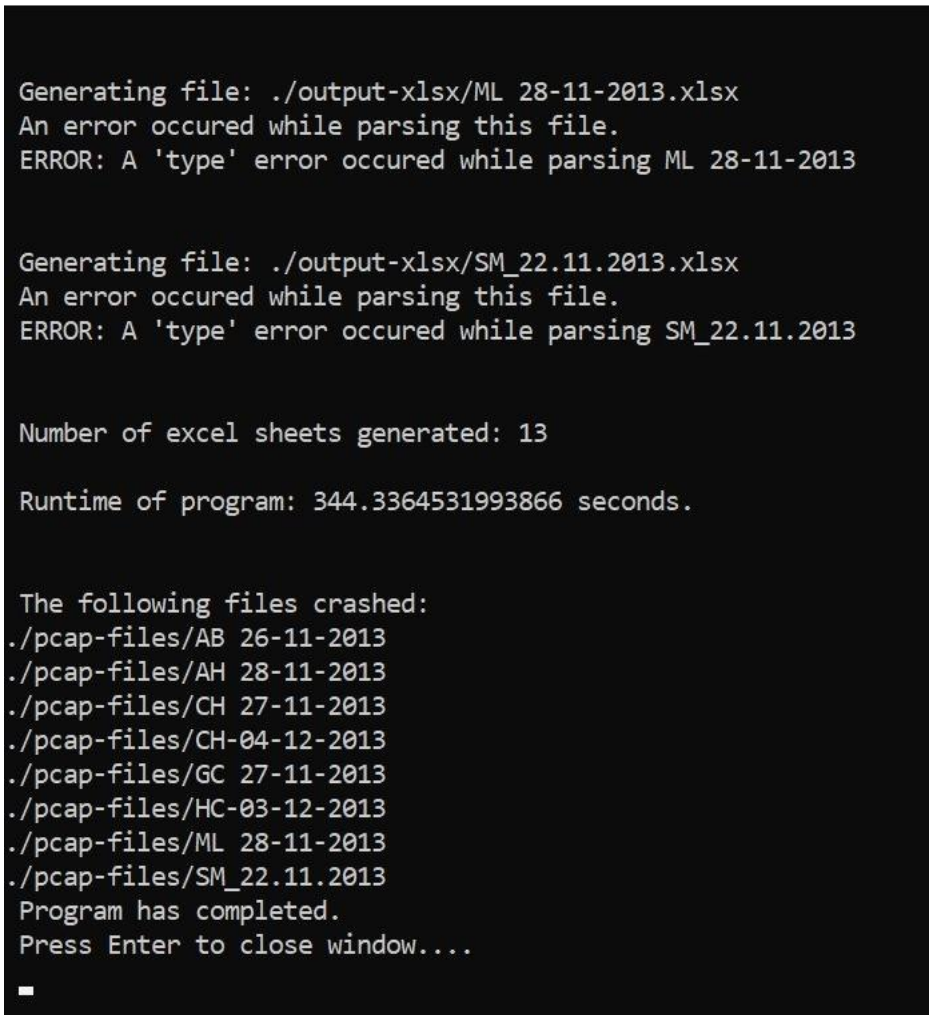

Figure 5. Workflow for raw PCAP file and experimental results for Unique I.P. addresses with Time complexity.

Based on this analysis, we make one more IPYNB file (Interactive Python Notebook) known as the Jupyter notebook. Jupyter is a free, open-source, interactive web tool known as a computational notebook. Researchers can 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 scheme, containing an ordered list of input/output cells which can have code, text (using MARKDOWN), mathematics, plots, and rich media, usually ending with the IPYNB extension [31–33]. This file consists of an algorithm performing data clustering of Unique I.P.s found in the Excel sheet of the previous phase. The data clusters of I.P.s form based on the number of I.P.s facing a particular type of attack. This specific type of attack is determined based on the number of factors identified during the analysis. The IPYNB file is collecting all the unique I.P.s as input and extracting the information like on which layer they are operating, what type of ports and protocols are compromised when they are attacking the source I.P.s of end-users, and what information they extracted from the particular environment of the V.M.s, etc. Based on the analysis, the model designed the group of all the threat agents into particular categories concerning their attacking behaviours identified during the analysis.

Figures 6–8 show the histogram of the bar chart with the help of the IPYNB algorithm for each Excel sheet generated during the first phase. Note that we have demonstrated the experimental results of only three PCAP files, and similarly, we can show this for the other PCAP file. 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 develops the histograms on the cumulative data of all the files in the folder. For every file in the output Excel sheet, three histograms have been generated, and all these three histograms consist of common data at the y-axis, i.e., the number of unique I.P.s. Figures 6a, 7a and 8a show the protocols being used by the attackers and the number of unique I.P.s using these protocols. Figures 6b, 7b and 8b show the ports on the host targeted and the number of unique I.P.s that targeted them. This histogram highlights the vulnerable ports. Figures 6c, 7c and 8c show the time spent as a function of the number of unique I.P.s. This histogram highlights how much time an attacker will usually spend to attack a host. These histograms for the protocols, ports, and time spent on the network will help evaluate the three main attributes for the threat agents, i.e., motivation, opportunity, and capability. Once we identify the port open during the network access, 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 identify the protocols accessed by the threat agents, evaluate the hacker's potential capability, and level of skills acquired by threat actors.

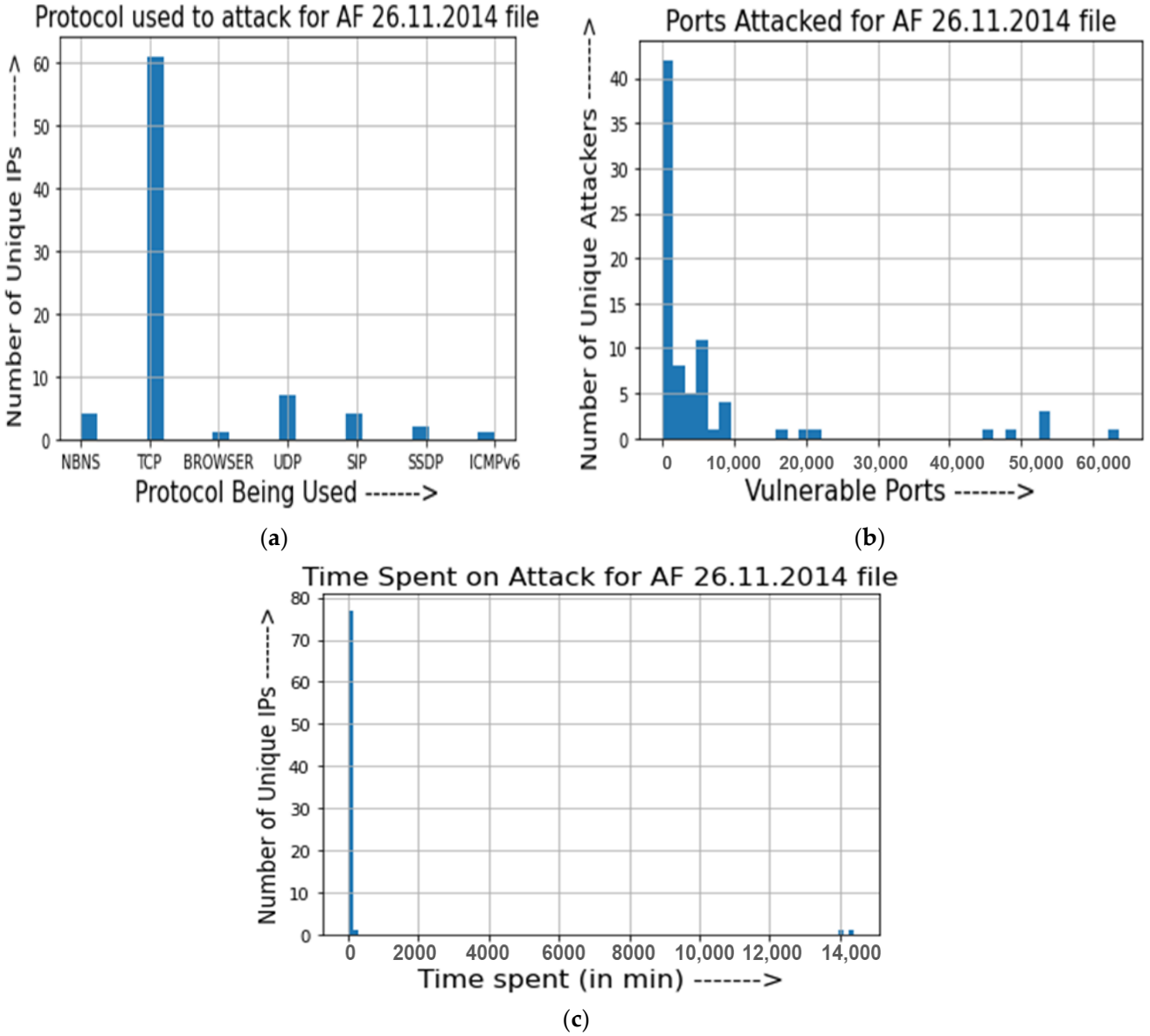

Figure 6. Experimental Results for PCAP file (AF 26.11.2014). (**a**) Number of Unique I.P.s vs. Protocol being used; (**b**) Number of Unique Attackers vs. Vulnerable Ports; (**c**) Number of Unique I.P.s vs. Time Spent.

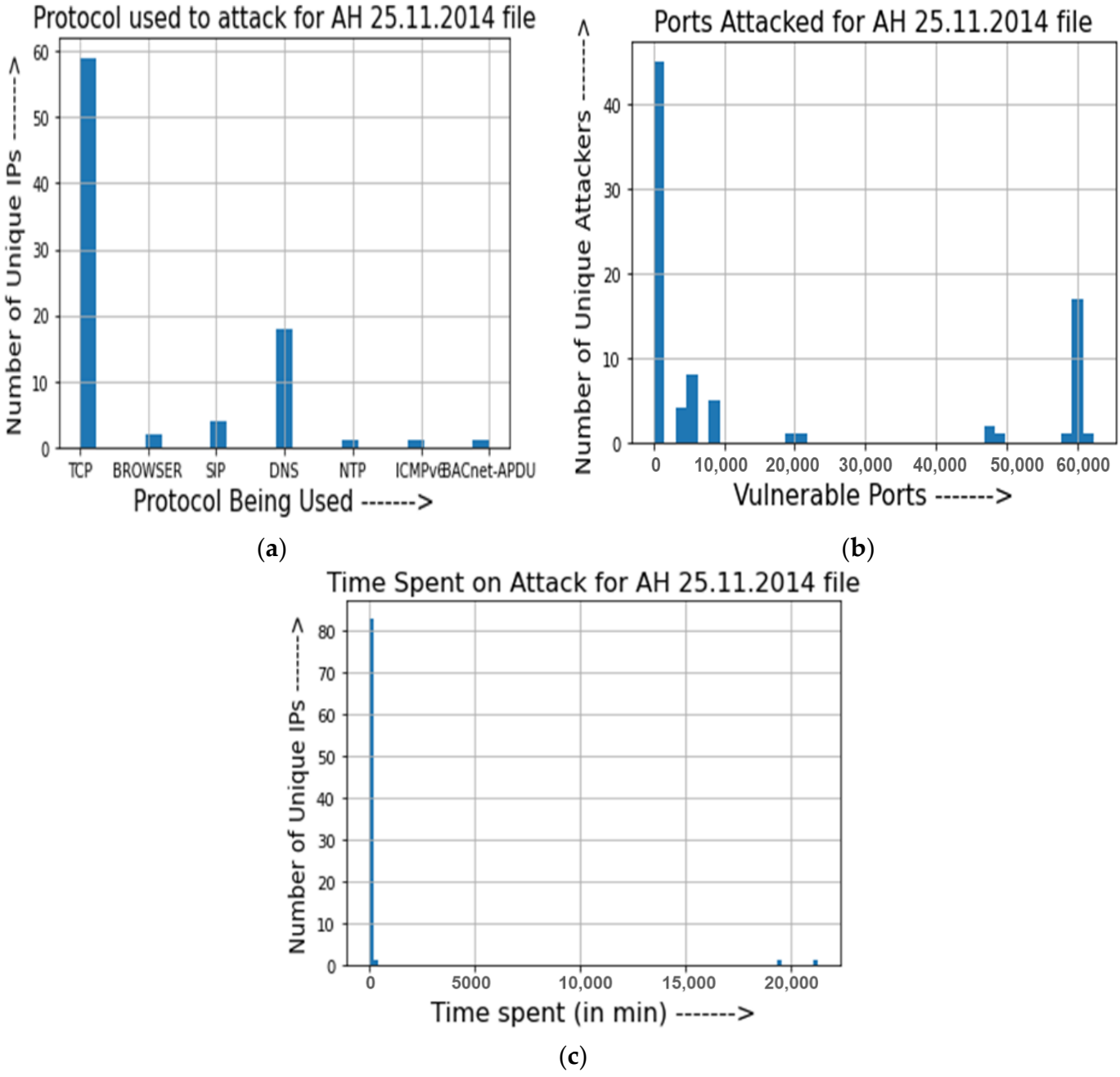

Figure 7. Experimental Results for PCAP file (AH 25.11.2014). (**a**) Number of Unique I.P.s vs. Protocol being used; (**b**) Number of Unique Attackers vs. Vulnerable Ports; (**c**) Number of Unique I.P.s vs. Time Spent.

From this analysis, we can identify the particular groups of threat agents accessing a specific protocol for penetration of the network. For example, in Figure 8, the TCP protocol is used by most of the I.P.s and mainly targets the network layers. So, we can conclude that in this analysis, the threat agents have primarily distributed denial of services (DDOS) type of attacks.

Figure 9 histograms are based on the accumulated data in the potential output produced in the Excel sheets. They are 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. Figure 9a shows how many packets are sent to which port on the host machine, and Figure 9b shows the volume of packets for every Protocol used to attack the host.

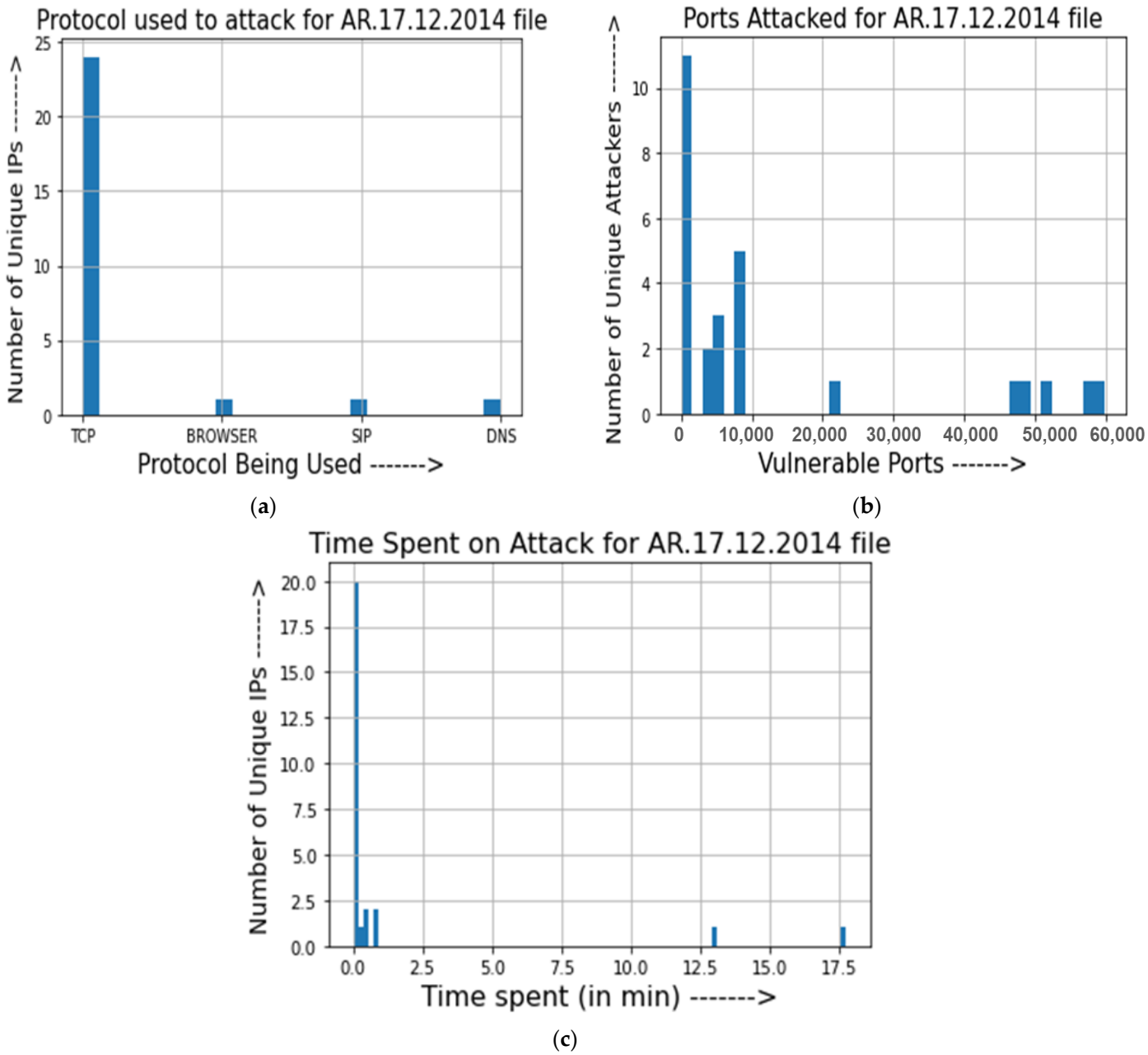

Figure 8. Experimental Results for PCAP file (AR 17.12.2014). (**a**) Number of Unique I.P.s vs. Protocol being used; (**b**) Number of Unique Attackers vs. Vulnerable Ports; (**c**) Number of Unique I.P.s vs. Time Spent.

Figure 10 represents the histogram between the total data collected from each unique I.P., whole time spent on the network, and protocols used to attack the network. Figure 10a highlights the amount spent by the attacker for every Protocol used to attack the host. In Figure 10b, 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.

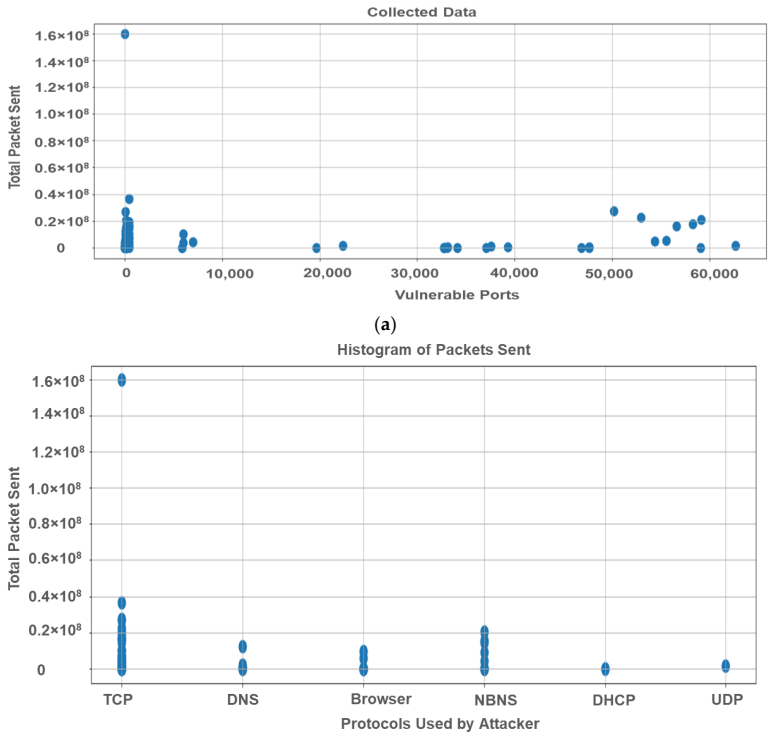

Figure 9. Histogram for (**a**) Total Packets sent vs. Vulnerable Ports, (**b**) Total Packets sent vs. Protocol used by Attackers.

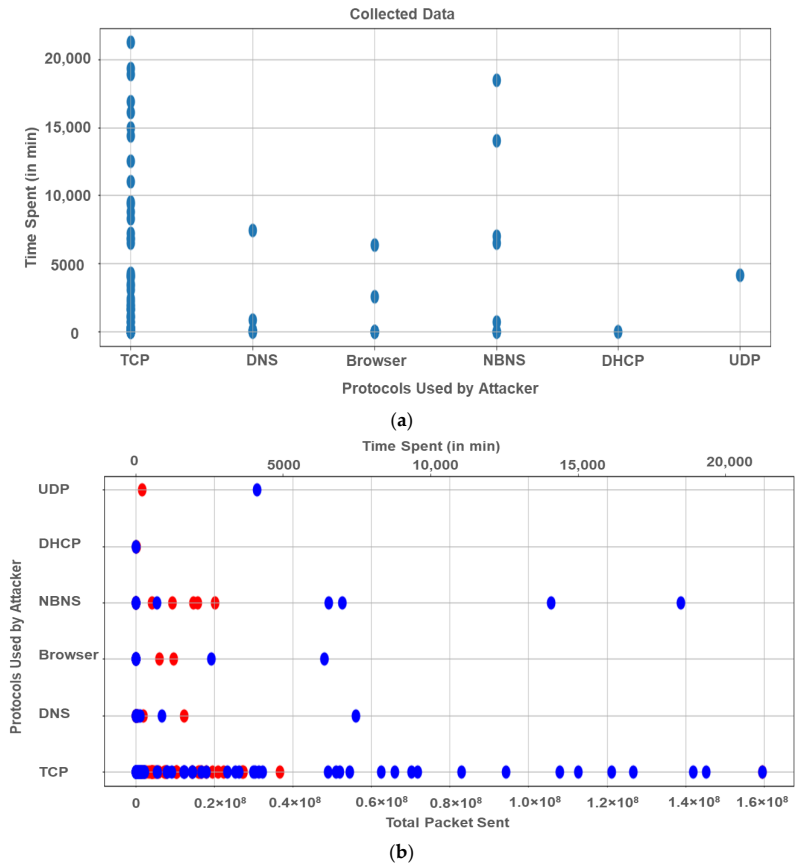

Figure 10. Histogram for (**a**) Time Spent vs. Protocol used by Attackers, (**b**) Protocol used by Attackers vs. Total Packets sent.

5. Conclusions and Future Work

Threats and threat agent's risks are emerging in threat assessment of a network for an organisation and business of the companies. The 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 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 and 2019, which illustrates the value of assets stored on the server, and the motivation, opportunity, and capability of the threat agents while accessing the network. We evaluate the situational awareness data through this semi-automatic threat assessment model by exploring the threat profiles for the historically captured data with the aid tools. Furthermore, we provide the threat agent practitioners with an idea of using an automatic model for threat assessment of a network. This research's findings will support decision makers, management, and software developer practitioners regarding the building of threat agent profiling for their historical data. Critical Threat Intelligence feeds for the threat agent's groups might be helpful for the evaluation of new threats found in the network. In the future, we aim to build an automatic machine learning–based threat and vulnerability analysis security reference model as a security risk management tool to evaluate the security needs of networks with sequential requirements of the next to real-time informational environment.

Author Contributions: Conceptualisation, G.S. and S.V.; methodology, G.S.; software, G.S.; validation, G.S., S.V., and C.M.; formal analysis, G.S.; investigation, G.S.; resources, S.V.; data curation, G.S.; writing—original draft preparation, G.S.; writing—review and editing, G.S., S.V., C.M., N.A., and S.K.; visualisation, G.S.; supervision, S.V., C.M. and N.A.; project administration, S.V.; funding acquisition, G.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author and are available on GitHub [33].

Acknowledgments: We are grateful for the anonymous reviewers' hard work and comments that allowed us to improve the quality of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Iglesias, J.A.; Angelov, P.; Ledezma, A.I.; Sanchis, A. Modelling evolving user behaviours. In Proceedings of the 2009 IEEE Workshop on Evolving and Self-Developing Intelligent Systems, Nashville, TN, USA, 30 March–2 April 2009; pp. 16–23. [CrossRef]
2. Xue, M.; Yuan, C.; Wu, H.; Zhang, Y.; Liu, W. Machine Learning Security: Threats, Countermeasures, and Evaluations. *IEEE Access* **2020**, *8*, 74720–74742. [CrossRef]
3. Jones, A. *Identification of a Method for the Calculation of the Capability of Threat Agents in an Information Environment*; School of Computing, University of Glamorgan: Pontypridd, UK, 2002; pp. 1–134.
4. Mavroeidis, V.; Bromander, S. Cyber Threat Intelligence Model: An Evaluation of Taxonomies, Sharing Standards, and Ontologies within Cyber Threat Intelligence. In Proceedings of the 2017 European Intelligence and Security Informatics Conference (EISIC), Athens, Greece, 11–13 September 2017; pp. 91–98.
5. Atote, B.S.; Saini, T.S.; Bedekar, M.; Zahoor, S. Inferring emotional state of a user by user profiling. In Proceedings of the 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), Greater Noida, India, 14–17 December 2016; pp. 530–535.
6. Asgari, H.; Haines, S.; Rysavy, O. Identification of Threats and Security Risk Assessments for Recursive Internet Architecture. *IEEE Syst. J.* **2017**, *12*, 2437–2448. [CrossRef]
7. Azaria, A.; Richardson, A.; Kraus, S.; Subrahmanian, V.S. Behavioral Analysis of Insider Threat: A Survey and Bootstrapped Prediction in Imbalanced Data. *IEEE Trans. Comput. Soc. Syst.* **2014**, *1*, 135–155. [CrossRef]
8. Vidalis, S.; Jones, A.; Blyth, A.; Jones, A. Assessing cyber-threats in the information environment. *Netw. Secur.* **2004**, *2004*, 10–16. [CrossRef]
9. Cappelli, D.M.; Moore, A.P.; Trzeciak, R.F. *The CERT Guide to Insider Threats: How to Prevent, Detect, and Respond to Information Technology Crimes (Theft, Sabotage, Fraud)*; Addison-Wesley: Boston, MA, USA, 2012.

10. Susukailo, V.; Opirskyy, I.; Vasylyshyn, S. Analysis of the attack vectors used by threat actors during the pandemic. In Proceedings of the 2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT), Zbarazh, Ukraine, 23–26 September 2020; Volume 2, pp. 261–264.
11. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [CrossRef]
12. Legg, P.A.; Moffat, N.; Nurse, J.R.; Happa, J.; Agrafiotis, I.; Goldsmith, M.; Creese, S. Towards a conceptual model and reasoning structure for insider threat detection. *J. Wirel. Mob. Netw. Ubiquitous Comput. Dependable Appl.* **2013**, *4*, 20–37.
13. Bishop, M.; Conboy, H.M.; Phan, H.; Simidchieva, B.I.; Avrunin, G.S.; Clarke, L.A.; Osterweil, L.J.; Peisert, S. Insider Threat Identification by Process Analysis. In *2014 IEEE Security and Privacy Workshops*; IEEE: Piscataway, NJ, USA, 2014; pp. 251–264.
14. Morakis, E.; Vidalis, S.; Blyth, A. Measuring vulnerabilities and their exploitation cycle. *Inf. Secur. Tech. Rep.* **2003**, *8*, 45–55. [CrossRef]
15. Vidalis, S.; Jones, A. Threat Agents: What InfoSec officers need to know. *Mediterr. J. Comput. Secur.* **2006**, *1*, 1–12.
16. Sogbesan, A.; Ibidapo, A.; Zavarsky, P.; Ruhl, R.; Lindskog, D. Collusion threat profile analysis: Review and analysis of MERIT model. In Proceedings of the World Congress on Internet Security (WorldCIS-2012), Guelph, ON, Canada, 10–12 June 2012; pp. 212–217.
17. Casillo, M.; Coppola, S.; De Santo, M.; Pascale, F.; Santonicola, E. Embedded intrusion detection system for detecting attacks over CAN-BUS. In Proceedings of the 2019 4th International Conference on System Reliability and Safety (ICSRS), Rome, Italy, 20–22 November 2019; pp. 136–141.
18. Lombardi, M.; Pascale, F.; Santaniello, D. EIDS: Embedded Intrusion Detection System using Machine Learning to Detect Attack over the CAN-BUS. In Proceedings of the 30th European Safety and Reliability Conference and 15th Probabilistic Safety Assessment and Management Conference, Venice, Italy, 21–26 June 2020; p. 2028.
19. Erola, A.; Agrafiotis, I.; Happa, J.; Goldsmith, M.; Creese, S.; Legg, P. RicherPicture: Semi-automated cyber defence using context-aware data analytics. In Proceedings of the 2017 International Conference On Cyber Situational Awareness, Data Analytics and Assessment (Cyber SA), London, UK, 19–20 June 2017; pp. 1–8.
20. Deore, U.D.; Waghmare, V. Cybersecurity automation for controlling distributed data. In Proceedings of the 2016 International Conference on Information Communication and Embedded Systems (ICICES), Chennai, India, 25–26 February 2016; pp. 1–4.
21. Legg, P.A.; Buckley, O.; Goldsmith, M.; Creese, S. Automated Insider Threat Detection System Using User and Role-Based Profile Assessment. *IEEE Syst. J.* **2015**, *11*, 503–512. [CrossRef]
22. Pogrebna, G.; Skilton, M. The Twelve Principles of Safe Places. In *Navigating New Cyber Risks*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 171–197.
23. Iskandar, A.; Virma, E.; Ahmar, A.S. Implementing DMZ in Improving Network Security of Web Testing in STMIK AKBA. *Int. J. Eng. Technol.* **2018**, *7*, 99–104. [CrossRef]
24. Vidalis, S.; Jones, A. Analysing Threat Agents and Their Attributes. In Proceedings of the 4th European Conference on Information Warfare and Security 2005 (ECIW 2005), Glamorgan, UK, 11–12 July 2005.
25. Rubini, R.; Porta, A.; Baselli, G.; Cerutti, S.; Paro, M. Power spectrum analysis of cardiovascular variability monitored by telemetry in conscious unrestrained rats. *J. Auton. Nerv. Syst.* **1993**, *45*, 181–190. [CrossRef]
26. Shin, B.; Lowry, P.B. 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. Secur.* **2020**, *92*, 101761. [CrossRef]
27. Chen, R.-C.; Cheng, K.-F.; Chen, Y.-H.; Hsieh, C.-F. Using Rough Set and Support Vector Machine for Network Intrusion Detection System. In Proceedings of the 2009 First Asian Conference on Intelligent Information and Database Systems, Dong Hoi, Vietnam, 1–3 April 2009; pp. 465–470.
28. Available online: https://github.com/Gauravsbin/Excell-sheets-of-pcap-files-and-results-of-ThreatAssessment-analysis (accessed on 12 June 2021).
29. Rynes, A.; Bjornard, T. *Intent, Capability, and Opportunity: A Holistic Approach to Addressing Proliferation as a Risk Management Issue*; Idaho National Laboratory (INL): Idaho Falls, ID, USA, 2011.
30. Rossebo, J.E.; Fransen, F.; Luiijf, E. Including threat actor capability and motivation in risk assessment for Smart GRIDs. In Proceedings of the 2016 Joint Workshop on Cyber-Physical Security and Resilience in Smart Grids (CPSR-SG), Vienna, Austria, 12 April 2016; pp. 1–7.
31. Saygili, G.; Rathje, E.M.; Wang, Y.; El-Kishky, M. Cloud-Based Tools for the Probabilistic Assessment of the Seismic Performance of Slopes. In *Geotechnical Earthquake Engineering and Soil Dynamics V*; American Society of Civil Engineers (ASCE): Houston, TX, USA, 2018; pp. 19–26.
32. Van Veen, H.; Saul, N.; Eargle, D.; Mangham, S. Kepler Mapper: A flexible Python implementation of the Mapper algorithm. *J. Open Source Softw.* **2019**, *4*, 1315. [CrossRef]
33. Narkar, S.; Thomson, B.L.; Fox, P.A. Designing for 2030: The Impact and Potential of Virtual Laboratories. In Proceedings of the American Geophysical Union, Fall Meeting 2020, 1–17 December 2020.