Detecting Malware in Cyberphysical Systems Using Machine Learning: a Survey

Montes F.[1], Bermejo J.[1], Sánchez L. E.[2], Bermejo J. R.[1], and Sicilia J. A.[1]*
[1] Escuela Superior de Ingeniería y Tecnología, Universidad Internacional de La Rioja
Avda. de La Paz 137, 26006 Logroño, La Rioja, Spain
[e-mail: francisco.montes529@comunidadunir.net, javier.bermejo@unir.net, juanramon.bermejo@unir.net, juanantonio.sicilia@unir.net]
[2] Research Group GSyA, University of Castilla-la Mancha, Paseo de la Universidad, 4 13071 Ciudad Real, Spain
[e-mail: luisenrique@sanchezcrespo.org]
*Corresponding author: Sicilia J. A.

Received July 21, 2020; revised August 17, 2020; accepted August 27, 2020; published March 31, 2021

Abstract

Among the scientific literature, it has not been possible to find a consensus on the definition of the limits or properties that allow differentiating or grouping the cyber-physical systems (CPS) and the Internet of Things (IoT). Despite this controversy the papers reviewed agree that both have become crucial elements not only for industry but also for society in general. The impact of a malware attack affecting one of these systems may suppose a risk for the industrial processes involved and perhaps also for society in general if the system affected is a critical infrastructure.

This article reviews the state of the art of the application of machine learning in the automation of malware detection in cyberphysical systems, evaluating the most representative articles in this field and summarizing the results obtained, the most common malware attacks in this type of systems, the most promising algorithms for malware detection in cyberphysical systems and the future lines of research in this field with the greatest potential for the coming years.

Keywords: Cyber-physical System, IoT, Malware, Machine Learning, Detection

1. Introduction

With the rise of CPS in all areas of society researchers are focusing their efforts on the security challenges posed by these types of systems, focusing mainly on research into the vulnerabilities they possess and their attacks and the countermeasures that need to be deployed to stop them.

Within the different types of attacks that can suffer CPS, this article focuses on those originated by malware. This is a field in constant evolution, where researchers are constantly proposing new methods to counteract the effects of malware, and the creators of this type of software are developing increasingly sophisticated and efficient malware to achieve their goals. According to [1] malware is no longer static software designed to be used only once. Rather, behind its development there is a business model that seeks to maximize its investments through the evolution of its software. In this way, the author of a malicious software tries to amortize as much as possible the time and effort spent in the creation of malware, reusing as many pieces as possible in each iteration at the same time that evolves it so that it can be used in several attacks with new features to evade the controls that security tools implement with each iteration.

CPS have become ubiquitous in recent years [2] and are at the heart of today's critical infrastructure and industrial applications, so ensuring that these systems are secure is a major concern [3].

Cyber-physical systems have a direct interaction with industrial processes that may involve the presence of multiple components such as machines, robots or other moving objects and humans. In this context, according to [4, 5], the safety of the human individuals involved in the process is crucial, so low system latency is particularly important for real-time response.

In this sense, the detection of a threat, such as a malware attack, which may pose a risk of malfunctioning or degradation in response times in a cyberphysical system is a crucial aspect [6].

Perhaps one of the most relevant works to illustrate the risk that malware can pose to a CPS is that carried out by [7]. It analyses malware that takes advantage of the use of machine learning techniques to maximize its ability to cause damage to the target system. It is precisely the aim of this article to deepen the state of the art of malware detection in CPS environments using machine learning algorithms to boost its automatization to know what the scientific advance in the field and the most promising future areas of work is.

The rest of the article is organized as follows. Section II describes the relationship between the IoT and CPS, with the most commonly used definitions of each of these terms and the properties shared by them. A review of the commonly accepted definitions for malware and their classification ontology is introduced in section III. Section IV gives way to the introduction of the particularities of malware in CPS and IoT. Section V reviews the different methods of malware analysis, identifying the differences between them and the benefits of using each of them. Section VI reviews the different malware detection methods and their differences. Section VII is a review of the state of the art of applying machine learning to malware detection in cyberphysical environments with a selection of the most relevant articles and the main results obtained in each of them. Finally, section VIII concludes the article with a summary of the most relevant observations and future work that can be developed in this field.

2. Internet of Things and Cyber-physical Systems

The Internet of Things (IoT), takes technological interaction capabilities to a new level in an attempt to link the physical world, through devices equipped with sensors and actuators, and the digital world, through communication networks.

The origin of the term IoT is attributed to Kevin Ashton in 1999 in an attempt to promote the use of RFID technology among c-level executives in the organisation where he worked. However, the term spent years being unknown to the general public and it was not until 2011 that its popularity began to grow when Gartner included it in its study "*Hype Cycle for Emerging Technologies, 2011"*.

This popularity was reinforced when in 2013 IDC published a report entitled *"Internet of Things (IoT) 2013 to 2020 Market Analysis: Billions of Things, Trillions of Dollars (IDC #243661)"* indicating that it would be an industry with an 8.9 trillion-dollar market in 2020. The term became commonplace among the general public in 2014 after Google's purchase of Nest and the prestigious consumer electronics trade show (CES), held annually in Las Vegas, used the term IoT as its main topic.

It is surprising that since the term IoT became popular among the general public, the number of scientific articles dealing with malware for IoT devices have tripled compared to those dealing with the same topic for cyberphysical systems, as can be seen in **Fig. 1**. Reversing the trend of previous years where publications of articles dealing with malware in CPS were always higher.

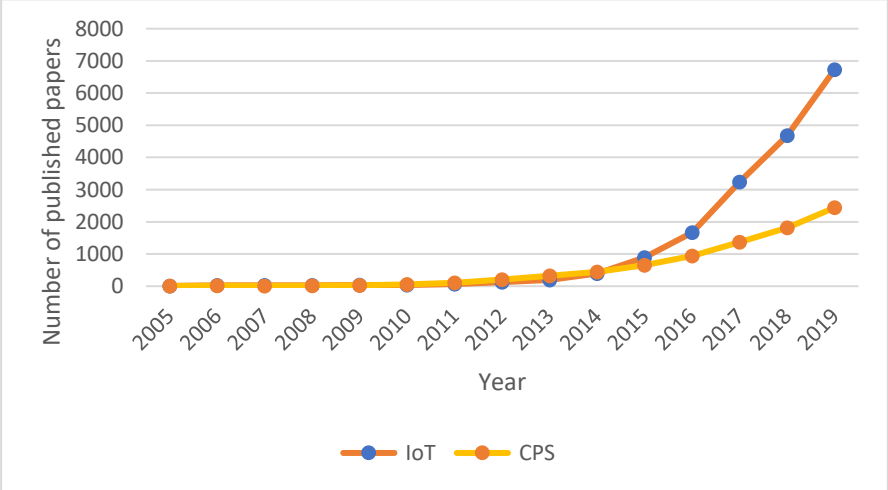

Fig. 1. Evolution of the publication of scientific articles on IOT and CPS

The term IoT has ended up capturing the attention of the general public to define the scenario of interconnection between the physical and digital world, eclipsing some other terms such as M2M, Industry 4.0, Industrial Internet or IIoT, which may been seen as subsets of the IoT that encompass specific sectors [8].

Therefore, it is not surprising that there are several definitions of IOT that have been formulated by different researchers or standardization bodies. Among them, the following should be highlighted:

- an open and complete network of intelligent objects that have the capacity to self-organize, share information, data and resources, reacting and acting on situations and changes in the environment is described in [9].

- use the term as an umbrella to cover various aspects related to the extension of the Internet and the Web in the physical world, through the deployment of devices with identification, detection and action capabilities as described in [10].

Although there is no common definition to describe the IOT, all the revised definitions point in the same direction of expanding the virtual world, providing everyday devices with new sensing and computing capabilities to participate in an evolution of what we now know as the Internet.

In contrast to the definition of IoT that seeks to extend the world of the Internet and Web to physical systems. In 2007, the concept of cyber-physical systems emerged in the industrial control systems environment, due the importance that this type of systems was gaining in the automation of industrial processes.

The most accepted definition for this type of system is that of [11], which defines them as a set of computational and communication elements which, using information and knowledge of the processes, independently control physical systems.

In spite of such a different origin, in [8] it is proposed that between both terms the scientific literature collects overlaps in 4 dimensions as indicated in **Fig. 2**. For [12] between the IoT and the CPS there is only a partial overlap, since it considers that between both there are obvious differences because IoT emphasizes the interconnection however CPS do it in the exchange of information and feedback. In [11, 13] both concepts are treated as equivalent and the authors conclude that there are no clear distinctions between the two. Whereas [14, 15] consider the IOT as a subset within CPS, since the latter can be highly interconnected systems internally, but without strictly necessary network connectivity beyond the boundaries of their own system, as is the case with autonomous vehicles.

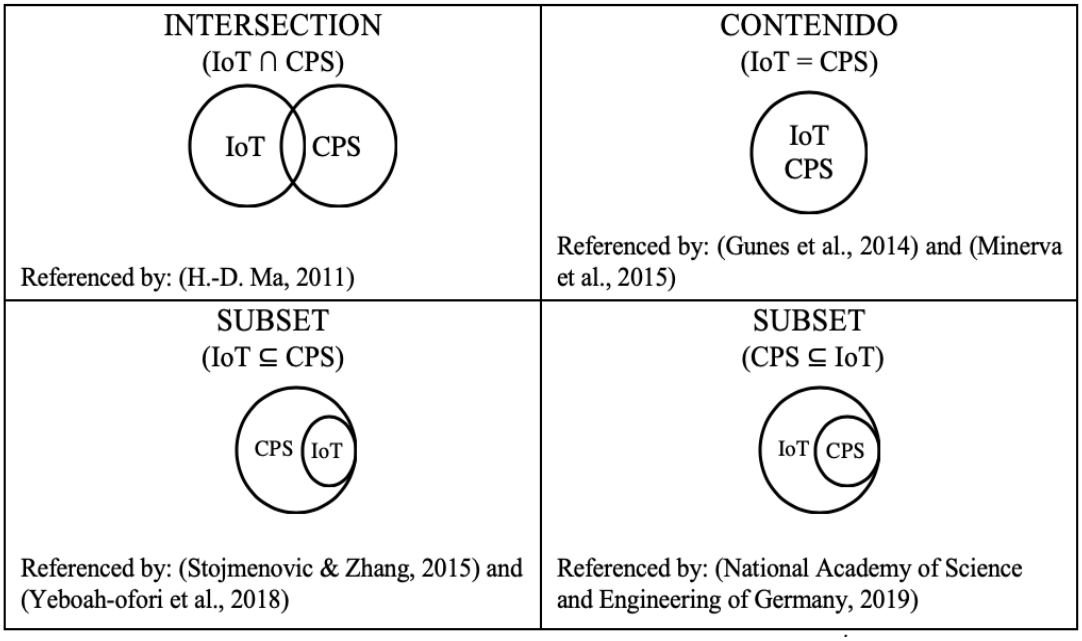

Fig. 2. Overlap quadrant between IoT and CPS

On the other hand, there are other authors such as [16], who consider that CPS are a subset of the IoT environments, since they consider that the former serve as a platform or building block for IoT.

Despite the obvious benefits that CPS and the IoT can provide, [17] highlights that they will also face the following challenges:

- Interoperability: the success of the Internet has been based, among other things, on the high level of interoperability between the systems that were connected. Today, in the different industries, there are different standards to support the specific applications of each of them, so interoperability will be a key element for the success of the CPS and the IoT.
- Economies of scale: the cost of sensors and actuators must be reduced to reach a point where integrating them into a large number of physical objects is economically viable.
- Efficiency: one of the essential challenges for a massive use of the IoT is to achieve energy efficiency.
- Security: this is possibly the main challenge for it to be the main axis on which the Internet of the future is built, especially when industries such as connected cars or intelligent medical devices are becoming increasingly prominent.

To overcome the safety challenges that CPS and IoT will face, according to [18], it is necessary to ensure a secure exchange of information, both between different devices that will be part of the systems, and with the people who will use these devices. Above all, it is necessary to ensure that these devices have sufficient protection mechanisms against potential attacks.

In [19] a systematic review of the literature on the security of cyberphysical systems is carried out and the main security challenges in these systems are identified. This concept is also developed in [6], in their article the authors identify that one of the omnipresent needs of this type of system is the need to ensure security against cyber-attacks. And among the variety of attacks they can suffer, malware is the one that poses the greatest risk [20].

3. Malware

One of the most accepted formal definitions is that of [21] which defines malware as *"All software or firmware that is inserted into a system without the user's knowledge, allowing the theft of information, corrupting the functions of the equipment or evading the mechanisms implemented to control access to it"*.

Outside of the academic environment, widely accepted industry definitions of malware can be found mainly in reference guides and glossaries of terms from the National Institute of Standards and Technology (NIST). Malware is defined as "*A program that is inserted into a system, usually covertly, with the intent to compromise the confidentiality, integrity, or availability of the victim's data, applications, or operating system, or to otherwise annoy or disrupt the victim" in* [22]. Malware is considered as *"Software designed and operated by an adversary to violate the security of a computer (includes spyware, viruses, rootkits and Trojans)" in* [23]. While in [24] is defined as *"Software or firmware intended to perform an unauthorized process that will have an adverse impact on the confidentiality, integrity or availability of an information system"*.

The above definitions focus primarily on defining the intentions of the malware authors, however, there are also detractors of such definitions, such as [25], who argue that this approach may be inappropriate because it is impossible to know in advance what intentions an unknown binary code has. They therefore suggest an alternative definition for the term malware that covers the malicious aspects of this type of software: *"Malicious programs are code that runs on a computer system and whose presence or behaviour is unknown to system*

administrators; if system administrators were aware of the code and its behaviour, they would not allow it to be executed. Furthermore, they compromise the confidentiality, integrity or availability of the system by exploiting existing vulnerabilities in a system or by creating new ones".

Perhaps the first part of this definition is too focused on fitting into the business environment, where there are figures who assume different roles in the administration of Information Technology (IT) that may have a broad knowledge of the IT environment and the code that runs on it. But it seems wrong to think that this definition also applies to a particular user in a domestic environment, who generally will not have enough computer knowledge to know the code that the software running on his computer has. In this case, when the user installs a software, he usually does it without thinking about how its code is and what its behaviour is, and the usual criterion for the installation is the trust placed in the source from which he has downloaded it.

However, the second part of the definition which refers to the three pillars of Information Security known as the CID triad, does better capture the essence of what malware does. When a user uses a computer system, he is assuming that only authorized users will have access to the data, that the data will be correct and will not be altered without the owner's consent, and that it will be available for use whenever necessary.

According to the authors the CID triad it´s compromised by exploiting vulnerabilities in a system. Within these vulnerabilities there are two types, the first are widely known by the industry as they may have been included in repositories of vulnerabilities open to the public, while the second are those that pose the greatest risk, as it´s indicated in [26], since they have not yet been publicized and would only be accessible to the researchers who discovered them or, in the case that they have been put up for sale, the buyer.

Although the second part of the definition more closely resembles the reality of malware, it is necessary to point out that not all malware takes advantage of vulnerabilities in a software or system to run, as some types of malware try to trick the user to some extent to get him to run it on his own system voluntarily.

Based on the different definitions seen so far, we propose a new and complete definition of malware, considering it as "*any software that compromises the confidentiality, integrity or availability of a system, using a vulnerability, system failure or by tricking a user with permissions on the system in order to be executed*.

Once the definition of malware has been reviewed, it is necessary to mention that there are different subcategories of malware depending on some common characteristics of the malware [27]. The main ones are:

- Virus: have the ability to duplicate themselves and in turn insert code into executable programs so that it can propagate from one to another [28].
- Worms: once it infects one computer it has the ability to spread to infect other computers and all of them can then be used collectively to carry out denial of service attacks, phishing or information theft [29].
- Ransomware: are capable of encrypting the victim's computer through the use of public and private key infrastructure and thus restrict access to certain types of file by requesting payment of a fee in exchange for keys that allow decryption [30].
- Cryptojacking: uses the computing power of the infected computer to perform crypto mining on behalf of the malware owner and without explicit consent of the owner of the infected computer. The compromised computer may not exhibit any strange behaviour, but the malware will take advantage of its power and processing

consumption to allow the malware owner to profit economically from the crypto currencies obtained through the mining [31].

- Spyware: attempt to spy on the user's activities in order to access sensitive personal information and share it with third parties, usually without the user's knowledge and without altering the behaviour of the computer [32].
- Adware: show the user advertising on their system to make certain advertisers receive a higher flow of visits, without altering the behaviour of the computer [33]. A quantitative study of the relationship between adware and malware is carried out in [34], concluding with a list of adware families that are likely to be malicious, and will therefore be considered a type of malware.
- Trojans: are capable of behaving like legitimate programs while including unknown or unwanted features. This type of software usually impersonates legitimate software to trick the user into installing it and then activates additional functionality to seek out information theft [35].
- Botnet: allow to remotely control a set of infected computers without the user being aware that the computer is infected. This type of malware is generally used in spam or distributed denial of service campaigns [36].
- Rootkit: once successfully installed on the system, it allows to run files remotely, perform an exfiltration of sensitive data, modify system settings and alter the functionality of the security software. Its detection and prevention is very difficult because it tries to hide its presence by making the detection and removal tasks by the security tools more difficult. Therefore, its detection generally depends on manual methods [37].
- Advanced Persistent Threat (APT): specifically designed, generally using "Zero Day" vulnerabilities, to achieve a complex attack carried out by a well-funded and highly capable organization or state. It tries to achieve persistence in the attacked organization during long periods of time, trying to go unnoticed in the systems to carry out the theft of valuable information during the maximum possible time [38].

This classification tries to define the characteristics of each malware according to its type, but it is necessary to point out that in the scientific field there is still no consensus when it comes to classifying and defining the characteristics of the different types of malware and it is an open area for further research.

4. Malware in Cyber-physical Systems

An analysis of the insecurity of network devices published on the Internet is presented in [39], where 540,000 devices (representing 13% of the total number of devices identified) are identified with the default access credentials. These devices included firewalls, routers, IPTV devices, printers and video conferencing systems among others. In addition, it was found that 96% of the vulnerable devices still had them 4 months after their initial identification.

The next article from this author [40] also focuses on this line of work, where a large-scale review of the security of the firmware of 32,000 embedded devices is carried out, applying static code analysis techniques to 1.7 million individual files obtained from this firmware. Also in a later article [41] makes a systematic review of the existing cyber security threats on CCTV and video surveillance devices and the same author [42] makes a dissertation on some aspects or directions that could be used to improve the security in IoT devices.

Table 1. Articles analysing the malware security of cyberphysical system components

Author	Year	Purpose of the study	Devices studied
Cui & Stolfo	2010	Security analysis of network devices published on the Internet	Firewalls, routers, IPTV, printers and videoconference systems
Costin et al.	2014	Review of security in embedded devices firmware	Embedded devices
Costin	2016	Cybersecurity threats on CCTV devices	CCTV and video surveillance
Costin	2018	Proposal for improvements in the safety of embedded devices	Embedded devices

Therefore, it is not surprising that with the importance that CPS are gaining, they represent a new target also for malware designers, who see in them a new way to continue expanding their business.

In [43] there is a review of the state of the art of malware for embedded devices based on Unix systems, paying special attention to the main Botnets that were created with this type of devices between 2006 and 2010. In this same article, the malware called PSYB0T is identified as the first one for embedded devices.

This author had previously published a paper [44] in which an exhaustive analysis is made of the botnet called "Chuck Norris" that had infected a large number of DSL modems and home routers using Linux as an operating system since the end of 2009.

In June 2010 a new malware was detected that was specifically targeted at IoT systems and was named Stuxnet [45]. This malware was specially designed to attack an Iranian nuclear infrastructure in Natanz and the attack destroyed the nuclear centrifuges while causing the monitoring system to show normal behaviour [46].

In 2013, Linux-Darlloz malware [47] appeared, a botnet of IoT devices that took advantage of a vulnerability in PHP present in many of these devices. While in 2015 BASHLITE exploited default passwords and known vulnerabilities in the bash shell to hijack devices running BusyBox [48]. Later the source code was released and by 2016 it is estimated that about 1 million devices were infected, 96% of them being IoT devices.

In 2016, Mirai malware was made public through a series of attacks that affected some of the most heavily trafficked sites on the Internet such as GitHub, Twitter, Reddit and Netflix [49]. After an anonymous source published Mirai's code on GitHub, it was found to be malware specifically designed to target IoT devices and create botnets with them to carry out distributed denial of service attacks [50].

Table 2. Major malware attacks in CPS or IoT environments

Author	Year	Malware type	Malware family	Detection date	Malware objectives
Čeleda et al.	2010	Botnet	Chuck Norris	2009	Routers with Linux
Čeleda et al	2011	Botnet	PSYBOT	2006-2010	Unix embedded devices
Langner	2011	APT	Stuxnet	2010	Iran´s nuclear facilities
Bertino & Islam	2017	Botnet	Linux-Darlloz	2013	Devices exposed to Internet with PHP vulnerabilities
Marzano et al.	2018	Botnet	Bashlite	2015	Devices exposed on the Internet with a vulnerable version of Busybox
Sinanović & Mrdovic	2017	Botnet	Mirai	2016	Malware specifically designed to affect IoT devices

The differences between gaining unauthorized access to a system and causing damage to that system are considered in [51]. In the particular case of CPS, and unlike in an IT environment, an attacker should know how the physical process is controlled and this includes, among other things, equipment failure conditions, process behaviour and signal processing. Specifically, in their work they consider a specific type of malware, namely a denial of service attack, at a specific time could affect the processing of signals from the cyberphysical system to alter a certain value of the process above or below a threshold that would cause a failure in the system.

In [52] the authors even go further as they propose a framework for creating botnets with devices that form the CPS, which allows attacks, both real and simulated, to be carried out on water distribution systems.

A demonstration with a proof of concept a type of denial of service attack that is feasible over CPS is described in [53]. Taking as a reference point the popularity that Deep Neural Networks (DNN) are gaining in recent years. The authors exploit a feature that allows code to be mixed with data in DNN models and can be triggered by a specific physical object without impairing the accuracy of DNN network inference. Through proof of concept, the authors manage to turn a DNN network into an evasive, self-contained malware in a CPS.

So far, the malware attacks that have been seen could have an impact on a system and this could have consequences for people. But an indication of the constant evolution and increased risk that malware could pose to cyberphysical systems can be found in [6]. In the article they point the level of vulnerabilities that modern automotive systems present to attacks. One of the examples they raise is that of an attack, executed under controlled conditions, to a Jeep driving at 70 mph on a highway in St. Louis (USA) where the car was remotely hijacked by the attackers to show how various electronic control units, from the windshield wipers to the braking and engine systems, can be remotely manipulated through the vehicle's built-in cellular connection. Although this attack was executed in a controlled environment, it is claimed that the remote attack of a vehicle is a real threat that could have consequences affecting the lives of the vehicle occupants.

Also in [7], the authors demonstrate a new type of malware that takes advantage of Machine Learning techniques to maximize its impact. In their work, they use the Raven-II surgical robot to exploit its vulnerabilities through intelligent malware that is able to learn and track the movements of the robot's arms and trigger an attack when the robot is in a critical stage of a (hypothetical) surgical procedure.

The last example is the most extreme case of why CPS urgently need to be protected against malicious software. This requires a study of malware so that its identification and detection can be improved and the effects of malware on CPS can be stopped before the consequences are catastrophic.

Table 3. Exploring the impacts of malware on CPS

Author	Year	Malware type	Attack type	Impact
Krotofil et al.	2014	DoS	Denial of service on the signal processing of a CPS	General failure in the physical system due to denial of service at a key point in the process
Antonioli et al.	2018	Botnet	Framework for building botnets with the devices that form a CPS	Attacks, real and simulated, on water distribution CPS
Liu & Wen	2019	DoS	Proof of Concept for Denial of Service on CPS Systems	Turning a deep neural network (DNN) into an elusive, self-contained malware for CPS

Dibaji et al.	2019	APT	Proof of concept for remote access to a vehicle on the road	Full control over the vehicle and access to critical vehicle systems such as the engine or brakes
Chung et al.	2019	APT	Malware that uses machine learning techniques to learn the movements of a surgical robot	Full control over the surgical robot in critical phases of an operation

5. Malware Analysis Methods

Within the malware analysis there are several working approaches: static, dynamic, hybrid and RAM memory analysis.

5.1 Static Analysis

This type of analysis seeks to evaluate, without running the malware, those components that are fixed in a malicious software and try to infer properties or behaviours from it simply by observation. As indicated in [54] within the static analysis there are several sub-techniques such as file format inspection, text string extraction, scanning by different antivirus engines or reverse engineering.

According to research by [54], static analysis has limitations that make it insufficient in itself to identify malware, so it is necessary to complement it with other techniques, such as dynamic or hybrid analysis.

5.2 Dynamic Analysis

According to [55], dynamic analysis studies, among others: file activities (e.g. which files were created), registry activities (e.g. which registry values were added or modified), network activities (e.g. which files were downloaded or which information was exchanged and process activities (e.g. which processes were launched or executed).

Within the dynamic analysis, as indicated [56], there are several sub-techniques such as, among others, function call monitoring, function parameter analysis, information flow tracking, instruction tracing and start points.

Dynamic analysis is more effective in revealing the natural behaviour of malware, which is generally difficult to achieve by static analysis. According to [31] it is also more effective than static in detecting unknown malware. This makes malware obfuscated by encryption, compression, metamorphic or polymorphic techniques unable to evade dynamic detection measures [26].

5.3 Memory Analysis

Since any process or object in an Operating System will have to go through its RAM at some point, some researchers have begun to consider RAM as the ideal place to conduct their malware analysis [57].

5.4 Malware Analysis in Cyber-physical Systems

Everything mentioned so far refers to generic malware analysis. But the differences between different operating systems make it necessary to devote a particular section to the analysis of malware on UNIX/LINUX-based systems [58], which are the most commonly used for cyberphysical systems.

The challenges of performing malware analysis for Linux environments described above, while providing a higher barrier to entry than other environments, have not prevented solutions that attempt to facilitate the automation of malware analysis for these environments, as well as the emergence of specific malware analysis features for IoT devices [59].

6. Malware Detection

As in the case of malware analysis, there are different ways of detecting it, the two main ones being signature-based detection and behaviour-based detection.

6.1 Signature based

The signature-based detection method uniquely identifies a piece of malware through the use of mathematical functions applied to a malware binary [60]. Using this method any file on a system can be analysed, with a low false positive rate [61], to see if it matches any of the known signatures. This mode of operation is mainly used by anti-viruses and is a fast and effective as long as the signatures for the malware have been previously identified and added to the signatures database [62].

However, malware developers have devised tactics to evade signature-based detection [63]. These measures cause the resulting binary to be modified without altering its operation and thus changing the signature of the binary. Additionally, if the file being analysed is a new sample, it has not yet been registered in the signature database and cannot be recognized as malware.

6.2 Behaviour based

The behaviour-based detection method arises to solve the problems that have been mentioned for the signature-based detection method. In this case, in an initial training phase, the aim is to study the activities performed by the malware. In a later phase, the files are analysed for behavioural patterns that match known patterns from the initial detection [60]. This method can detect obfuscated files or new samples as malware if their behaviour matches a pattern that has already been analysed.

Although the mode of operation of this method seems promising, it must be taken into account that as a disadvantage it has a higher number of false positives and a longer detection time as indicated [61]. In addition, it requires a greater computing effort, which may cause it to be incompatible with some resource-constrained environments.

The behaviour-based approach is very focused on machine learning, where models can be created that learn from behaviour to refine themselves and increasingly improve malware detection.

7. Application of Machine Learning to the Detection of Malware in Cyberphysical Systems

According to [64] the rapid evolution in the last decade of cybersecurity, where the number of threats is increasing on the same way as the efforts of malware designers to stay ahead of detection systems, is making traditional cybersecurity solutions irrelevant in detecting and mitigating attacks. In that sense, advances in Artificial Intelligence, more specifically in the fields of Machine Learning and Deep Learning, show promise in helping to counteract the ever-changing threats of these environments.

The increasing volume of new malware samples being made public on a daily basis requires a new approach to automating malware detection. This section will review the state of the art of malware detection in cyberphysical environments using machine learning techniques. Which is, according to [65] where most of the research work is currently concentrated.

7.1 State of the Art in Detecting Malware in Cyber-physical Systems Using Machine Learning Techniques

The work conducted in [66], although without explicit reference to Cyber-physical Systems, revolves around the application of the SVM (Support Vector Machine) for the detection of malware in Android devices for reliable IoT services. For this purpose, the performance of SVM was compared with other Machine Learning classifiers and showed, through experimental validation, how SVM improves performance to the rest of classifiers.

The paper by [67] is one of the first identified in the literature talking specifically about the detection of attacks in CPS, although it did not focus only on malware and had a more holistic approach to the possible attacks that could suffer this type of environment. The authors propose an approach that uses behavior-based Machine Learning for intrusion detection and the use of a specific test bench for CPS that replicates the specific components of a water treatment facility to obtain data for training of a supervised model. The results of the work suggest the effectiveness of the proposed approach by validating the results with the proposed dataset, although there are certain limitations. The validation of results is especially relevant, ideally with a broader dataset that includes information from other types of CPS.

The detection of malware in home routers, using Machine Learning is discussed in [68]. The article compares the effectiveness of three different malware detection algorithms based on the behaviour in home routers by observing the system kernel calls in these devices.

In the work of [69] is proposed the detection of anomalies based on different algorithms of Machine Learning to find malicious traffic using as dataset communication data from industrial environments (Modbus/TCP) synthetically generated that allowed learning supervised by the model. The Machine Learning algorithms used at work were Support Vector Machine (SVM), Random Forest, k-nearest neighbour and k-means clustering. The results obtained show that SVM and Random Forest behaved well with all the data sets used, while k-nearest neighbour and k-means did not provide the expected results. As future work, the authors propose several possibilities to extend the methods proposed. For example, obtaining information from different sources to combine it with the synthetic data and improve the results. In this sense, one of the most important needs is the generation of data with attacks that are specific to industrial applications and specifically in this particular case to Modbus.

Also, in 2018 is published [70], this paper proposes a model supported by cloud environments for malware detection. Firstly, building a malware detection model based on SVM with data exchange in the cloud security platform. The number of infected nodes is then calculated, based on the transmission attributes of the WMS. On this basis, a malware detection and removal algorithm are presented based on the modified model and the calculation of optimal strategies. The numerical results and comparisons obtained in the paper show that the proposed algorithm can increase the usefulness of the WMS efficiently and effectively.

Another paper from 2018 [71], in which the authors propose an advanced multi-level malware detection system for virtualized machines in CPS. This system is also identified by its acronym AMMDS (Automated Multilevel Malware Detection System) and takes advantage of memory forensic analysis (MFA) techniques to predict the early symptoms of malware execution by detecting hidden stealthy processes in a host operating system. More specifically,

the AMMDS system detects and classifies malicious executables that are running. The system proposed in the article has been evaluated against a large dataset combining malware and benign executables and the results of the evaluation reached 100% accuracy and zero false positive rate (FPR) in the classification of unknown malware with a maximum performance overhead of 5.8%. Future work pending in this article is the validation of the accuracy of the results with other data sets, to confirm the ratios obtained by the authors.

The future problem of the estimated 30 trillion IoT devices worldwide by 2020 and the new trends in malware such as crytojackers is identified in [72]. Although industrial systems are generally systems with limited resources, the authors consider that the low security that this type of devices usually have, results in a significant risk due to the large park of devices that could be subject to this type of malware. To avoid this situation, the authors propose a lightweight cryptojacking classifier model for the detection of this type of malware in CPS based on the dendritic cell algorithm.

That same year, a paper was published [73], in which the authors, alerted by the growing trend in CPS to replace personal computers with mobile devices and making those key parts of these systems, decided to implement a model for the detection of malware in mobile applications developed for those environments. By integrating a semi-supervised approach and Deep Learning and combining the benefits of both methods, they manage to overcome the limitations posed by supervised systems for this type of environment. Although the proposed model provides satisfactory performance for the authors, there is still room for improvement in terms of accuracy level. As future work, it is proposed to train the model using static and dynamic capabilities and to use a more recent dataset containing tagged and untagged data so that the model will be able to detect new or unknown malware.

Another article this year [74] is focused on the detection of one of the most widespread and dangerous malware elements, botnets and their relationship to the IoT. In it, the authors propose a botnet detection methodology based on Deep Learning applied to a new and specific SDN (Software-Defined Networking) dataset that achieves an accuracy in the classification of more than 97%. To achieve these results, the authors have used TensorFlow, the free software platform for Machine Learning and some high-level APIs specific to this platform. As a future work, it is possible to reproduce their results and apply them to other datasets to validate the accuracy of the classifier.

Also, in 2019 it is published [75] which considers the independent component analysis (ICA), the global K-means clustering and the exponentially weighted multivariate moving average (MEWMA) to extract the behavioural indicators that cluster the malware. A monitored SVM-based system is used for the detection system, allowing malware behaviour patterns to be updated automatically. Performance comparison of the experimental results summarizes that semi-supervised models can detect more accurately than existing supervised models, where accuracies are increased to 100% when using SVM and semi-supervised models based on the random forest.

Following the proliferation of scientific articles on the topic studied in this article, in 2019 [76] proposed a work that helps predict malware attacks in CPS and more specifically in the supply chain. The authors use Machine Learning techniques, such as Decision Tree and SVM algorithms, and a dataset obtained from Microsoft Malware Prediction for the validation of their results, which show that these algorithms can be used in the supply chain to detect and predict future trends in malware attacks.

In [77] the work is focused on the application of machine learning in the detection of intrusions in aerospace cyberphysical systems. Although the work did not focus only on malware detection, an interesting fact is extracted from its conclusions, which is that the

application of machine learning in detecting attacks in cyberphysical systems requires the availability of data on previous attacks to this type of environment, which is currently a major limitation.

Table 4. Analysis of the main results and future work identified in the state of the art

Author	Year	Environment	Classification algorithm	Results	Future work
Ham et al.	2014	IoT	SVM	SVM improves the performance of other classifiers in detecting malware for Android devices in IoT environments.	-
Junejo & Goh	2016	CPS	Several	The proposed method succeeds in identifying the occurrence of an attack and the specific type of attack in a water treatment environment.	Extend the dataset used
An et al.	2018	IoT	SVM n-grams PCA	SVMs and n-grams behave better with small fragments. All three algorithms have 100% effectiveness and 0 false positives with large fragments	Extend the work with smaller packages
Anton et al	2018	CPS	SVM Random forest k-nearest neighbour k-means clustering	With SVM and Random forest you get the best detection results	Extend the dataset used
Zhou & Yu	2018	IoT	SVM	SVM shows positive results to increase the usefulness of WMS	-
Ajay Kumara & Jaidhar,	2018	CPS	AMMDS	AMMDS achieves 100% results with zero false positives and 5.8% CPS performance overhead	Extend the dataset used
Ahmad et al.	2019	CPS	DCA	Lightweight model to detect and classify Cryptojackers in CPS	-
Sharmeen et al	2019	IoT	Deep Learning	Improved performance of supervised models through the Deep Learning approach and a semi-supervised model for malware detection on Android devices for IoT environments	Extend the datasets used with tagged and untagged data
Letteri et al.	2019	IoT	Deep Learning	Applying Deep Learning with TensorFlow to discover botnets in IoT environments achieves over 97% detection accuracy	Extend the datasets used with more current data
Huda et al.	2019	CPS	ICA k-nearest MEWMA Random Forest SVM	SVM and Random Forest achieve results of up to 100% detection accuracy	-
Yeboah-Ofori & Boachie	2019	CPS	Decision Tree SVM	Success using decision tree and SVM algorithms in detecting malware in the supply chain	-
Maleh	2020	CPS	-	The application of Machine Learning in the detection of attacks on CPS requires the availability of data on previous attacks on this type of environment, which is currently a major limitation	Extend the availability of datasets

Table 4 summarizes the results obtained in the articles analysed, as well as the main lines of work identified by their authors.

8. Conclusion

This paper presents a review of the state of the art in the detection of malware in Cyber-physical systems using different machine learning algorithms. One of the main challenges while reviewing the scientific literature in this field has been to identify what is considered to be a CPS and its relationship to the Internet of Things. Currently in the scientific literature up to four different answers to this question can be found. This makes it necessary to study in depth which are the characteristics and properties that Cyber-physical systems should have and if these are equivalent, complementary or different from those of the IoT systems. Answering these questions could serve to link two fields of study that are not completely connected at the moment and that could possibly benefit mutually from the advances in scientific knowledge obtained in the other.

The heterogeneity of the Cyber-physical systems is one of the main problems faced by the researchers of the papers consulted. This heterogeneity, together with the lack of data on past malware attacks in this type of environment, means that one of the future works that commonly appears in many of the papers is the expansion of the datasets used. In some cases, these datasets correspond to specific data from specific subsystems, making it necessary to expand the data. While in other cases, the datasets are old, and researchers claim a lack of more recent data to be able to expand their research.

Just as there is a deficit in the datasets available for cyberphysical systems, there is also no systematic review of the different ways malware affects cyberphysical systems, and this is a key aspect in defining the security measures to be implemented in these systems to protect them from malware.

Based on the work reviewed in this paper, it can be concluded that the greatest risk to cyberphysical systems is being attacked by a malware to turn them into botnets. After botnets, denial of service and advanced persistent threats are the two other types of malware that have also been identified, while the other types of malware have not been mentioned in any of the articles reviewed. As future work, it is proposed to further investigate the detection of botnets in Cyber-physical systems, as this seems to be the trend that malware developers are moving towards.

In the review of the state of the art carried out in this article, no papers have been identified that mentioned the application of machine learning algorithms in the analysis of malware for cyberphysical systems. Similarly, no articles were found that referred to signature-based malware detection through the application of machine learning algorithms.

All articles reviewed talk about behaviour-based malware detection, so this would confirm that this is the line with the most future work.

Within the Machine Learning techniques and how they are applied to malware detection in Cyber-physical systems, the common denominator in six of the twelve articles reviewed is the identification of SVM (Support Vector Machine) as the classification algorithm that gives the best results. From the results of the work analysed, it cannot be concluded whether Random Forest, n-grams or decision trees are effective classification algorithms in detecting malware in cyberphysical systems. Although they have shown promising results in some of the papers analysed, the shortcomings in the datasets used and the heterogeneity of the systems do not make it possible to conclude whether the same results could be obtained in any Cyber-physical systems.

Additionally, the classification of malware is an important area for future study, since the scientific community has not yet reached a consensus on numerous issues such as the number of existing malware families, what characteristics the members of these families must meet, and the criteria for grouping them with scientific rigor. Achieving this consensus is important in order to have clear groupings that facilitate homogeneity in the results of scientific studies focused on the analysis or detection of malware.

References

[1] A. Lakhotia, V. Notani, and C. LeDoux, "Malware Economics and its Implication to Anti-Malware Situational Awareness," in *Proc. of 2018 International Conference On Cyber Situational Awareness, Data Analytics And Assessment*, pp. 1-8, June 2018. Article (CrossRef Link)

[2] C. S. Wickramasinghe, D. L. Marino, K. Amarasinghe, and M. Manic, "Generalization of Deep Learning for Cyber-Physical System Security: A Survey," in *Proc. of the 44th Annual Conference of IEEE Industrial Electronics Society*, pp. 745-751, 2018. Article (CrossRef Link)

[3] R. Mitchell and I. R. Chen, "A survey of intrusion detection techniques for cyber-physical systems," *ACM Computing Surveys*, vol. 46, no. 4, pp. 55:1-55:29, Mar. 2014. Article (CrossRef Link)

[4] Y. Pan, J. White, D. Schimidt, A. Elhabash, L. Sturm, J. Camelio, and C. Williams, "Taxonomies for Reasoning About Cyber-physical Attacks in IoT-based Manufacturing Systems," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 3, p. 45, 2017. Article (CrossRef Link)

[5] N. Nikolakis, V. Maratos, and S. Makris, "A cyber physical system (CPS) approach for safe human-robot collaboration in a shared workplace," *Robotics and Computer Integrated Manufacturing*, vol. 56, pp. 233-243, Apr. 2019. Article (CrossRef Link)

[6] S. M. Dibaji, M. Pirani, D. B. Flamholz, A. M. Annaswamy, K. H. Johansson, and A. Chakrabortty, "A systems and control perspective of CPS security," *Annual Reviews in Control*, vol. 47, pp. 394-411, Jan. 2019. Article (CrossRef Link)

[7] K. Chung, X. Li, R. K. Lyer, and T. Kesavadas, "Smart Malware that Uses Leaked Control Data of Robotic Applications: The Case of Raven-II Surgical Robots," in *Proc. of the 22nd International Symposium on Research in Attacks, Intrusions and Defenses*, pp. 337-351, 2019. Article (CrossRef Link)

[8] C. Greer, M. J. Burns, D. A. Wollman, and E. R. Griffor, "Cyber-Physical Systems and Internet of Things," *National Institute of Standards and Technology*, 2019. Article (CrossRef Link)

[9] S. Madakam, R. Ramaswamy, and S. Tripathi, "Internet of Things (IoT): A Literature Review," *Journal of Computer and Communications*, vol. 3, no. 5, May 2015. Article (CrossRef Link)

[10] R. Minerva, A. Biru, and D. Rotondi, "Towards a Definition of the Internet of Things (IoT)," *IEEE Internet of Tnings*, no. 1, pp. 1-86, May 2015. Article (CrossRef Link)

[11] H.-D. Ma, "Internet of Things: Objectives and Scientific Challenges," *Journal of Computer Science and Technology*, vol. 26, no. 6, pp. 919-924, Nov. 2011. Article (CrossRef Link)

[12] V. Gunes, S. Peter, T. Givargis, and F. Vahid, "A Survey on Concepts, Applications, and Challenges in Cyber-Physical Systems," *KSII Transactions on Internet and Information Systems*, vol. 8, no. 12, pp. 4242-4268, Dec. 2014. Article (CrossRef Link)

[13] I. Stojmenovic and F. Zhang, "Inaugural issue of "cyber-physical systems," *Cyber-Physical Systems*, vol. 1, no. 1, pp. 1-4, Jan. 2015. Article (CrossRef Link)

[14] A. Yeboah-ofori, J. D. Abdulai, and F. Katsriku, "Cybercrime and Risks for Cyber Physical Systems: A Review," Apr. 2018. Article (CrossRef Link)

[15] National Academy of Science and Engineering, Germany, "Acatech Position Paper: Cyber-Physical Systems Driving Force for Innovation in Mobility, Health, Energy and Production," *National Academy of Science and Engineering*, Germany, 2019. Article (CrossRef Link)

[16] O. Vermesan and P. Friess, Internet of Things: Converging Technologies for Smart Environments and Integrated Ecosystems, River Publishers, Denmark, 2013.

[17] S. Tweneboah-Koduah, K. E. Skouby, and R. Tadayoni, "Cyber Security Threats to IoT Applications and Service Domains," *Wireless Personal Communications*, vol. 95, no. 1, pp. 169-185, July 2017. Article (CrossRef Link)
[18] A. Humayed, J. Lin, F. Li, and B. Luo, "Cyber-Physical Systems Security-A Survey," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1802-1831, Dec. 2017. Article (CrossRef Link)
[19] Y. Ashibani and Q. H. Mahmoud, "Cyber physical systems security: Analysis, challenges and solutions," *Computer and Security*, vol. 68, pp. 81-97, July 2017. Article (CrossRef Link)
[20] R. Sihwail, K. Omar, and K. A. Zainol Ariffin, "A survey on malware analysis techniques: static, dynamic, hybrid and memory analysis," vol. 8, pp. 1662-1671, Jan. 2018. Article (CrossRef Link)
[21] M. Tracy, W. Jansen, K. Scarfone, and J. Butterfield, "Guidelines on Electronic Mail Security," *NIST Special Publication*, vol. 2, Feb. 2007. Article (CrossRef Link)
[22] E. B. Barker, M. Smid, and D. Branstad, "A Profile for U. S. Federal Cryptographic Key Management Systems," *National Institute of Standards and Technology (NIST SP 800-152)*, Oct. 2015. Article (CrossRef Link)
[23] C. Paulsen and P. Toth, "Small Business Information Security: The Fundamentals," *National Institute of Standards and Technology (NIST IR 7621r1*), Oct. 2016. Article (CrossRef Link)
[24] O. Or-Meir, N. Nissim, Y. Elovici, and L. Rokach, "Dynamic Malware Analysis in the Modern Era-A State of the Art Survey," *ACM Computing Survey*, vol. 52, no. 5, Sep. 2019. Article (CrossRef Link)
[25] L. Bilge and T. Dumitraș, "Before we knew it: an empirical study of zero-day attacks in the real world," in *Proc. of the 2012 ACM Conference on Computer and Communications Security*, pp. 833-844, Oct. 2012. Article (CrossRef Link)
[26] J. B. Higuera, C. A. Aramburu, J. R. Bermejo Higuera, M. A. Sicilia Urban, and J. A. Sicilia Montalvo, "Systematic Approach to Malware Analysis (SAMA)," *Applied Science*, vol. 10, no. 4, Jan. 2020. Article (CrossRef Link)
[27] R. Sharp, "An Introduction to Malware," DTU Library, 2017.
[28] S. Bardhan, D. Montgomery, J. Filliben, and A. Heckert, "A general methodology for deriving network propagation models of computer worms," *National Institute of Standards and Technology*, Feb. 2019. Article (CrossRef Link)
[29] N. Scaife, H. Carter, P. Traynor, and K. Butler, "CryptoLock (and Drop It): Stopping Ransomware Attacks on User Data," in *Proc. of IEEE 36th International Conference on Distributed Computing Systems*, pp. 303-312, June 2016. Article (CrossRef Link)
[30] K. Jayasinghe and G. Poravi, "A Survey of Attack Instances of Cryptojacking Targeting Cloud Infrastructure," in *Proc. of 2nd Asia Pacific Information Technology Conference*, pp. 100-107, 2020. Article (CrossRef Link)
[31] Y. Ye, T. Li, D. Adjeroh, and S. S. Iyengar, "A Survey on Malware Detection Using Data Mining Techniques," *ACM Computing Surveys*, vol. 50, no. 3, June 2017. Article (CrossRef Link)
[32] J. Landage and M. P. Wankhade, "Malware and Malware Detection Techniques: A Survey," *International Journal of Engineering Research and Technology*, 2013. Article (CrossRef Link)
[33] J. Gao, L. Li, P. Kong, T. F. Bissyandé, and J. Klein, "Should You Consider Adware as Malware in Your Study?," in *Proc. of IEEE 26th International Conference on Software Analysis, Evolution and Reengineering (SANER)*, pp. 604-608, 2019. Article (CrossRef Link)
[34] I. Memon, R. A. Shaikh, H. Fazal, H. Tunio, and Q. A. Arain, "The World of Hacking: A Survey," *University Sindh Journal of Information and Communication Technology*, vol. 4, no. 1, Mar. 2020. Article (CrossRef Link)
[35] M. Chowdhury, A. Rahman, and R. Islam, "Malware Analysis and Detection Using Data Mining and Machine Learning Classification," in *Proc. of International Conference on Applications and Techniques in Cyber Security and Intelligence*, pp. 266-274, 2018. Article (CrossRef Link)
[36] A. Zaki and B. Humphrey, "Unveiling the kernel: rootkit discovery using selective automated kernel memory differencing," in *Proc. of Virus Bulletin Conference*, pp. 239-256, 2014. Article (CrossRef Link)

[37] W. Niu, X. Zhang, G. Yang, J. Zhu, and Z. Ren, "Identifying APT Malware Domain Based on Mobile DNS Logging," *Mathematical Problems in Engineering*, pp. 1-9, 2017. Article (CrossRef Link)
[38] A. Cui and S. J. Stolfo, "A quantitative analysis of the insecurity of embedded network devices: results of a wide-area scan," in *Proc. of the 26th Annual Computer Security Applications Conference*, pp. 97-106, 2010. Article (CrossRef Link)
[39] A. Costin, J. Zaddach, A. Francillon, and D. Balzarotti, "A Large-Scale Analysis of the Security of Embedded Firmwares," in *Proc. of the 23rd USENIX Security Symposium*, pp. 95-110, 2014. Article (CrossRef Link)
[40] A. Costin, "Security of CCTV and Video Surveillance Systems: Threats, Vulnerabilities, Attacks, and Mitigations," in *Proc. of the 6th International Workshop on Trustworthy Embedded Devices*, pp. 45-54, 2016. Article (CrossRef Link)
[41] A. Costin, "IoT/Embedded vs. Security: Learn from the Past, Apply to the Present, Prepare for the Future," in *Proc. of the 22nd Conference of Open Innovations Association*, pp. 412-414, 2018. Article (CrossRef Link)
[42] P. Čeleda, R. Krejč, and V. Krmíček, "Revealing Botnets Using Network Traffic Statistics," *Security Protection of Information*, pp. 7-16, 2011. Article (CrossRef Link)
[43] P. Čeleda, R. Krejčí, J. Vykopal, and M. Drašar, "Embedded Malware - An Analysis of the Chuck Norris Botnet," in *Proc. of 2010 European Conference on Computer Network Défense*, pp. 3-10, 2010. Article (CrossRef Link)
[44] R. Langner, "Stuxnet: Dissecting a Cyberwarfare Weapon," *IEEE Security and Privacy*, vol. 9, no. 3, pp. 49-51, May 2011. Article (CrossRef Link)
[45] J. P. Farwell and R. Rohozinski, "Stuxnet and the Future of Cyber War," *Survival*, vol. 53, no. 1, pp. 23-40, Feb. 2011. Article (CrossRef Link)
[46] E. Bertino and N. Islam, "Botnets and Internet of Things Security," *Computer*, vol. 50, no. 2, pp. 76-79, Feb. 2017. Article (CrossRef Link)
[47] A. Marzano, D. Alexander, O. Fonseca, E. Fazzion, C. Hoepers, K. S. Jessen, M. Chaves, I. Cunha, D. Guedes, and W. Meira, "The Evolution of Bashlite and Mirai IoT Botnets," in *Proc. of IEEE Symposium on Computers and Communications (ISCC)*, pp. 00813-00818, 2018. Article (CrossRef Link)
[48] H. Sinanović and S. Mrdovic, "Analysis of Mirai malicious software," in *Proc. of the 25th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pp. 1-5, 2017. Article (CrossRef Link)
[49] G. Kambourakis, C. Kolias, and A. Stavrou, "The Mirai botnet and the IoT Zombie Armies - IEEE Conference Publication," in *Proc. of IEEE Military Communications Conference*, pp. 267-272, 2017. Article (CrossRef Link)
[50] M. Krotofil, A. A. Cárdenas, B. Manning, and J. Larsen, "CPS: driving cyber-physical systems to unsafe operating conditions by timing DoS attacks on sensor signals," in *Proc. of the 30th Annual Computer Security Applications Conference*, pp. 146-155, 2014. Article (CrossRef Link)
[51] D. Antonioli, G. Bernieri, and N. O. Tippenhauer, "Taking Control: Design and Implementation of Botnets for Cyber-Physical Attacks with CPSBot," *ArXiv180200152 Cs*, Jan. 2018. Article (CrossRef Link)
[52] T. Liu and W. Wen, "Deep-evasion: Turn deep neural network into evasive self-contained cyber-physical malware: poster," in *Proc. of the 12th Conference on Security and Privacy in Wireless and Mobile Networks*, pp. 320-321, 2019. Article (CrossRef Link)
[53] H. V. Nath and B. M. Mehtre, "Static Malware Analysis Using Machine Learning Methods," in *Proc. of International Conference on Security in Computer Networks and Distributed Systems*, pp. 440-450, 2014. Article (CrossRef Link)
[54] A. Moser, C. Kruegel, and E. Kirda, "Limits of Static Analysis for Malware Detection," in *Proc. of the 23rd Annual Computer Security Applications Conference (ACSAC 2007)*, pp. 421-430, 2017. Article (CrossRef Link)

[55] U. Bayer, E. Kirda, and C. Kruegel, "Improving the efficiency of dynamic malware analysis," in *Proc. of the 2010 ACM Symposium on Applied Computing*, pp. 1871-1878, 2010. Article (CrossRef Link)
[56] M. Egele, S. Theodoor, K. Engin, and C. Kruegel, "A survey on automated dynamic malware-analysis techniques and tools," *ACM Computing Survey*, vol. 44, no. 2, Mar. 2008. Article (CrossRef Link)
[57] T. Teller and A. Hayon, "Enhancing Automated Malware Analysis Machines with Memory Analysis," *Teller Enhancing AM*, pp. 1-5, 2014. Article (CrossRef Link)
[58] E. Cozzi, M. Graziano, Y. Fratantonio, and D. Balzarotti, "Understanding Linux Malware," in *Proc. of IEEE Symposium on Security and Privacy (SP)*, pp. 161-175, May 2018. Article (CrossRef Link)
[59] D. Uhricek, "LiSa – Multiplatform Linux Sandbox for Analyzing IoT Malware," *Excel FIT*, pp. 1-6, 2019. Article (CrossRef Link)
[60] A. Damodaran, F. D. Troia, C. A. Visaggio, T. H. Austin, and M. Stamp, "A comparison of static, dynamic, and hybrid analysis for malware detection," *Journal of Computer Virology and Hacking Techniques*, vol. 13, no. 1, pp. 1-12, Feb. 2017. Article (CrossRef Link)
[61] Z. Bazrafshan, H. Hashemi, S. M. H. Fard, and A. Hamzeh, "A survey on heuristic malware detection techniques," in *Proc. of the 5th Conference on Information and Knowledge Technology*, pp. 113-120, May 2013. Article (CrossRef Link)
[62] M. Alazab, S. Venkataraman, and P. Watters, "Towards Understanding Malware Behaviour by the Extraction of API Calls," in *Proc. of 2010 Second Cybercrime and Trustworthy Computing Workshop*, pp. 52-59, July 2010. Article (CrossRef Link)
[63] I. You and K. Yim, "Malware Obfuscation Techniques: A Brief Survey," in *Proc. of International Conference on Broadband, Wireless Computing, Communication and Applications*, pp. 297-300, Nov. 2010. Article (CrossRef Link)
[64] S. Zeadally, E. Adi, Z. Baig, and I. A. Khan, "Harnessing Artificial Intelligence Capabilities to Improve Cybersecurity," *IEEE Access*, vol. 8, pp. 23817-23837, 2020. Article (CrossRef Link)
[65] D. Ucci, L. Aniello, and R. Baldoni, "Survey of machine learning techniques for malware analysis," *Computers and Security*, vol. 81, pp. 123-147, Mar. 2019. Article (CrossRef Link)
[66] H. S. Ham, H. H. Kim, M. S. Kim, and M. J. Choi, "Linear SVM-Based Android Malware Detection for Reliable IoT Services," *Journal of Applied Mathematics*, Sep. 03, 2014. Article (CrossRef Link)
[67] K. N. Junejo and J. Goh, "Behaviour-Based Attack Detection and Classification in Cyber Physical Systems Using Machine Learning," in *Proc. of the 2nd ACM International Workshop on Cyber-Physical System Security*, pp. 34-43, May 2016. Article (CrossRef Link)
[68] N. An, A. Duff, G. Naik, M. Faloutsos, S. Weber, and S. Mancoridis, "Behavioral anomaly detection of malware on home routers," in *Proc. of the 12th International Conference on Malicious and Unwanted Software (MALWARE)*, pp. 47-54, Oct. 2017. Article (CrossRef Link)
[69] S. D. Anton, S. Kanoor, D. Fraunholz, and H. D. Schotten, "Evaluation of Machine Learning-based Anomaly Detection Algorithms on an Industrial Modbus/TCP Data Set," in *Proc. of the 13th International Conference on Availability, Reliability and Security*, pp. 1-9, 2018. Article (CrossRef Link)
[70] W. Zhou and B. Yu, "A cloud-assisted malware detection and suppression framework for wireless multimedia system in IoT based on dynamic differential game," *China Communications*, vol. 15, no. 2, pp. 209-223, Feb. 2018. Article (CrossRef Link)
[71] M. Ajay Kumara and C. Jaidhar, "Automated multi-level malware detection system based on reconstructed semantic view of executables using machine learning techniques at VMM," *Future Generation Computer Systems*, vol. 79, pp. 431-446, Feb. 2018. Article (CrossRef Link)
[72] A. Ahmad, W. Shafiuddin, M. N. Kama, and M. M. Saudi, "A New Cryptojacking Malware Classifier Model Based on Dendritic Cell Algorithm," in *Proc. of the 3rd International Conference on Vision, Image and Signal Processing*, vol. 84, pp. 1-5, 2019. Article (CrossRef Link)

[73] S. Sharmeen, S. Huda, and J. Abawajy, "Identifying Malware on Cyber Physical Systems by incorporating Semi-Supervised Approach and Deep Learning," in *Proc. of IOP Conference on Earth Environmental Science*, vol. 322, Sep. 2019. Article (CrossRef Link)

[74] I. Letteri, G. D. Penna, and G. D. Gasperis, "Security in the internet of things: botnet detection in software-defined networks by deep learning techniques," *International Journal of High Performance and Networking*, vol. 15, no. 3, 2019. Article (CrossRef Link)

[75] S. Huda, J. Abawajy, B. Al-Rubaie, L. Pan, and M. M. Hassan, "Automatic extraction and integration of behavioural indicators of malware for protection of cyber–physical networks," *Future Generation Computing Systems*, vol. 101, pp. 1247-258, Dec. 2019. Article (CrossRef Link)

[76] A. Yeboah-Ofori and C. Boachie, "Malware Attack Predictive Analytics in a Cyber Supply Chain Context Using Machine Learning," in *Proc. of 2019 International Conference on Cyber Security and Internet of Things (ICSIoT)*, pp. 66-73, May 2019. Article (CrossRef Link)

[77] Y. Maleh, "Machine Learning Techniques for IoT Intrusions Detection in Aerospace Cyber-Physical Systems," *Machine Learning and Data Mining in Aerospace Technology*, vol. 836, pp. 205-232, 2020. Article (CrossRef Link)

Francisco Montes García is a PhD student at Universidad Internacional de La Rioja (UNIR), in Spain. He graduated in 2009 in Telematics at Universidad de Oviedo and received a master's degree in Cyber Security in 2016 from the Universidad Internacional de La Rioja (UNIR). His research interests include cyber physical systems, Internet of Things, malware analysis and classification and machine learning.

Javier Bermejo Higuera received the B.S. degree from Alcala University and his Ph.D. degree from Army Polytechnic School. Currently, he is Professor in the Escuela Superior de Ingeniería y Tecnología at Universidad Internacional de La Rioja (UNIR). He is an author of several publications His research interests include the fields of software security, cybersecurity and malware analysis.

Luis Enrique Sánchez Crespo holds a PhD in Computer Science from the University of Castilla-La Mancha (Spain), a MSc in Computer Science from the Polytechnic University of Madrid (Spain). He is Certified Information System Auditor by ISACA and Leader Auditor of ISO27001 by IRCA. He is Professor at the University International of Rioja and University of Castilla-la Mancha. He participates at the GSyA research group of the Department of Information Technologies and Systems at the Castilla-La Mancha University.

Juan Ramón Bermejo Higuera is currently chief of cybersecurity unit at National Institute of Aerospace techniques of Spain. He is a professor of applications security in the International University of La Rioja and he also is an associate professor of Ciberdefence Master Science degree in the Alcala de Henares University. He received the M.Sc in computer engineering degree from Distance Education National University of Spain in 1998 and the Ph.D degree from Distance Education National University at 2014. He is an author of several publications. His research interests include cybersecurity and cyber defense.

Juan Antonio Sicilia Montalvo received the B.S. degree and his Ph.D. degree from Universidad de Zaragoza. Currently, he is Professor in the Escuela Superior de Ingeniería y Tecnología at Universidad Internacional de La Rioja (UNIR). His research interests include combinatorial optimization, computer security, software development based on mathematical algorithms, numerical methods and heuristic techniques for solving engineering problems.