Article

A New Proposal on the Advanced Persistent Threat: A Survey

Santiago Quintero-Bonilla ^{1,*} and Angel Martín del Rey ² ¹ Department of Applied Mathematics, University of Salamanca, 37008 Salamanca, Spain² Institute of Fundamental Physics and Mathematics, Department of Applied Mathematics, University of Salamanca, 37008 Salamanca, Spain; delrey@usal.es

* Correspondence: santiago.quintero@usal.es

Received: 11 May 2020; Accepted: 29 May 2020; Published: 3 June 2020

Abstract: An advanced persistent threat (APT) can be defined as a targeted and very sophisticated cyber attack. IT administrators need tools that allow for the early detection of these attacks. Several approaches have been proposed to provide solutions to this problem based on the attack life cycle. Recently, machine learning techniques have been implemented in these approaches to improve the problem of detection. This paper aims to propose a new approach to APT detection, using machine learning techniques, and is based on the life cycle of an APT attack. The proposed model is organised into two passive stages and three active stages to adapt the mitigation techniques based on machine learning.

Keywords: advanced persistent threats; machine learning; malware detection; cybersecurity

1. Introduction

Cybersecurity is responsible for establishing security policies; these policies set out the steps to follow for data to be managed within the technological infrastructure in an organisation. However, some security flaws and vulnerabilities (e.g., the use of outdated equipment, use of policies that are not reviewed continuously, failing to install updates at time, awareness deficiency) allow attackers to realise an intrusion in an organisation.

The increasing development of sophisticated tools used by cybercriminals, such as zero-day vulnerabilities and denial of service (DoS) attacks, conventional solutions cannot cope with the current complexity of these types of threats.

Nowadays, advanced persistent threat attacks represent a real threat to public and private entities around the world and will continue to do so in the future [1]. These attacks are an imminent threat, whose main problem is the difficulty of early detection because attackers use different techniques, both to stay as long as possible undetected, and to evade efficiently.

The differences between an advanced persistent threat (APT) and a common cyber-attack are significant. For example, the number of resources of all kinds necessary to carry out the attack. A common cyber-attack can be directed to entities or organisations with null or deficient cybersecurity policies in order to steal customer or financial activity data of a company [2]. These attacks are usually detected, and the damage caused is not usually critical. However, an APT can be focused on large organisations and industry sectors, causing severe damage, e.g., intellectual property theft, failure of essential services, and destruction of critical infrastructure. These attacks are usually undetected, and the damage caused can be critical.

In recent years, the number of reported APT related cases has increased significantly [3,4]; one of the main objectives of APT attackers is to remain undetected. Some researchers have proposed different approaches to understand and detect this type of threat. They can be observed that the

life cycle of this attack is an indicator to understand how these attacks work [5–7]. Furthermore, the machine learning techniques have allowed for the collection and analysis of tools used by attackers to improve the early detection of these attacks.

An example of the scope that APT can have is that actors take advantage of current issues that generate interest in the general population. The actual situation of the COVID-19 has generated the right scenario for actors to launch their attacks. In this case, the lure has been advisory information about healthcare situation in different countries, for these techniques such as spear-phishing, exploits with remote access tools and ransomware have been used [8].

The aim of this paper is to propose a new approach to APT detection using machine learning techniques, and based on the life cycle of an APT attack. The proposed new five-stage model more precisely matches the passive and active components of an APT attack.

The rest of paper is organised as follows, Section 2 gives an overview of APT attacks process. In Section 3, the techniques and algorithms used for APT detection are described. In Section 4, the APT life cycles are analysed. In Section 5, the discussion is presented, and finally, in Section 6 the conclusions are shown.

2. Advanced Persistent Threat

An advanced persistent threat is a selective attack that obtains unauthorised access to information and communication systems in order to filter confidential data or cause damage to a company, industry, or government organisation [2,9]. Since the appearance of Stuxnet [10], APT attacks are more cautious and damaging, showing the ease of intrusion into high-profile systems, evading many of the more sophisticated defence tools used to protect the computing environment. Currently, many of these threats remain undetected. Many of these threats, once detected, reappear with modifications to achieve their objective; for example, FIN6 [11], APT10 [12], APT41 [13] have been attacks that have caused significant losses of money, confidential information, and intellectual property.

Next, in Section 2.1 the characteristics of APT are defined, in Section 2.2 the attack process is presented, in Section 2.3 the methods and techniques are described, and finally, in Section 2.4 the attribution problems are listed.

2.1. Characteristics of Advanced Persistent Threats

In 2006, United States Air Force (USAF) analysts coined the term “advanced persistent threat” to facilitate discussion of intrusive activities [9] with their uncleared civilian counterparts. In this sense, military teams could discuss the characteristics of the attack without revealing classified identities. The components of the terminology defined by the USAF are explained below:

- Advanced: the enemy is familiar with the tools and techniques of intrusion, able to develop custom exploits.
- Persistent: the enemy intends to fulfil a purpose, receive orders, and attack specific goals.
- Threat: the enemy is coordinated, supported, and motivated.

APT attackers have purposes and objectives that differ from common computer criminals due to their targeted nature. For example, espionage in different sectors, such as, industrial, military, economic, technical and intellectual property, financial extortion, and political manipulation. The authors in [2] have summarised the differences between traditional threats and APT attacks. For this, the characteristics that have been considered are attacker, target, purpose, and approach (see Table 1).

Table 1. Differences between an advanced persistent threat (APT) attack and common malware attacks [2].

Feature	APT Attacks	Common Malware Attacks
Definition	APT is a sophisticated, targeted and highly organised attacks. (e.g., Stuxnet)	Malware is malicious software used to attack and disable any system. (e.g., ransomware)
Attacker	Government actors and organised criminal groups	A cracker (a hacker in illegal activities)
Target	Diplomatic organisations, information technology industry and others sectors	Any personal or business computer
Purpose	Filter confidential data or cause damage to a specific target	Personal recognition
Attack life cycle	Maintain persistence as possible using different ways	It ends when it is detected by the security actions (e.g., anti-virus software)

2.2. APT Attack Process

Different approaches can characterise an APT; each APT campaign acts differently, and attacks are customised for specific victim or organisation. Generally, the first step is to create a point to gain access to the network. Then, personalised malware creates a communication network to maintain access, that allows attackers to inject malicious code multiple times. This malware moves sideways through the system (in a stealthy way), detecting vulnerabilities it can exploit and infecting other hosts on the network. It also makes copies of itself to maintain persistence within the system. The APT malware can establish other outgoing connections as they gain access to the system and obtain as much data as possible.

One example life cycle approach has been described in the FireEye research about APT1; In the APT1 analysis, Mandiant (actually FireEye) has presented a report with an overview of the life cycle model of an ATP attack, consisting of eight stages: (1) Initial Recon, (2) Initial Compromise, (3) Establish Foothold, (4) Escalate Privileges, (5) Internal Recon, (6) Move Laterally, (7) Maintain Presence, and (8) Complete Mission. The stages between “Establish Foothold” and “Complete Mission” do not have to occur in this order every time [14]. This report is best known for identifying and understanding these types of threats.

As some APT campaigns are discovered, it is observed that its anatomy is diverse and changes according to the specific objective for which it has been designed. The diversification of the attack vectors makes the detection of these threats a complicated task. Consequently, in Section 5, the different existing approaches of APT attack process are described. Then, a new model to improve early detection is proposed, based on the life cycle of an APT attack and applying machine learning techniques.

2.3. Methods and Techniques

APTs use a variety of methods and techniques. The attack process begins with a study of the victim; in many cases, spear-phishing or emails are used together with social engineering, with the aim of the victim downloading an infected file. Then, the attacker compromises the computer and gains access to other computers within the organisation through the network.

The methods that characterise the most “advanced” APT groups are the use of zero-day exploits and unknown previously unidentified infection vectors. This methods engages several government organisations in several countries to successfully steal confidential information for a long time without first being discovered.

The techniques commonly used to carry out an APT attack are adapted or combined depending on the target. Some examples of these techniques are the following:

- Social engineering: Getting a user to compromise information systems. This technique is directed to people with privileged access, manipulating them to divulge personal information to carry out a malicious attack through control and persuasion, instead of involved aleatory attacks on systems [15].
- Spear-phishing: This technique is an attempt that primarily targets a specific organisation in order to collect user credentials, financial information, or other confidential information [16].
- Watering hole: It is similar to spear-phishing in cyberespionage. The attacks are adapted to the needs of the victims. In order to do this, attackers try to obtain information about the victim considering his/her personal interests [17].
- Drive-by-download: this technique performs the unintentional download and execution of malicious software when a malicious web page is visited [18]. The malware is downloaded “stealthily” without the knowledge of users, taking advantage of security breaches, browser exploits, or integrated plugins such as ActiveX, Java/JavaScript, or Adobe Flash player [19].

2.4. Attribution Problem

The attribution of a cyberattack or a particular campaign to an actor has become a problem. This problem is more complicated when it is trying to correlate an APT to a particular group or state. Experts can observe different evidence to identify the attackers when analysing these types of threats, such as IP addresses, e-mails, or the malicious code used. These attackers often use the concept of a false flag, which consists of posing as a third party to camouflage their operations. In recent years, attacks attributed to government actors and organised groups has shown a significant increase.

The main actors can be divided into two large groups: government actors and organised criminal groups. These APT actors will be briefly described in the following.

2.4.1. Actors

The cyberattacks carried out by governments and nation-states are becoming more frequent. The suspicions of interference in elections or interruption of electricity supply in other countries are generating widespread public concern due to the high cybernetic capabilities of these actors.

- China: Chinese cyberattacks have been observed as being focused on industrial espionage, and aimed to steal intellectual property. APT1 have been the most persistent cyberthreat of this actor [14].
- United States: This actor could have perpetrated the most sophisticated cyberattacks. Attacks have been harmful, and advanced technologies have used, which means considerable resources for the development of this type of attack. The APT campaigns have mainly been used to enforce geopolitical interests. One example is the world-famous operation Stuxnet [10], which targeted SCADA (Supervisory Control and Data Acquisition) systems to cause substantial damage in the nuclear program of Iran.
- Russia: This actor is very active in terms of state-sponsored APT activity. These groups have been involved in high profile intrusions and because of this has been the subject of intense investigations [4]. Recently, spear-phishing attacks from APT28 have been detected by Microsoft; their targets have been the employees of the German government. This group has attempted to gain access to employee credentials and infect sites with malware [20].
- Iran: In the Middle East, this actor controls the most attack capacity attributed to the country with several incidents perpetrated by diverse groups [4]. Experts have monitored APT33 operations because this group has recently upgraded its infrastructure. The main objectives of this group have been the aviation industry and energy companies with connections to petrochemical production. The latest malware campaigns have been targeted at organisations in the United States, the Middle East, and Asia [21].
- North Korea: The cyber groups associated with this actor have conducted numerous operations, including conventional espionage, banking hacks, and destructive attacks [3]. One example employed by this actor is the ransomware WannaCry [22].

- Israel: This actor has been identified as a possible co-author of the Stuxnet [10] attack. It is publicly known as the high potential of the intelligence services of this country, one example is Unit 8200 [23] of the Israeli army, the equivalent to the US intelligence agency NSA. The Duqu 2.0 [24] attack was state-sponsored by this actor and it has infected numerous systems in several countries in recent years. This malware used zero-day vulnerabilities, and for sending data to the command and control (C&C) servers; different techniques were used to take the computers on the network.

2.4.2. Campaigns

Campaigns are the actions, methods, and customised techniques that an actor performs against a target to execute an APT in order to extract highly sensitive data, for example, zero-day malware, social engineering, and data extraction through C&C servers. In addition to the actors mentioned above, there are groups of cybercriminals organised with private funding and not responding to government interests; these groups have carried out different campaigns. In recent years, new APT campaigns have been discovered; these campaigns (see Table 2) are mostly still active, and the number of targets affected is unknown. They use different propagation methods, e.g., exploits, infected files, and custom malware. These campaigns are designed for cyberespionage, and the main targets are diplomatic organisations and the information technology industry.

The investigation of these campaigns was carried out by Kaspersky, using a 15-step methodology, in which samples of malware, generated traffic, and communication protocols used by the attackers in an incident were dissected; this incident could be classified as APT [25].

Table 2. Latest APT campaigns discovered (adapted from [25]).

Discovery Date	First Known Sample	Name	State	Targeted Platform
2019	2019	Topinambour	Active	Windows
2019	2013	TajMahal	Active	Windows
2018	2018	ShadowHammer	Inactive	Windows
2018	2018	FruitArmor	Active	Windows

3. Machine Learning

Machine learning (ML for short) is a sub-field of artificial intelligence (AI for short) that gives the computational process of automatically inferring and generalising a learning model from sample data. ML studies algorithms and techniques to automate solutions to complex problems that are difficult to program using conventional programming methods. ML models use mathematical and statistical functions and techniques to describe data dependencies, causalities, and correlations between input and output data.

The multiple utilities of ML serve to handle day-to-day problems and support decision-makers by bringing together researchers from different areas of knowledge. Some problems that ML can solve are the following: facial recognition, detection of false news, sentiment analysis, recommendation systems, fraud detection systems, language translation, and chatbots.

Next, in Section 3.1, the ML techniques and algorithms commonly used in cybersecurity are described, in Section 3.2, the applications of ML in cybersecurity used in APT detection are detailed and, in Section 3.3, the approaches used for APT detection are analysed.

3.1. Techniques and Algorithms

The concept of labelled and unlabelled data is necessary to be introduced before describing the ML models. When the correct answer to a data-related question is known, labelled data is obtained; however, when the correct answer is unknown, one has unlabelled data.

ML algorithms derive their power from the ability to learn from available data. The main ML models can be classified into supervised learning and unsupervised learning (see Figure 1).

Figure 1. Machine learning algorithms.

3.1.1. Supervised Learning

The goal of supervised ML is to build a model that create evidence-based predictions in the presence of uncertainty. These algorithms take a known data set (input) and known responses to the data (output), then train the model and generate analytical predictions in response to the new data. An example of this algorithm is used in weather forecasting.

Supervised learning uses classification and regression techniques to develop predictive models. The most popular supervised machine learning methods are artificial neural networks, support vector machine, decision trees, bayesian networks, *k*-nearest neighbour, and hidden Markov models [26]. These algorithms are explained below:

- Artificial neural networks (ANN) are computational brain-inspired models and interlinked by a lot of interconnections (artificial synapses) of artificial neurons (nodes) capable of performing specific calculations at their inputs [27]. An artificial neuron is composed of three or more layers, an input layer, one or more hidden layers, and an output layer. An ANN is capable of creating non-linear models to obtain the relationships between input attributes and label classification [28]. The main characteristics of ANN are adapting from experience, learning capability, generalisation capability, data organisation, fault tolerance, distributed storage, and facilitated prototyping [29]. These algorithms are useful for speech and pattern recognition [30], climate forecasting [31], and disease diagnosis [32]; although this model also solves classification and regression problems.
- Support vector machine (SVM) is one of the most accurate and robust methods of ML algorithms. This classifier works by identifying a hyperplane between two classes of labelled data in a set of training data. The SVM classifier uses several types of methods, e.g., non-linearity and use of kernels, separability, and margins or risk minimisation. Non-linearity and kernel usage are some of the pioneering discoveries in the field of ML; this method permits that a non-linear problem can be transformed into a linear problem. Several types of separating hyperplanes can be realized using a kernel, such as radial basis function (RBF), polynomial, linear, or sigmoid. Risk minimisation can be applied to cases that do not fit into the traditional SVM architecture, such as problems with missing data or unlabelled data [33–35].
- Decision tree (DT) models are accurate, stable, and straightforward to interpret. Their construction is based on decision rules that are represented in the form of a tree. The result of these models can represent non-linear relationships for problem-solving. Decision trees and random forests are the most remarkable because they are more precise and elaborate. Their predictive capacity

is higher because of these characteristics, but their performance is low. Most commonly used algorithms for building decision trees are CART (Classification and Regression Tree), ID3 (Iterative Dichotomiser), and CHAID (Chi-Squared Automatic Interaction Detector) [33,34,36].

- Bayesian networks (BN) are probabilistic graphical models used to describe and analyse multivariate distributions. These variables can be continuous or discrete, however, when all variables are discrete, the notation is represented as a series of sums and products. In the graphic representation of a BN, the nodes represent an observable variable or state, and the edges symbolise the conditional dependencies between the nodes. BN has been used in different areas, for example, Microsoft Windows System, NASA mission control, and bioinformatics applications [34,37,38].
- k -nearest neighbour (k -NN) can be used for both regression and classification problems. Due to simplicity, effectiveness, and intuitiveness of the concept, this model can be used to identify the nearest neighbours for a given data point based on a distance measure [39,40]. The assumption is that similar elements are closer together. The idea of closeness is a measure of distance, which can be a simple Euclidean distance between two points. In this case, the classification decision may be influenced by the sensitivity of k , especially in small data sets with outliers. Numerous families of distance measurements exist, and the following can be highlighted: Minkowski, Inner product, Square Chord, Shannon entropy, and Vicissitude [41].
- Hidden Markov model (HMM) is a stochastic probabilistic model of discrete events and a variation of the Markov chain, a chain of linked states or events, where the next state depends only on the current state of the system. HMM is used to analyse features or observations to predict the most likely state sequence; these hidden states represent an unobserved attribute of the process. HMM have been used to solve problems of financial analysis, genetic sequencing, image processing, and natural language processing [34,42].

3.1.2. Unsupervised Learning

Unsupervised learning does not have a training dataset. Some unlabelled data are presented, and the model itself must learn from them, and then predict future results [43]. This type of learning model is the most appropriate when the problem requires a large amount of data that is unlabelled. Unsupervised learning aims to find hidden patterns or specific structures in the data. It is used to extract inferences from datasets consisting of input data without labelled responses.

This learning model uses dimensionality reduction (e.g., principal component analysis or PCA) and clustering techniques (e.g., k -means, Fuzzy c -means, and hierarchical) to develop predictive models. An example of the application of unsupervised ML model is the detection and classification of unwanted mail or spam. These algorithms are explained below:

- Principal component analysis (PCA) is a procedure of dimension reduction. This statistical method is useful when there are a large number of variables, where each variable has more or less importance. PCA generates a score matrix T called a score matrix where the correlation between variables is displayed in a maximum of two or three dimensions. This procedure is used to assign a set of interrelated variables to a smaller set of non-linearly correlated variables while representing as much variance as possible in the original data set [44]. Some examples of applications of this method are feature extraction [45], social science, medicine, and genome [46].
- k -means is a clustering algorithm. This technique consists of selecting the input data into k clusters for a predefined k group. Each data point in the input set is unlabelled data. The interpretation for each of the k groups is that the mean value of the group is representative of all elements in that group. Alternatively, each k groups could represent a type of input data. The user defines the number k of clusters. This algorithm uses computational distances to find the distance between two points, for example, the Euclidean distance. Also k -means can be used in intrusion detection systems (IDS) [28].

- Fuzzy c-means is a soft clustering algorithm. This method randomly selects the number of clusters; then, each data point is assigned a cluster membership. This process is continuously reviewed to minimise the distance and degree of cluster membership [47].
- Hierarchical clustering is used to cluster data points when the data is unlabelled. This method can be classified into two categories: divisive and agglomerative. In the divisive approach, the data points are considered as one large cluster and then divided into smaller clusters. In the agglomerative approach, each data point is considered as an individual element, and then it is added to a cluster [48].

3.2. Role of Machine Learning in Cybersecurity Applications for APT Detection

Nowadays, massive and targeted attacks are more frequent. These attacks can cause damage to users or organisations such as the loss of sensitive information. Researchers are studying different approaches to prevent or minimise the risk of attacks. Some of the methods and techniques that researchers have used are directly related to machine learning.

The prevention measures require higher capacity for analysis and response in the shortest possible time, due to the large volume of data and the rapid evolution of current threats. For this reason, automated tools have been created to assist cybersecurity administrators. Machine learning techniques are a useful tool in the field of cybersecurity. For example, models of network traffic behaviour can be created to detect anomalous activity, reduce the number of false positives on alarms, and detect threats in real-time [49]. However, machine learning can be used to create attacks, for example, on sending fraudulent emails or password cracking software [50]. Machine learning applications in cybersecurity can be classified as follows [51]:

- Detection: These are the tools that allow the detection of abnormal behaviour to generate alerts in real-time, and to facilitate decision-making.
- Protection: Detect vulnerabilities to install security fixes automatically.
- Prediction: Techniques and algorithms to predict attacks and develop anti-malware techniques.
- Termination: Automatically eliminating the threat.

Machine learning techniques applied to cybersecurity can help system administrators find unusual behaviour in the network an organisation, for example, an APT.

Some keys approaches to detecting APT are: (1) Observe unusual alerting patterns to detect malware with malicious load recognition, known components, and remote control activities. (2) Monitoring suspicious outbound traffic on the network can display significant parameters such as infected computers, C&C centres, and data exfiltration. (3) Monitoring unexpected internal traffic on the network could reveal escalating privileges, lateral movements, and malware propagation. Some of these cybersecurity applications using ML techniques are described below.

Spam and Phishing Detection

Spam is mail that has not been solicited. Usually, they come from unknown senders for advertising or commercial purposes, so it is essential to distinguish them from legitimate emails. Phishing is one of the most widely used attack vectors, where an entry point is established between the attacker and the network of a company. Social engineering is used to trick the victim into visiting a fraudulent site to steal credentials. The detection of phishing is becoming increasingly difficult due to the advanced evasion strategies used by attackers, such as open redirects to avoid spam filters [52,53]. For this purpose, different ML classification techniques can help to detect spam. The classification between an authentic mail and a fraudulent one is necessary to distinguish different criteria, allowing the algorithm used to learn to identify any email, between the training dataset. The authors in [54] proposed a scoring technique to detect lateral spear-phishing emails using a combination of various features. A practical, deployable, and real-time detection system for such attacks has been created.

Malware Detection

Modern malware creates executable files that can cause damage to systems on a network or steal information without users' permission. Usually, the malware uses communication to a C&C server through randomly generated IP or URL addresses. For this reason, creating blacklists is an inefficient method. In this way, machine learning algorithms have been used to detect malicious communication addresses. Some studies proposed by researchers for the detection of malware with machine learning techniques are discussed in [35]. The authors in [55] have presented a novel proposal to detect C&C channels used in APT attacks. This process consists of observing specific communication patterns within web browsing in order to identify and detect the malware used in these attacks. Another approach to malware detection has been detailed in [56]. The objective of this work is the detection of malware based on the analysis of DNS traffic and malicious traffic through traffic monitoring at the egress point of the network.

Intrusion Detection

This method allows the monitoring of network traffic to analyse data flows for unusual behaviour patterns; for example, intrusion detection systems and intrusion prevention systems are used. This method can be divided into misuse and anomaly detection. Anomaly detection uses techniques modelling the network and identifying abnormal data flow behaviour on the network. Misuse detection uses signature-based techniques (hash) on known attacks to detect possible attacks [57]. In [58], the authors review the machine learning techniques used for these detection methods. The authors in [5] propose the detection of lateral movement based on anomalies in malicious remote desktop protocol (RDP) sessions in the Windows operating systems. In this paper, taking advantage of system event logs, several supervised machine learning techniques were evaluated to classify RDP sessions and detect malicious session entries.

3.3. Approaches Used for APT Detection

The volume of data generated by information systems has increased in recent years. This growth has made malware and network attacks more difficult to detect. However, several approaches have been proposed to solve this problem, such as dynamic analysis [59], context-based [60], independent access [61], contextual information [62], and information flow tracking [63]. These data must be analysed in the shortest possible time to identify an attack. Consequently, researchers have begun to use machine learning techniques to improve the rate of true positives in detecting APT attacks [64]. Some proposed approaches are detailed below.

A novel machine learning-based system called MLAPT was presented in [6]. This model was detected APT attacks through early alerts that are analysed by ML algorithms. These alerts have been created from a correlation framework between several detection modules. MLAPT is based on the analysis of a six-phase APT life cycle: (1) Intelligence Gathering, (2) Point of entry, (3) C&C Communication, (4) Lateral Movement, (5) Asset/Data Discovery, and (6) Data Exfiltration. The MLAPT framework works in three phases:

- Threat detection: The network traffic is scanned by eight detection modules to find techniques used by APT. The output of this phase consists of alerts, known as events.
- Alert correlation: The events generated by detection modules are correlated, and the output can be two types of alerts.
- Attack prediction: A machine learning-based prediction module is used to detect APT techniques.

A novel distributed framework architecture for APT detection (DFA-AD) is described in [65]. This framework classifies events in a distributed environment and correlation between them to detect techniques used by APT. Intrusion detection is realised in a distributed environment on the trusted platform module (TPM). DFA-AD has been designed in three phases:

- Network traffic, traffic flow is collected, processed, and analysed by a method of recognition using machine learning algorithms.
- Correlation event, through specific rules given by an administrator, the events generated in the previous phase are collected to be evaluated.
- Voting service, the previous information are analysed, and the alert is generated if an APT attack is detected.

Fractal-based anomaly classification mechanism is presented in [66]. This method has used k -NN and correlation fractal dimension (FD) as anomaly classification algorithms to test the dataset and comparison of the results. In the first step, two datasets with normal network traffic and APT attack traffic packages were combined. Then, vector characteristics were extracted through the analysis of the TCP (Transmission Control Protocol) session data. Next, the dataset noise has been removed, and the result dataset will be used on the anomaly classification algorithm to detect an attack. Finally, the authors demonstrated that the algorithm based on the Euclidean dimension is less effective than the algorithm based on the fractal dimension, giving better results.

An APT detection system based on the big data architecture process was proposed in [67]. This model used k -NN algorithms with big data about network data, system logs, and security information. This system was divided into four steps:

- APT system architecture: Network data and information system was collected to be analysed.
- Big data processing technology: A Hadoop cluster was used to improve the analysis of an APT attack.
- APT analysis technology: The detection of malicious attack was detected from vulnerabilities and suspicious connections with anomalous behaviour.
- APT detection algorithm: This method used the tool Mahout because it can process big data and k -NN algorithm can be used for the detection. This model was divided into four phases: retrieve, reuse, revise, and retain.

An anomaly-based approach for the detection of malicious RDP (remote desktop protocol) sessions was detailed in [5]. This model proposed RDP sessions as an intrusion method used in the lateral movement phase of the APT life cycle. The host and network logs were used to identify anomalous events that may match a trace of an APT attack. For this purpose, two real datasets were used, which were divided into five different types of logs: authentication, process, flow, DNS, and red team logs. These datasets were evaluated with the following ML techniques: logistic regression, Gaussian-Naive Bayes, decision tree, random forest, and LogitBoost. The authors concluded that the LogitBoost algorithm is the most effective for the detection of anomalies in the RDP sessions.

An attack scenario method over mining IDS security logs to detect APT was proposed in [7]. This method uses the four-phase kill chain intrusion (IKC) model: information collection, intrusion, latent expansion, and information theft. The attack events were classified according to the purpose of each of the phases of the IKC model. These events were then correlated with IDS logs, using fuzzy clustering to form the attack chain. Finally, this model creates scenarios that serve as a guide for the detection and defence of these targeted attacks.

A detection system of APT that permits an early discovery of the attack was detailed in [33]. This model used a dataset where four categories of attacks were identified: DoS, probe, R2L (unauthorised remote machine connection), and U2R (unauthorised access as local user administrative privilege). The correlation of the variables was analysed with PCA; the number of variables was reduced to 94 characteristics. Then, four classification algorithms: SVM, NB, DT, and multilayer perceptron (MLP) were used. The dataset was analysed with different parameters of each algorithm. Results show that the algorithm with the most effective precision was SVM-RBF or MLP-AS ($N = 4$).

In summary, the proposed models use different machine learning methods for malware detection. The most used algorithms were k -NN, SVM, and DT. In Table 3, the approaches and their phases, ML algorithms, detection accuracy, and life cycle that were used in each work can be seen.

Table 3. Comparison of machine learning (ML)-based APT detection approaches.

Authors	Algorithm	Approach	Approach Detail	APT Life Cycle Used	Detection Accuracy
Ghafir et al. [6]	DT, SVM, <i>k</i> -NN and Ensemble learning	MLAPT	Phases: Threat detection Alert correlation Attack prediction	6 phases	81,8%
Sharma et al. [65]	Genetic programming, classification and regression tree, dynamic bayesian game model and SVM.	DFA-AD	Phases: Network traffic Correlation event Voting service	Non-specified	98.5%
Siddiqui et al. [66]	<i>k</i> -NN and Correlation fractal dimension.	Fractal-based anomaly.	Steps: Combined packet capture (pcap files) Feature vector extraction Noise removal Anomaly classification with ML algorithms	Non-specified	93.58% (FD), 92.83% (<i>k</i> -NN)
Shenwen et al. [67]	<i>k</i> -NN	Detection based on Big Data	Phases: retrieve reuse revise retain	Non-specified	Non-specified
Bai et al. [5]	LR, GNB, DT, RF and LB	RDP-based LM detection	Steps: Preprocessing of dataset Defining metrics Apply ML techniques Compare results	1 phase	99.99% (LB)
Chu et al. [33]	PCA, SVM, NB, DT and MLP	Early discovery of APT attack	Steps: Dataset preprocessing Dimension reduction Classifier	Non-specified	97.22% (SVM)
Zhang et al. [7]	Fuzzy clustering	APT attack scenarios	Steps: Data preprocessing Attack event classification Fuzzy clustering Attack scenario mining	IKC model (4 phases)	Non-specified

4. Advanced Persistent Threat Life Cycle Analysis

The life cycle is fundamental to understanding how an APT attack works and identifying the most commonly used malicious techniques; there are several ways in which APT attack campaigns use their resources to stay undetected. In recent years, researchers have proposed life cycles organised in stages. These stages are composed of techniques, methods, and tools used to perform a targeted intrusion. The number of stages of a life cycle varies according to the proposed approach; for example, a life cycle can be organised from three stages [68] to eleven stages [1].

4.1. Three Stage Attack

The authors in [68] proposed a life cycle described in three-stages, based on the analysis of different methods and techniques of 22 APT campaigns. Each stage contemplates at least three characteristics or techniques that are used to carry out the attack. The stages contemplated are:

1. Initial compromise (IC): In this stage, attackers attempt to access to the target system. The most commonly used techniques in this phase are spear-phishing (e.g., attaching an email or a link to a compromised server), watering-hole (malicious code on a regularly visited website), server-side attacks (exploiting vulnerabilities on servers or stealing brute-force credentials), and infected storage media (compromised USB, CD, or DVD).
2. Lateral movement (LM): Attackers attempt to compromise other services on the target system or network. The objective is to try legitimate credentials that will allow them to persist in the system. Some of the LM techniques used are standard operating system tools (e.g., RDP, PsExec, and Powershell), and exploit a vulnerability (zero-day exploit).
3. Command and control (C&C): When the system has been compromised, it is necessary to establish an external connection to exfiltrate data. Attackers use services such as HTTP, HTTPS, or FTP. Also, they can use tools such as remote connection tools like VNC (Virtual Network Computing) or RDP.

4.2. Four Stage Attack

The intrusion kill chain (IKC) is a four-stage model that identifies the behaviours and purposes of an APT attack [7]. The stages are described below:

1. Information collection: In this initial stage, the recognition of the network is made using scanning or social engineering tools.
2. Intrusion: In this stage spear-phishing techniques, malicious email attachments or backdoors are used to obtain access privileges.
3. Latent expansion: The attacker attempts to maintain control in order to obtain data that will allow the attacker to continue with the expansion within the network.
4. Information theft: The attacker establishes a connection to a server, and the stolen data is transferred. Encryption techniques can be used to camouflage the extracted data.

Another approach to the four-stages life cycle was detailed in [61]. In this approach, the stages are described as follows:

1. Initial compromise: The techniques used are social engineering and spear-phishing.
2. C&C: A communication channel is established between a committed server and the target.
3. Lateral movement: Attackers seek to collect internal information and move between hosts with critical vulnerabilities.
4. Attack achievement: The attack is completed, and the theft of sensitive information begins.

4.3. Five Stage Attack

In this work [69], a model for analysing the life cycle of APT organised in five stages was proposed; this model is called attack chain. The five stages are as follows:

1. Delivery: Spear-phishing is used to send emails to recipients within the network.
2. Exploit: The vulnerabilities of the services, system or applications are exploited.
3. Installation: In this stage, it is possible to install malware such as RAT (Remote Access Tool).
4. Command and control: The attacker has remote access to a compromised host or server.
5. Actions: The actions carried out consist of gaining access to other hosts or servers on the same network to extract confidential information.

The authors in [9] describe another five-stage model as follows:

1. Recount: The target is selected; the information related to the target that is published is sought.
2. Incursion: The attacker obtains access to the network through stolen credentials with techniques such as SQL injection or with the use of malware.
3. Discovery: The attacker searches for confidential data in the system.
4. Capture: The attacker installs an undetectable rootkit to collect confidential data for an extended period.
5. Exfiltration: The collected data is sent to the C&C servers.

4.4. Six Stage Attack Model

The authors in [70,71] proposed a six-stage life cycle model to describe an APT attack. This model emphasises that attackers must trick a person into running malware and exploit any zero-day vulnerability. Attackers then access the corporate network from the compromised computer and execute a cycle of hard-to-reach manoeuvres to achieve their ultimate goals. The six stages of this life cycle are as follows:

1. Information Gathering: The objective of this stage is to gather information on the structure of the organisation through public social network profiles.
2. Point of entry: Social engineering, spear-phishing and zero-day exploit are the most used techniques for the victim to allow the attacker to gain access to the computer.
3. Command and control server: The attacker establishes a connection from the compromised host to the C&C server to maintain the connection. Secure Sockets Layer (SSL) encryption is the method usually used to send traffic to the C&C server.
4. Lateral movement: The attacker can move through the network to find a vulnerable host when access has been gained.
5. Data of interest: Critical information on hosts or servers is identified.
6. External server: the data of interest is transmitted to the C&C servers of the attackers.

The authors in [2] have adopted a six-stage life cycle based on the intrusion kill chain model. This model organises the stages as follows:

1. Reconnaissance and weaponization is a preparation stage to study and collect technical information from the target organisation. Some techniques used are social engineering and open-source intelligence (OSINT).
2. Delivery: The attackers send the exploits to the targets directly or indirectly, for example, a direct technique can be through spear-phishing and in an indirect way through watering-hole attack.
3. Initial intrusion: The information obtained in the previous stage (such as credentials), allows attackers to gain access to the target, execute malicious code and exploit vulnerabilities.
4. Command and control: The attackers establish a mechanism to take control of the compromised hosts; for this, the attackers create social networking sites, TOR anonymity networks or use remote access tools.
5. Lateral movement: When attackers have established a connection to their C&C servers, they move around the network of the organisation looking for useful information to gain access to other systems.
6. Data exfiltration: Attackers send critical encrypted information to servers.

4.5. Seven Stage Attack Model

In the article [72], a general approach of an APT attack of seven stages was presented. These stages are:

1. Research: The attackers seek publicly available information about the victim.
2. Preparation: The attackers prepare an initial attack to exploit the vulnerabilities using network scanning to create custom exploits.
3. Intrusion: The attackers launch the first attack which usually consists of spear-phishing.
4. Conquering the network: Remote access tools or backdoors to control the system are used when the attacker has compromised at least one host.
5. Hiding presence: The attacker seeks to remain hidden in the network for a long time. The attack can have periods of inactivity.
6. Gathering data: The attacker looks for data of interest and masks it as legitimate traffic to be slowly extracted.
7. Maintaining access: The attacker can modify or create exploits, remote access tools and C&C servers, to obtain prolonged access to the network.

The Lockheed Martin company proposed a seven-stage life cycle called cyber kill chain (CKC) [73]. This model seeks to understand how an attack works to enrich the understanding of tactics, techniques, and procedures used by attackers. These stages are described below:

1. Reconnaissance: The attacker performs a preliminary reconnaissance of the network of the organisation, using spear-phishing techniques, port scanning, and social engineering.
2. Weaponization: The attacker builds a payload that is sent to the victim. It usually consists of an exploit with a RAT/trojan delivery.
3. Delivery: The payload created is sent to the victim through mail, websites or a removal devices.
4. Exploitation: The attacker executes the exploit that has been sent to the victim.
5. Installation: A Trojan and/or remote access trojan (RAT) is installed when the attacker gains access to the system.
6. Command and control: The remote access software connects to C&C of the attacker.
7. Actions and objectives: The attacker performs data exfiltration compromising the integrity and availability of the data. This stage can last weeks, months or even years.

4.6. Eight Stage Attack Model

Mandiant (now FireEye), proposed an eight-stages model after analysing the APT1 campaign [14] organised as follows:

1. Initial recon: Initial recognition of the target.
2. Initial compromise: Describes the methods used for the first intrusion of the target, e.g., spear-phishing.
3. Establish foothold: Consists of ensuring control of the target from outside the network, for example, C&C servers.
4. Escalate privileges: The attacker looks for credentials that permit access to more resources within the system.
5. Internal recon: In this stage, the attacker collects all the possible information about the victim.
6. Move laterally: The attacker can connect and share resources using legitimate credentials.
7. Maintain presence: The attacker performs actions to remain for an extended period within the network without being detected.
8. Complete mission: The information of interest is compressed to be sent to the C&C servers.

4.7. Eleven Stage Attack Model

ATT and CK analyses of tactics are shown to be the distinct stages of an attack that a threat actor works through to accomplish the strategic goal [1]. ATT and CK matrix describes the next tactics:

1. Initial access: Consists of the initial contact with the target to search for patient zero.
 2. Persistence: The attacker seeks to gain access for a long time in the target.
 3. Privilege escalation: To obtain privileges in the network is necessary to install malware or gain access to confidential data.
 4. Discovery: Consists of obtaining relevant information from the target, such as system location or usernames.
 5. Lateral movement: Refers to how the attacker moves within the network to search for important vulnerable information or services.
 6. Collection: Collecting relevant information.
 7. Exfiltration: Extracting the collected data.
- The following stages achieve the objective of the attack, and can be executed in parallel with the previous seven stages.
8. Execution: The execution of malware through remote connections that are carried out between the initial access stage and lateral movement.
 9. Defence evasion: Consists of not being detected by the defence and detection mechanisms, for example, firewall or logs.

10. Credential access: Refer to accessing the compromised system with valid credentials.
11. Command and control: Consists of creating a C&C channel to communicate the attacker servers with the compromised systems of the target.

The proposed life cycles have similarities in the methods and techniques used by the attackers in each stage. Consequently, one stage of a life cycle can be divided into several stages in other approaches to explain in more detail how the APT attack works. For this reason, researchers can select the life cycle that suits their work or take a proposed life cycle as the basis for creating a new life cycle. Each APT attack has unique characteristics, and several may use a very similar life cycle.

5. A Novel Proposal

Recently, several life cycles were proposed to describe an APT attack. In each phase of these cycles, the tactics, techniques, and procedures (TTP) used by the attackers (actors) are specified. The basis of some of these proposed models are IKC, CKC, and attack chain models. CKC is a well-known model and has been used as the basis for the seven-stages life cycles analysed; however, IKC was used as the basis for both four-stage and six-stage models.

In Table 4, a comparison of the life cycles of an APT attack was made. It can be seen that cycles with the same amount of phases have different ways to explain the behaviour of the APT. A three-stage model can describe the same steps as a five-stage to eleven-stage model. For this reason, stages with similar characteristics were grouped together. For example, the stage of initial compromise of the three-stage model can be similar to the stages of initial access, persistence, and privilege escalation of the eleven-stage model.

Another point to note is that some authors indicate that the C&C connection is made before starting the network scan. Other authors place this stage at the end of the cycle when the data is extracted. An eleven-stage model describes that there are stages that can be developed in parallel with the main stages of the cycle to maintain persistence in the target and extract critical information when it is found.

The reviewed approaches agree that the first steps are the study and analysis of the target. Then an exploitation of the vulnerabilities occurs to compromise one or more hosts within the target. Finally, the extraction of the data to a C&C server is performed in a stealthy manner by the attackers. The Mandiant life cycle describes the cleanup as a final stage that when executed, the organisation may not detect that it has been attacked. It is important to remark that the life cycles are analysed to provide an idea of how an APT attack works, however, each attacker can carry out the stages in any order and use the TTPs that are adapted to meet the objectives.

APT attacks are targeted and work in stealthy actions; therefore, it is difficult to detect them early. A five-stage life cycle model is proposed in this paper. In each stage, the most commonly used related TTP, according to MITRE [1], have been included.

In this case, an APT attack was divided into passive and active actions. These actions extend from social engineering attacks to specific attacks such as unauthorised access to servers. Therefore, actions that do not modify data or interfere with the transmission of information were considered as passive actions, e.g., port scanning techniques; and the actions that modify data, remove information, or change the flow of packets were considered as active actions, e.g., distributed denial of service.

ML techniques provide a solution for the analysis of large amounts of data, such as IDS alerts, logs, or unauthorised remote connections; analysis of this data can help IT administrators identify anomalous behaviour on the network, which can be associated with misuse of computer resources, common malware installed on a network host, or an APT attack.

Table 4. Comparison between different proposed APT life cycle approaches.

3 Stages [68]	4 Stages [7]	4 Stages [61]	5 Stages [9]	5 Stages [69]	6 Stages [70]	6 Stages [2]	7 Stages [72]	7 Stages [73]	8 Stages [14]	11 Stages [1]
Initial Compromise	Information Collection Intrusion phase	Initial Compromise	Reconnaissance Incursion	Delivery	Intelligence gathering Initial Compromise	Reconnaissance and weaponization Delivery	Research Preparation	Reconnaissance Weaponization	Initial recon Initial compromise	Initial access Persistence
Lateral movement	Lateral expansion	C&C Lateral movement	Discovery Capture	Exploit Installation	C&C Lateral movement Assets/Data discovery	Initial intrusion C&C Lateral movement	Intrusion Conquering network Hiding presence	Delivery Exploitation Installation	Establish foothold Escalate privileges Internal recon	Privilege Escalation Discovery Lateral movement
Command and control	Information theft phase	Attack achievement	Ex-filtration	C&C Actions	Data ex-filtration	Data ex-filtration	Gathering data Maintaining access	C&C Actions on objective	Move laterally Maintain presence Complete mission	Collection Exfiltration Stages executed in parallel: Execution, Defence evasion, Credential access, and Command & control

This model aims to detect an APT attack early and efficiently; however, some APT may go unnoticed by one or more stages, so detection solutions are proposed from the beginning to the end of the active attack using ML techniques. The stages of this model are described below:

- * Target discovery: This stage consists of the passive exploration of the network organisation, to obtain the approximate details of the IT structure to be attacked. To achieve this goal, the attacker can perform port scanning techniques (e.g., Nmap tools), search for indexed services on the Internet (web surveillance cameras, servers or SCADA systems, with tools such as Shodan), public profiles in social networks of the employees, and OSINT reconnaissance tools (e.g., spider foot).

These types of techniques used to recognising the resources of an organisation are difficult to detect by ML techniques because these attacks are usually made passively. A passive attack does not modify or interfere with communication but rather listens to or monitors the information that was transmitted. Information that can be found on the Internet can be collected for sale on the darknet; these attacks can require the use of multiple specialised tools over a long period.

Therefore, it is recommended to close unused ports, use firewalls, IDS, and secure private virtual connections (VLAN and VPN), create password policies, and user awareness of the organisation.

1. Exploitation toolset: This stage objective is to gain access on the target network through the vulnerabilities detected in the previous stage, or by tricking an employee of the organisation. The process starts with the elaboration of a method to reach the target. For this, the attacker uses techniques such as spear-phishing in different ways, such as valid accounts or replication through USB. Later, the attacker exploits the detected vulnerability using scripting, Powershell, and user execution; then, remote management tools are used to establish a connection with the target network.

To prevent an employee from being attacked, it is recommended to avoid using personal devices within the network of an organisation and to avoid opening suspicious files when in doubt. However, ML techniques allow for the creation of automated solutions to detect possible attacks at an early stage. For example, a module can be created that scans email for malicious links or malicious files.

Another solution would be to scan network traffic for remote connection packets from unauthorised servers, the analysis of logs to detect anomalous activity within the network, and finally, software updates. The implementation of these ML solutions requires a training dataset in the normal flow of the organisation and another dataset with anomalous network flow. Then, the ML algorithm that provides the best accuracy must be chosen. Finally, tests must be performed in a controlled environment.

The ML algorithms that have given the best results have been k -NN and SVM. During initial training or retraining of the algorithm, datasets with flows from other attack techniques can be added to improve detection.

2. Internal intrusion: When the attacker has compromised the first host on the network organisation, the next objective is to escalate privileges to access confidential and critical information. For this, the attacker must be able to maintain persistence during an extended period since this stage is the longest one. Persistence on a network can be done through redundant access, account manipulation, or a web shell. Access to credentials can be obtained through brute force techniques, account manipulation, forced authentication, or credential dumping. Another essential step performed by the attacker is the evasion of defence systems (e.g., IDS, IPS, and firewall); this can be done through proxy connections and the obfuscation of files or information.

Some solutions are to use ML techniques for the analysis of logs generated by IDS/IPS for the detection of possible APT attack patterns, (failed access to SSH, FTP, or telnet services), analysis of system logs (unauthorised program installations, directories, and files with coded names,

unknown hosts on the network). Some ML algorithms that can be used are *k*-means, NB, and SVM.

3. Set data extraction channels: This stage consists of establishing a connection with the C&C attacker server to send all the collected information, usually sending the data compressed and encrypted and limiting the size of the packets. The data are usually sent during hours of lower network bandwidth usage. The attacker can use fast-flux techniques to make the connections. Data can be stored on a host within the network and sent to the C&C server when the target is completed or sent in small packets at different times.

Some techniques for data collection are automated collection, email collection, and man in the browser. Data extraction can be automated and on different media (e.g., alternative protocols, network medium, and physical medium). The tools used in C&C servers are domain generation protocols, remote access tools, and multilayer encryption.

As a solution for the detection of sending data to C&C servers, ML techniques can be used to search for hosts with encrypted data, connections with random IP addresses and DNS, and encrypted data flows to unknown and unauthorised servers. In this stage, *k*-NN and *k*-means algorithms can be used for APT detection.

- * Eliminate footprints: When the attacker has completed the mission, the next step is to remove all possible attack traces on the network and compromised systems, for example, these traces can be logs, compressed files, installed software, or malware. If the attacker has reached this stage, the organisation may not know that it has been compromised and attacked with an APT. Therefore, it would be difficult to check how much information the attacker has extracted and how long it has remained within the network. For this reason, the attack must be identified early.

The proposed five-stage model matches well the steps followed in an APT attack. The first and last stages of the model are considered passive actions because in most cases, they do not identify a real attack. The three active stages of the model have defined some of the most used techniques in the attacks and have detailed possible measures to mitigate the attack. It is important to remember that each organisation must include in its cybersecurity plan the security policies that are adapted to its infrastructure, without forgetting that users must be made aware frequently.

Another advantage of this proposed model is that attacks have been considered in all stages. In contrast to other models studied it has been observed that the authors have proposed a life cycle, but the detection of APT is carried out in only one stage of the life cycle; moreover, this stage does not always correspond to the first stage of the cycle.

The identification of possible attacks by stages in our model, facilitates the detection of APT, helping to anticipate these anomalous behaviours on the network.

6. Conclusions and Future Work

Advanced persistent threats are sophisticated targeted and personalised attacks. Attackers are often called actors and are classified into government and private actors. These actors use various techniques to execute the attack; the techniques become more sophisticated when the attack progresses successfully. The machine learning techniques and models frequently used to detect an APT attack are SVM, *k*-NN, and DT.

Furthermore, the life cycles of an APT attack were analysed; different stages form these cycles. The stages of the different cycles have similarities that are able to be grouped; however, these stages represent a non-linear order of attack behaviour. Finally, a five-stage life cycle model was described, the most used techniques were identified, and possible mitigation techniques were proposed. The use of ML techniques that have given good results was recommended.

The advantage of this model is that both passive and active stages in the life cycle were considered, and it simplifies the behaviour of an APT attack. One shortcoming is that the datasets for the training of the different ML algorithms have not been obtained.

As future work, a framework based on the proposed five-stage model is proposed. Also, the creation of a dataset containing network flows (normal and malicious) to train the ML algorithms used in the framework is recommended. Finally, the effectiveness of the framework can be tested by simulating an APT attack in a controlled space.

Author Contributions: Conceptualization, S.Q.-B. and A.M.d.R.; methodology, S.Q.-B.; writing—original draft preparation, S.Q.-B.; writing—review and editing, S.Q.-B. and A.M.d.R.; supervision, A.M.d.R.; project administration, A.M.d.R.; funding acquisition, A.M.d.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been partially supported by Ministerio de Ciencia, Innovación y Universidades (MCIU, Spain), Agenda Estatal de Investigación (AEI, Spain), and Fondo Europeo de Desarrollo Regional (FEDER, UE) under project with reference TIN2017-84844-C2-2-R (MAGERAN) and the project with reference SA054G18 supported by Consejería de Educación (Junta de Castilla y León, Spain). S. Quintero-Bonilla has been supported by IFARHU-SENACYT scholarship program (Panama), and the educational leave by Technological University of Panama.

Acknowledgments: The authors would like to recognise the work of the anonymous reviewers for their valuable comments and suggestions.

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

References

1. Swisscom. *Targeted Attacks Cyber Security Report 2019*; Technical report; Swisscom (Switzerland) Ltd. Group Security: Bern, Switzerland, 2019.
2. Chen, P.; Desmet, L.; Huygens, C. A Study on Advanced Persistent Threats. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin, Germany, 2014; Volume 8735 LNCS, pp. 63–72.
3. Fireeye. *M-Trends 2019: Fireeye Mandiant Services Special Report*; Technical report; Fireeye: Milpitas, CA, USA, 2019.
4. Lemay, A.; Calvet, J.; Menet, F.; Fernandez, J.M. Survey of publicly available reports on advanced persistent threat actors. *Comput. Secur.* **2018**, *72*, 26–59. [CrossRef]
5. Bai, T.; Bian, H.; Daya, A.A.; Salahuddin, M.A.; Limam, N.; Boutaba, R. A Machine Learning Approach for RDP-based Lateral Movement Detection. In *Proceedings of the 2019 IEEE 44th Conference Local Computer Networks*, Osnabrueck, Germany, 14–17 October 2019; pp. 242–245.
6. Ghafir, I.; Hammoudeh, M.; Prenosil, V.; Han, L.; Hegarty, R.; Rabie, K.; Aparicio-Navarro, F.J. Detection of advanced persistent threat using machine-learning correlation analysis. *Futur. Gener. Comput. Syst.* **2018**, *89*, 349–359. [CrossRef]
7. Zhang, R.; Huo, Y.; Liu, J.; Weng, F. Constructing APT Attack Scenarios Based on Intrusion Kill Chain and Fuzzy Clustering. *Secur. Commun. Netw.* **2017**, *2017*, 1–9. [CrossRef]
8. Threat Intelligence Team, M.L. APT36 Jumps on the Coronavirus Bandwagon, Delivers Crimson RAT. Available online: <https://blog.malwarebytes.com/threat-analysis/2020/03/apt36-jumps-on-the-coronavirus-bandwagon-delivers-crimson-rat/> (accessed on 16 March 2020).
9. Jeun, I.; Lee, Y.; Won, D. A Practical Study on Advanced Persistent Threats. *Commun. Multimed. Secur.* **2012**, *8735*, 144–152.
10. Falliere, N.; Murchu, L.O.; Chien, E. W32. stuxnet dossier. *White Pap. Symantec Corp., Secur. Response* **2011**, *5*, 29.
11. FireEye. *Follow the money: Dissecting the Operations of the Cyber Crime Group FIN6*; Technical Report; FireEye: Milpitas, CA, USA, 2016.
12. Coopers, Pricewaterhouse. *Operation Cloud Hopper*; Technical report; PwC UK Cyber Security and Data privacy: London, UK, 2017.
13. FireEye. *Double Dragon: APT41, a Dual Espionage and Cyber Crime Operation*; Technical report; FireEye: Milpitas, CA, USA, 2019.
14. Mandiant. *APT1 Exposing One of China's Cyber Espionage Units*; Technical report; Mandiant: Alexandria, VA, USA, 2013.

15. Krombholz, K.; Hobel, H.; Huber, M.; Weippl, E. Advanced social engineering attacks. *J. Inf. Secur. Appl.* **2015**, *22*, 113–122. [\[CrossRef\]](#)
16. Aleroud, A.; Zhou, L. Phishing environments, techniques, and countermeasures: A survey. *Comput. Secur.* **2017**, *68*, 160–196. [\[CrossRef\]](#)
17. Symantec. *Internet Security Threat Report*; Technical Report 2; Symantec: Tempe, AZ, USA, 2016.
18. Tanaka, Y.; Akiyama, M.; Goto, A. Analysis of malware download sites by focusing on time series variation of malware. *J. Comput. Sci.* **2017**, *22*, 301–313. [\[CrossRef\]](#)
19. Paganini, P. Turla APT Group's Espionage Campaigns Now Employs Adobe Flash Installer and Ingenious Social Engineering. Available online: <https://www.cyberdefensemagazine.com/turla-apt-groups-espionage-campaigns-now-employs-adobe-flash-installer-and-ingenious-social-engineering/> (accessed on 20 August 2019).
20. ThaiCERT. Threat Group Cards: A Threat Actor Encyclopedia. Available online: https://www.thaicert.or.th/downloads/files/A_Threat_Actor_Encyclopedia.pdf (accessed on 24 June 2019).
21. Paganini, P. Iran-Linked APT33 Updates Infrastructure Following Its Public Disclosure. Available online: <https://securityaffairs.co/wordpress/87784/apt/apt33-updates-infrastructure.html> (accessed on 21 November 2019).
22. Adams, C. Learning the lessons of WannaCry. *Comput. Fraud Secur.* **2018**, *2018*, 6–9. [\[CrossRef\]](#)
23. Cordey, S. Trend Analysis: The Israeli Unit 8200—An OSINT-based study. Technical Report; Center for Security Studies (CSS), ETH Zürich: Zürich, Switzerland, 2019.
24. Kasperky Lab. *The Duqu 2.0-Technical Details (V2.1)*; Technical Report; Kasperky Lab: Moscow, Russia, 2015.
25. Kaspersky Lab. *Targeted Cyberattacks LOGBOOK*; Kasperky Lab: Moscow, Russia, 2019.
26. Dua, S.; Du, X. *Data Mining and Machine Learning in Cybersecurity*; Auerbach Publications: London, UK, 2011.
27. Kaviani, S.; Sohn, I. Influence of random topology in artificial neural networks: A survey. *ICT Express* **2020**. [\[CrossRef\]](#)
28. Khraisat, A.; Gondal, I.; Vamplew, P.; Kamruzzaman, J. Survey of intrusion detection systems: Techniques, datasets and challenges. *Cybersecurity* **2019**, *2*, 20. [\[CrossRef\]](#)
29. Da Silva, I.N.; Hernane Spatti, D.; Andrade Flauzino, R.; Liboni, L.H.B.; dos Reis Alves, S.F. *Artificial Neural Networks*; Springer International Publishing: Cham, Switzerland, 2017; pp. 1–307. [\[CrossRef\]](#)
30. Dahl, G.E.; Dong, Y.; Li, D.; Acero, A. Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition. *IEEE Trans. Audio. Speech. Lang. Process.* **2012**, *20*, 30–42. [\[CrossRef\]](#)
31. Valverde Ramírez, M.C.; de Campos Velho, H.F.; Ferreira, N.J. Artificial neural network technique for rainfall forecasting applied to the São Paulo region. *J. Hydrol.* **2005**, *301*, 146–162. [\[CrossRef\]](#)
32. Erkamaz, O.; Ozer, M.; Perc, M. Performance of small-world feedforward neural networks for the diagnosis of diabetes. *Appl. Math. Comput.* **2017**, *311*, 22–28. [\[CrossRef\]](#)
33. Chu, W.L.; Lin, C.J.; Chang, K.N. Detection and Classification of Advanced Persistent Threats and Attacks Using the Support Vector Machine. *Appl. Sci.* **2019**, *9*, 4579. [\[CrossRef\]](#)
34. Joshi, A.V. *Machine Learning and Artificial Intelligence*; Springer International Publishing: Cham, Switzerland, 2020; Volume 64; pp. 49A–60A. [\[CrossRef\]](#)
35. Martínez Torres, J.; Iglesias Comesaña, C.; García-Nieto, P.J. Review: machine learning techniques applied to cybersecurity. *Int. J. Mach. Learn. Cybern.* **2019**, *10*, 2823–2836. doi:10.1007/s13042-018-00906-1. [\[CrossRef\]](#)
36. Alloghani, M.; Al-Jumeily, D.; Hussain, A.; Mustafina, J.; Baker, T.; Aljaaf, A.J. Implementation of Machine Learning and Data Mining to Improve Cybersecurity and Limit Vulnerabilities to Cyber Attacks. In *Nature-Inspired Computation in Data Mining and Machine Learning*; Yang, X.S., He, X.S., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 47–76. [\[CrossRef\]](#)
37. Cleophas, T.J.; Zwinderman, A.H. *Modern Bayesian Statistics in Clinical Research*; Springer International Publishing: Cham, Switzerland, 2018.
38. von Davier, M.; Lee, Y.S. *Handbook of Diagnostic Classification Models*; Methodology of Educational Measurement and Assessment, Springer International Publishing: Cham, Switzerland, 2019; p. 646. [\[CrossRef\]](#)
39. Gou, J.; Ma, H.; Ou, W.; Zeng, S.; Rao, Y.; Yang, H. A generalized mean distance-based k-nearest neighbor classifier. *Expert Syst. Appl.* **2019**, *115*, 356–372. [\[CrossRef\]](#)
40. Pan, Y.; Pan, Z.; Wang, Y.; Wang, W. A new fast search algorithm for exact k-nearest neighbors based on optimal triangle-inequality-based check strategy. *Knowl.-Based Syst.* **2020**, *189*, 105088. [\[CrossRef\]](#)

41. Abu Alfeilat, H.A.; Hassanat, A.B.; Lasassmeh, O.; Tarawneh, A.S.; Alhasanat, M.B.; Eyal Salman, H.S.; Prasath, V.S. Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review. *Big Data* **2019**, *7*, 221–248. [\[CrossRef\]](#)
42. Awad, M.; Khanna, R., Hidden Markov Model. In *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*; Apress: Berkeley, CA, USA, 2015; pp. 81–104. [_5. \[CrossRef\]](#)
43. Portugal, I.; Alencar, P.; Cowan, D. The use of machine learning algorithms in recommender systems: A systematic review. *Expert Syst. Appl.* **2017**, *97*, 205–227. [\[CrossRef\]](#)
44. Olivieri, A.C., Principal Component Analysis. In *Introduction to Multivariate Calibration: A Practical Approach*; Springer International Publishing: Cham, Switzerland, 2018; pp. 57–71. [_4. \[CrossRef\]](#)
45. Joshi, V.B.; Raval, M.S.; Gupta, D.; Rege, P.P.; Parulkar, S.K. A multiple reversible watermarking technique for fingerprint authentication. *Multimed. Syst.* **2016**, *22*, 367–378. [\[CrossRef\]](#)
46. Wang, D.; Xu, J. Principal Component Analysis in the local differential privacy model. *Theor. Comput. Sci.* **2020**, *809*, 296–312. [\[CrossRef\]](#)
47. Yang, L.; Deng, M. Based on k-Means and Fuzzy k-Means Algorithm Classification of Precipitation. In Proceedings of the 2010 International Symposium on Computational Intelligence and Design, Hangzhou, China, 29–31 October 2010; Volume 1, pp. 218–221. [\[CrossRef\]](#)
48. Ahuja, R.; Chug, A.; Gupta, S.; Ahuja, P.; Kohli, S., Classification and Clustering Algorithms of Machine Learning with their Applications. In *Nature-Inspired Computation in Data Mining and Machine Learning*; Yang, X.S., He, X.S., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 225–248. [_11. \[CrossRef\]](#)
49. Guan, Z.; Bian, L.; Shang, T.; Liu, J. When Machine Learning meets Security Issues: A survey. *2018 IEEE Int. Conf. Intell. Saf. Robot.* **2018**, 158–165. [\[CrossRef\]](#)
50. Geluvaraj, B.; Satwik, P.M.; Ashok Kumar, T.A. The Future of Cybersecurity: Major Role of Artificial Intelligence, Machine Learning, and Deep Learning in Cyberspace. In *Lecture Notes on Data Engineering and Communications Technologies*; Springer Singapore: Singapore, 2019; Volume 15, pp. 739–747. [_67. \[CrossRef\]](#)
51. Mohanty, S.; Vyas, S. Cybersecurity and AI. In *How to Compete Age Artificial Intelligence*; Apress: Berkeley, CA, USA, 2018; pp. 143–153. [_6. \[CrossRef\]](#)
52. OWASP. Unvalidated Redirects and Forwards. 2019. Available online: https://cheatsheetseries.owasp.org/cheatsheets/Unvalidated_Redirects_and_Forwards_Cheat_Sheet.html (accessed on 19 September 2019).
53. Paganini, P. Phishers Continue to Abuse Adobe and Google Open Redirects. Available online: <https://securityaffairs.co/wordpress/91877/cyber-crime/adobe-google-open-redirects.html> (accessed on 11 October 2019).
54. Bhadane, A.; Mane, S.B. Detecting lateral spear phishing attacks in organisations. *IET Inf. Secur.* **2019**, *13*, 133–140. [\[CrossRef\]](#)
55. Lamprakis, P.; Dargenio, R.; Gugelmann, D.; Lenders, V.; Happe, M.; Vanbever, L. Unsupervised Detection of APT C&C Channels using Web Request Graphs. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin, Germany, 2017; Volume 10327 LNCS, pp. 366–387.
56. Zhao, G.; Xu, K.; Xu, L.; Wu, B. Detecting APT Malware Infections Based on Malicious DNS and Traffic Analysis *IEEE Access* **2015**, *3*, 1132–1142. [\[CrossRef\]](#)
57. Buczak, A.L.; Guven, E. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. *IEEE Commun. Surv. Tutorials* **2016**, *18*, 1153–1176. [\[CrossRef\]](#)
58. Liang, F.; Hatcher, W.G.; Liao, W.; Gao, W.; Yu, W. Machine Learning for Security and the Internet of Things: The Good, the Bad, and the Ugly. *IEEE Access* **2019**, *7*, 158126–158147. [\[CrossRef\]](#)
59. Su, Y.; Li, M.; Tang, C.; Shen, R. A Framework of APT Detection Based on Dynamic Analysis. In Proceedings of the 2015 4th National Conference on Electrical, Electronics and Computer Engineering, Xi'an, China, 12–13 December 2015; pp. 1047–1053.
60. Giura, P.; Wang, W. A Context-Based Detection Framework for Advanced Persistent Threats. In Proceedings of the 2012 International Conference on Cyber Security, Washington, DC, USA, 14–16 December 2012; pp. 69–74.
61. Wang, X.; Zheng, K.; Niu, X.; Wu, B.; Wu, C. Detection of command and control in advanced persistent threat based on independent access. In Proceedings of the 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, Malaysia, 22–27 May 2016; pp. 1–6. [\[CrossRef\]](#)

62. Aparicio-navarro, F.J.; Kyriakopoulos, K.G.; Ghafir, I.; Lambbotharan, S.; Chambers, J.A.; Technology, F. *Multi-Stage Attack Detection Using Contextual Information*; Loughborough University: Loughborough, UK, 2018; pp. 920–925.
63. Brogi, G.; Tong, V.V.T. TerminAPTor: Highlighting advanced persistent threats through information flow tracking. In Proceedings of the 2016 8th IFIP International Conference on New Technologies, Mobility and Security (NTMS), Larnaca, Cyprus, 21–23 November 2016.
64. Quintero-Bonilla, S.; del Rey, A.M. Proposed models for advanced persistent threat detection: A review. *Adv. Intell. Syst. Comput.* **2020**, *1004*, 141–148.
65. Sharma, P.K.; Moon, S.Y.; Moon, D.; Park, J.H. DFA-AD: A distributed framework architecture for the detection of advanced persistent threats. *Clust. Comput.* **2017**, *20*, 597–609. [\[CrossRef\]](#)
66. Siddiqui, S.; Khan, M.S.; Ferens, K.; Kinsner, W. Detecting Advanced Persistent Threats using Fractal Dimension based Machine Learning Classification. In Proceedings of the 2016 ACM on International Workshop on Security And Privacy Analytics, New Orleans, LA, USA, 11 March 2016; pp. 64–69.
67. Shenwen, L.; Yingbo, L.; Xiongjie, D. Study and research of APT detection technology based on big data processing architecture. In Proceedings of the 2015 IEEE 5th International Conference on Electronics Information and Emergency Communication, Beijing, China, 14–16 May 2015; pp. 313–316.
68. Ussath, M.; Jaeger, D.; Cheng, F.; Meinel, C. Advanced persistent threats: Behind the scenes. In Proceedings of the 2016 Annual Conference on Information Science and Systems (CISS), Princeton, NJ, USA, 16–18 March 2016; pp. 181–186.
69. Sexton, J.; Storlie, C.; Neil, J. Attack chain detection. *Stat. Anal. Data Min. ASA Data Sci. J.* **2015**, *8*, 353–363. [\[CrossRef\]](#)
70. Ghafir, I.; Prenosil, V. Proposed Approach for Targeted Attacks Detection. *Lect. Notes Electr. Eng.* **2016**, *362*, 73–80. [\[CrossRef\]](#)
71. Trend Micro. *The Custom Defense Against Targeted Attacks*; Technical report; Trend Micro: Tokyo, Japan, 2013.
72. Vukalovic, J.; Delija, D. Advanced Persistent Threats-detection and defense. In Proceedings of the 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 25–29 May 2015; pp. 1324–1330.
73. Lockheed Martin. *Cyber Kill Chain*; Lockheed Martin: Bethesda, MD, USA, 2009.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).