WWW.IJRPR.COM

**International Journal of Research Publication and Reviews**

Journal homepage: www.ijrpr.com ISSN 2582-7421

# A Novel Defense System in Cyber–Physical Systems Security Survey

***Mr. A Joshua Issac[1*], M A Reetha Jeyarani[2],C Muthukumaran[3], P B Aravind Prasad[4], Mr R Roshan Joshua[5]***

[1*] Research Scholar, Department of Computer Science & Engineering, SRM Institute of Science & Technology, Trichy, Tamil Nadu.
[2] Research Scholar, Department of Computer Science & Engineering, SRM Institute of Science & Technology, Trichy, Tamil Nadu.
[3] Assistant Professor, Department of Artificial Intelligence and Data Science, K Ramakrishnan College of Technology, Trichy, Tamil Nadu..
[4] Assistant Professor, Department of Artificial Intelligence and Data Science, K Ramakrishnan College of Technology, Trichy, Tamil Nadu.
[5]Assistant Professor, Department of Artificial Intelligence and Data Science, K Ramakrishnan College of Technology, Trichy, Tamil Nadu.
[1*]ad4640@srmist.ed.in, [2] rm7689@srmist.edu.in, [3]muthucg1984@gmail.com, [4]aravindprasadpb.ai@krct.ac.in,[5]roshanjoshuar.ai@krct.ac.in

**ABSTRACT—**

Cyber–physical systems (CPSs) are new types of intelligent systems that integrate computing, control, and com- munication technologies, bridging the cyberspace and physical world. These systems enhance the capabilities of our critical infrastructure and are widely used in a variety of safety-critical systems. CPSs are susceptible to cyber attacks due to their vul- nerabilities such that their security has become a critical issue. Therefore, it is important to classify and comprehensively inves- tigate this issue. Most of the existing surveys on it are conducted from a single perspective. In this article, we present a compre- hensive view of the security of CPSs from three perspectives: the physical domain; 2) the cyber domain; and 3) the cyber– physical domain. In the physical domain, we review some attacks that directly damage the physical components of CPSs such as sensors and discuss corresponding defenses. We also review the attacks that CPSs in the cyber domain may face and study methods to detect and defend against them. In addition, we sur- vey the intelligent attacks faced by CPSs and the corresponding defensive means. In the cyber–physical domain, we provide an overview of attacks that come from the cyber domain and even- tually damage the physical parts, and discuss the corresponding detection and defense methods. Finally, we present the challenges and future research directions. Through this in-depth review, we attempt to summarize the current security threats to CPSs and the state-of-the-art security means to provide researchers with a comprehensive overview.

***Index Terms*—Cyberattack, cyber–physical attack, cyber– physical systems (CPSs), defense, security, vulnerability.**

## 1. Introduction

THE RAPID development of information technology has put forward higher requirements on the physi-

cal world, which entails investigations into cyber–physical systems (CPSs). CPSs are intelligent systems that integrate computing, communication, and control. They form an impor- tant part of the Industrial Internet of Things and play an important role in Industry 4.0 [1]. They can sense the world around them and have the ability to adapt to and control the physical world [2]. They closely integrate cyber and phys- ical processes, and exchange data and information in real time. Physical processes are usually carried out by several tiny devices with sensing, computing, or communication capabil- ities. These physical devices can be identified with physical properties or information sensing devices and are connected to a cyber system, to send data to the computing system [3]. The development of CPSs has gone through different stages: embedded systems, intelligent embedded systems, systems of systems [4]. They are widely used in many dif- ferent fields in the current development stage, such as power transmission systems, agricultural systems, military systems, and autonomous systems [5] (unmanned aerial vehicles and autonomous driving systems, etc.), as well as other fields directly related to our daily life.

Although CPSs have many advantages and are developing fast and are being more widely used, attacks on CPSs can result in immeasurable losses. For example, in March 2019, the Venezuelas Gury Hydropower Station that provides 80% of its country's electricity, was destroyed, causing power out- ages in 18 of the country's 23 states. Large-scale blackouts paralyze traffic, interrupt communications, and prevent fighter jets from taking off and landing [3]. Therefore, it is important to establish robust security measures.

### *A. CPSs Definition and Architecture*

CPSs are generally considered to be multidimensional and complex systems that integrate computing, networks, and physical environments. The 3C technology is the col- lective name of communication, computation, and control technologies. The main purpose of CPSs is to use the com-

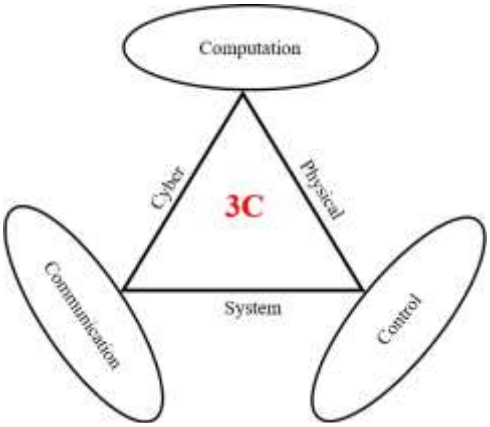


Fig. 1. 3C technology.

CPSs are first proposed as a new technology to integrate the physical world and virtual applications (such as cloud computing) into computing applications [10]. Gill [11] of the National Science Foundation provided a more complete defi- nition: CPSs are physical, biological, and engineering systems whose operations are integrated, monitored, and controlled by the computing core. Computation is deeply embedded in every physical component and may even be embedded in materials. The computing core is an embedded system that usually needs a real-time response and is usually distributed. The modern definition of CPSs is the integration of computing, communi- cation, and control capabilities to monitor and control entities in the physical world. The physical process is controlled and monitored by the cyber process, and the cyber process is also affected by the physical one [12].

In terms of their definition, researchers have reached a con- sensus, but there are still many different opinions about the architecture of CPSs. They have several mainstream architec- tures, such as the prototype architecture [13], "publish and subscribe" architecture [14], service-oriented architecture [15], and cloud-based architecture [16]. In this article, we focus on the three layer architecture. They are perception, network, and application layers, as shown in Fig. 2.

The first layer is the perception layer, also called the sensing or the recognition layer [17]. This layer includes sensors, actu- ators, global positioning systems (GPSs), and radio-frequency identification (RFID) tags along with other terminal devices for collecting real-time data to monitor or track the physical world and execute commands from the controller. The data collected can be sound, light [18], electric, biology or location, depending on the type of sensors.

The second layer is the network layer, also known as the transport layer [19] or transmission layer [1]. This layer transmits the sensory data through the network and exe- cutes control commands from the application layer. The data transmission uses local area network (LAN) and commu- nication networks, including 4G, 5G, infrared, Wi-Fi, and ZigBee, etc. This layer also uses routing devices, Internet gate- ways, firewalls, and intrusion detection systems to ensure data transfer [20].

The third layer is the application layer. Its tasks are to process information received from the network layer and issue commands that are executed by physical units such as actuators [21]. This layer also receives and processes information from the perception layer and then determines the automated actions that need to be performed [22]. Cloud com- puting and data mining algorithms are used to manage this layer of data [23]. In addition, this layer requires a robust multifactor authentication process to prevent unauthorized access [24].

### *B. CPSs Development History and Research Status*

In 2006, the American National Science Foundation proposed and described the concept of CPSs in detail, and then the construction of "New Science" began. CPSs have attracted much attention from governments, academia, and industry. In 2008, the United States established the CPS Steering Group to apply CPSs to energy, transportation, medical treatment, and agriculture. Germany also proposed "Industry 4.0," a core technology of which is the CPS [25]. By 2025, with the intro- duction of CPSs into Industry 4.0, the total gross value added of Germany is predicted to be 267 billion euros [26]. For Made in China 2025, CPSs are considered to be a comprehensive technology that promotes the integration and development of manufacturing and the Internet.

The emergence of CPSs has aroused widespread concerns in various countries. CPSs have been a priority issue for the United States, which seeks to seize the commanding heights of global industries. In 2013, the "German Industry 4.0 Implementation Recommendations" made CPSs the core tech- nology of Industry 4.0. South Korea tried to offer CPS courses as early as 2006 and focused on cross-platform research in computing, communications, and embedded objects. In Japan, the application of CPSs in smart medicine and robotics is led by the University of Tokyo. With the rapid integration and development of manufacturing and the Internet, CPSs are becoming core technology systems that support and lead a new round of global industrial change. In China, the Chinese Academy of Sciences initiated research on CPSs as early as 2007; it was not until 2016 that Germany used CPSs as a basic science, and it attracted domestic attention. The White Paper on China's CPSs focuses on "What are CPSs" and "Why are CPSs."

At present, CPSs theory is still under construction, and the related research still faces many problems that need to be solved. Since the National Science Council of China listed CPSs as an important area in 2006, it has held many relevant seminars internationally. Many journals have also published related special issues, which summarized the basic architec- ture of the system and the modeling, system testing and verification, information acquisition and

processing, commu- nication modes and protocols, intelligent computing methods, advanced control methods, information security and compre- hensive security analysis, and other theories and methods. Researches on CPSs in industrial control systems (ICSs), intel- ligent transportation systems, energy systems, and medical treatment have also attracted much attention.

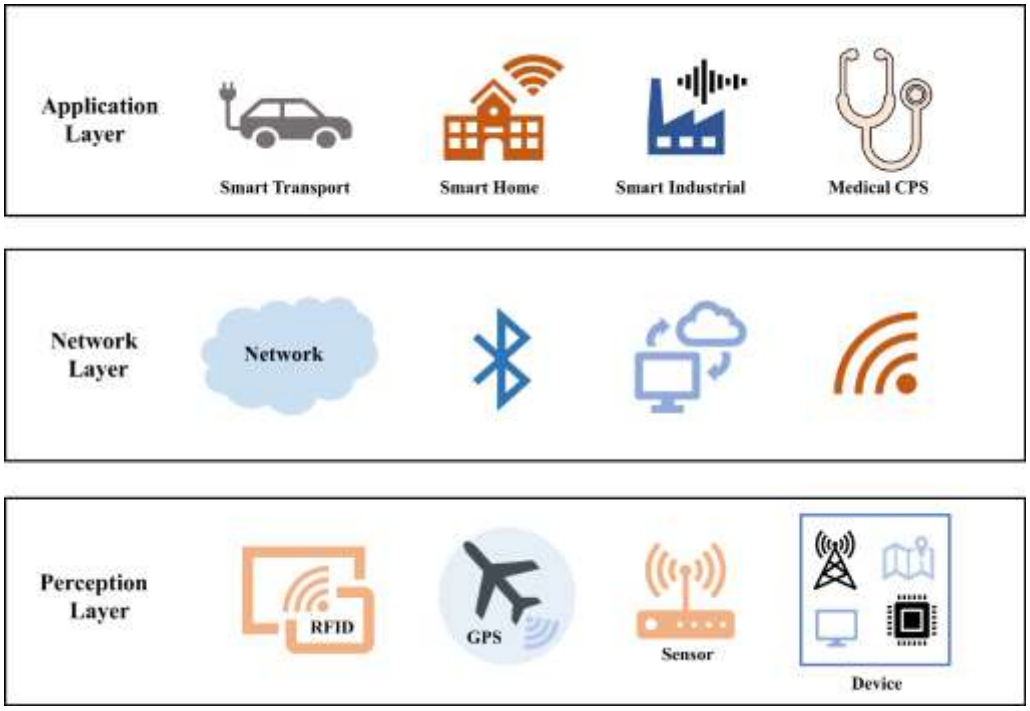


Fig. 2. CPSs layer structure.

### *C. Research on CPSs Security Issues*

In CPSs, data can be captured by physical objects or sensors and transmitted to a control system over a network. Physical devices are increasingly equipped with barcodes and RFID tags that can be scanned by smart devices, sending identi- fied information over the Internet to monitor and manage the physical environment [23]. At the same time, computing and processing units can be placed in the cloud, where decisions are generated and sent to physical objects [22]. The close integration of cyber and the physical world poses significant security challenges on CPSs.

In recent years, some researchers have studied the security issues of CPSs. Lu et al. [27] proposed a security framework for CPSs and analyze three aspects of the security objec- tives. Dibaji et al. [28] reviewed the security of CPSs from the perspective of system and control. Different CPSs secu- rity objectives are discussed in [29] and [3]. The security issues and challenges faced by CPSs are presented in [1] and [3]. As the integration of the cyber and physical processes in CPSs is becoming increasingly closer, CPSs may be attacked from the cyber domain, resulting in a series of consequences, such as hardware damage or certain failures. However, the existing studies have not divided the attacks faced by CPSs into specific domains (cyber, physical, and cyber–physical domains) to conduct a comprehensive analysis of the secu- rity of CPSs. Cyber–physical security is the difference of the security issues between CPSs and other systems or applica- tions. It means that an attack in cyberspace can impact on the physical equipment in ways that can be previously realized by physical means. Therefore, in the following, we analyze the security threats faced by CPSs from the above domains and propose corresponding solutions to the security attacks faced by CPSs.

### *D. Contributions of This Article*

In this article, we classify the security threats to CPSs into three domains: 1) physical; 2) cyber; and 3) cyber–physical domains, and review the attack mechanisms as well as detec- tion methods and defensive measures for each attack. The contributions include the following.

A comprehensive overview of the general background of CPSs, including the development of CPSs and the existing architectures is provided.

The security of CPSs is reviewed from a new per- spective, i.e., the physical domain, cyber domain, and cyber–physical domain.

The possible security threats to CPSs' intelligent systems caused by the widespread application of arti- ficial intelligence are analyzed, and the corresponding defensive measures are presented.

A comprehensive summary of security threats and defense methods for CPSs is provided, and the current challenges and future research directions are presented.

### *E. Organization*

Aside from the introduction, this article is divided into four main sections as follows. Section II details the key security threats that CPSs may face from the physical, cyber, and cyber–physical domains. Section III presents and analyzes the main CPSs security solutions that can be taken against the attacks from each domain. Finally, Section IV concludes this article.

## 2. Attacks on Cyber–Physical Systems

There are much related work on attack classification in CPSs [3], [4], [3], and this article classifies CPSs attacks

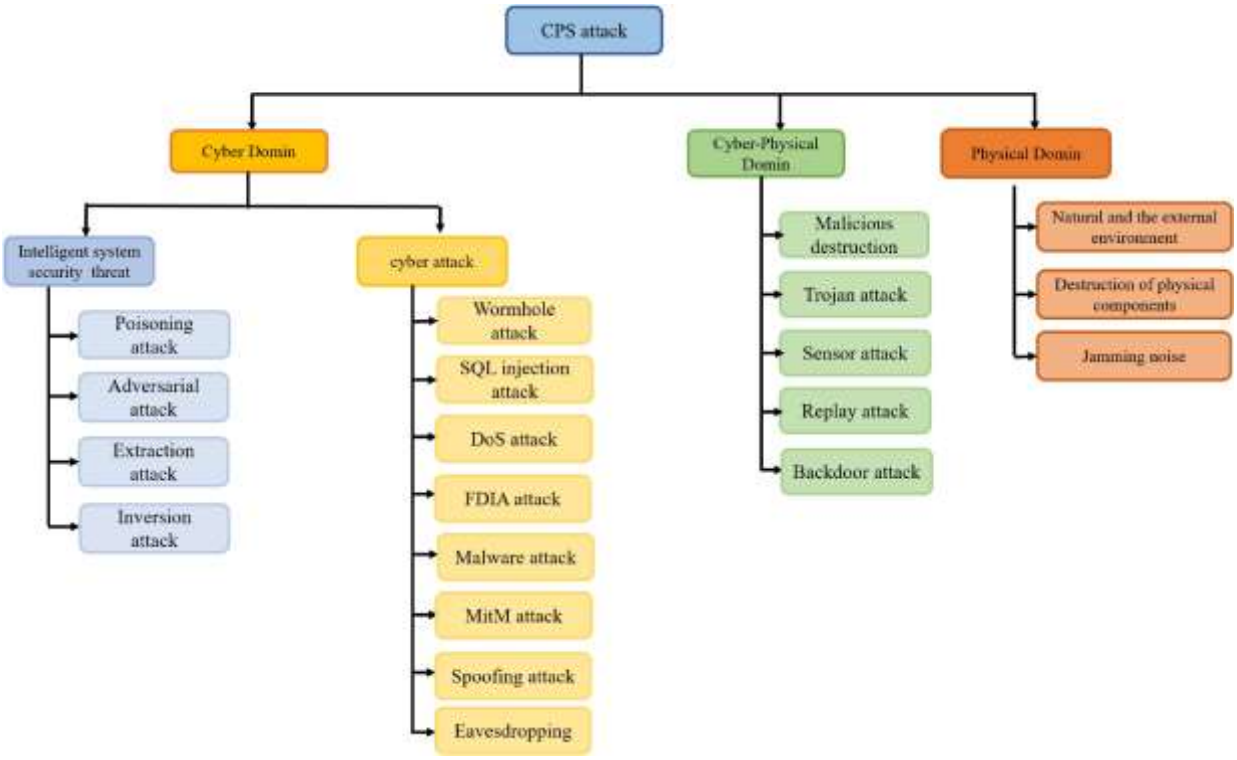


Fig. 3. Overall structure of a CPS attack.

TABLE I

Summary of Physical Attacks

| Types | Descriptions |
|---|---|
| Destruction of physical components | This attack directly destroys physical components. |
| Jamming noise | This attack causes the system to not work properly through system noise or signal interference. |
| Nature and environment | Some uncontrollable factors, such as weather or disasters, cause damage to physical components. |

from three domains. Fig. 3 shows the classification of attacks faced by CPSs in this article. Attackers can directly dam- age physical devices such as sensors and actuators that are called physical domain attacks. Cyber-domain attacks mainly refer to attacks on communication networks, such as wormhole and structured query language (SQL) attacks that may result in data leakage and transmission delays. Attackers can dam- age the physical domain, such as physical equipment, through the cyber domain, which we call cyber–physical attacks. We introduce these three types of attacks in this section.

### *A. Attacks on the Physical Domain*

1. *Natural and the External Environment:* The physical layer is composed of terminal devices, such as sensors and actuators, and these devices are generally located in an external or outdoor environment. Thus, they are susceptible to physi- cal attacks, such as stealing device components or replacing these devices. Common physical failures are equipment fail- ure, line failure, perceived data destruction, node capture, etc. A summary of the physical domain attacks is presented in Table I.

Natural and environmental factors include the impact of tor- nadoes, wildlife, and plants that may grow wildly. In [6], a hailstorm in Philadelphia that lasted several days make 75 000 people without electricity. In addition, there were more than 50 blackouts in the United States due to wildlife feeding on cables.

2. *Destruction of Physical Components:* The physical layer of CPSs consists of sensors and actuators, which are connected through a wired or wireless network [2]. The destruction of sensors, actuators, or the wires that connect them may cause CPSs to become unusable. However, due to physical or technical limitations, sensors and actuators are generally dis- tributed outdoors without much protection and are thus easily damaged. For example, a smart energy meter, i.e., the Intron centrum [3], can automatically calculate power and send results to a company. However, an attacker can easily access its hardware and destroy data by damaging sensing devices, thus causing financial loss to the company. Cardenas et al. [8] mentioned that attackers destroy some sensors or controllers to oscillate a physical system at its resonance frequency.

3. *Jamming and Noise:* System noise usually refers to the bombardment of a large number of radiated signals on an

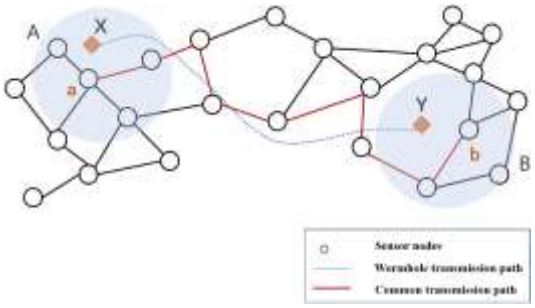


Fig. 4. Wormhole attack.

audio/video system. The system inevitably suffers from noise interference. Maheshwari [9] mentioned that by blocking the wireless channel between sensor nodes and remote base sta- tions, noise or signals of the same frequency can be introduced. These attacks may result in Denial of Service (DoS) by cre- ating intentional network interference [0]. An attacker can transmit interfering signals at the same frequency via a mali- cious device. If the interference continues in an area, all nodes in the area would be unable to communicate [24].

### *B. Attacks on the Cyber Domain*

*Cyber Attacks:*

*Wormhole attack:* According to [22], a wormhole attack makes a node transmit data by masquerading as the shortest channel; it is a malicious node in networks that cap- tures packets from one location, transmits them to another malicious node through a tunnel, and then replays these packets locally.

If a packet usually passes several hops from positions

*X* to *Y*, the packet transmitted through a wormhole near

*X* would arrive at *Y* before the packet passes through the multihop network. As shown in Fig. 4, the source node can send packets to node B through a wormhole link instead of adopting a multihop path. This kind of data packet transmit- ted through a tunnel can arrive faster or with fewer hops than that transmitted through conventional multihop routing [22]. Attackers can make nodes *a* and *b* believe that they are neighbors by forwarding routing messages, and then selec- tively discard the data messages to destroy the communication between nodes *a* and *b* [24].

Wormhole attacks are common in wireless sensor networks (WSNs). Attackers can create wormhole tunnels between two endpoints to replay messages observed in different regions [3], [2]. For cars in the Internet of Vehicles, two malicious vehicles in a network can cooperate and transmit packets from their dedicated tunnel. In addition, the first mali- cious node would generate a higher signal strength intensity to persuade legitimate nodes to believe that they are close to the destination [5].

Teng et al. [6] presented a wormhole attack detection algo- rithm related to the node trust optimization model against wormholes in WSNs. The proposed method owns a high detec- tion rate and a low false-positive rate for networks with high node density and high vulnerability, which ensures the safety and reliability of the WSNs.

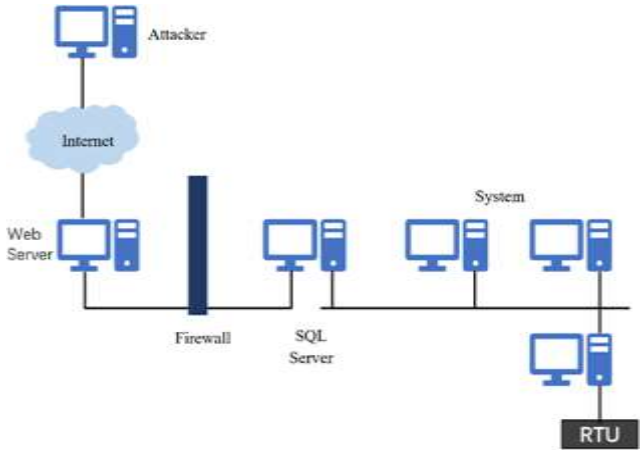


Fig. 5. SQL attack.

*SQL injection attack:* Many CPSs still rely on databases for data management. SQL injection attacks are commonly used by hackers to attack databases, and attack- ers can access data records without authorization. SQL comes from many different dialects, but most are based on the SQL- 92 ANSI standard [4]. SQL queries contain one or more SQL commands, such as SELECT, UPDATE, or INSERT. The type of SQL query makes the SQL language very popular and flex- ible. Hence, SQL attacks are prone to occur. SQL injection attacks target websites driven and managed by a CPS database to read sensitive data or delete data, resulting in database shutdown and other consequences [7].

Halfond et al. [8] mentioned some of the main types of SQL attacks. Most small industrial applications can use SQL for structural modification and content manipulation.

A supervisory control and data acquisition (SCADA) system is a typical CPS. Given the current data historians and Web accessibility in a SCADA system, SQL injection is one of the most important Web attacks and, thus, is of great signifi- cance to the security of a SCADA system [4]. In [9], SQL attacks against SCADA systems are studied (shown in Fig. 5). Even with a firewall installed, SQL attacks can still occur. An attacker may send commands to the SQL server through the Web server, which may compromise information such as user authentication inside the SQL server.

To address the threat of SQL-injection attacks (SQLIAs), Gowtham and Pramod [5] proposed SQLIA-prediction systems by using semantic features combined with the highly robust computing environment. Moreover, to alleviate com- putational burden, the authors introduce two feature selection algorithms called Mann–Whitney significance predictor test and principal component analysis.

This work in [1] focuses on a systematic review of machine learning and deep learning solutions that have been used to improve the detectability of SQL injection attacks. This systematic review allows researchers to understand the intersection between SQL injection attacks and artificial intel- ligence.

*DoS attack:* A DoS attack [5], [3] is a kind of resource exhaustion attack that makes communication

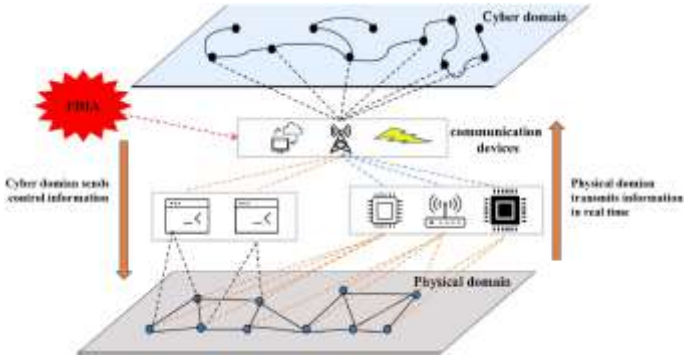


Fig. 6. FDIA.

networks or servers unable to provide services by using the defects of software or communication protocol or by send- ing a large number of useless requests to exhaust the server's resources [5]. In [5], some examples of DoS attacks that occur in CPSs are described. The work in [6] presents different types of DoS attacks.

A more serious DoS attack is a distributed DoS (DDoS) attack. In 2016, U.S. attackers launched the largest DDoS attack on the Dyn server through small CPSs and Internet of Things devices, causing downtime for Twitter, cable news network (CNN), and the Guardian [3].

In CPSs, DoS attacks mainly block the information exchange between controllers and actuators by consuming communication bandwidth. These attacks cut off their link, making the controller unable to obtain feedback from actua- tors, thus causing CPSs to be out of control [7]. Similarly, Koscher et al.

[8] proposed a DoS attack applied to intelli- gent vehicles, which disables controller area network (CAN) communication among vehicle body control modules (BCMs) and makes speedometers suddenly indicate 0. The attack also causes an instrument panel cluster (IPC) to freeze.

This article in [9] provides a structured and compre- hensive survey of the existing application-layer DoS attacks and defense mechanisms. This article classifies the exist- ing attacks and defense mechanisms into different categories, describes how they work, and compares them based on relevant parameters.

*False data injection attack:* Another potential threat to CPSs is false data injection attack (FDIA) [13], [14]. This type of attack mainly involves an attacker injecting false sen- sor data into a sensor or transmitting false data to trigger a malicious event [5]. Fig. 6 shows the process. The FDIA was originally introduced in smart grids. A smart grid is a typical CPS. An attacker modifies sensor readings in some way, and eventually an undetected error occurs. The FDIA can interfere with the process of power system state estimation. A success- ful FDIA may cause a state estimator to send an error message to system operators, resulting in a series of impacts on power systems [15]. The FDIA is a hot topic in the study of power system security, which is of great significance to the stability and safe operation of smart grids.

One form of FDIA is that an attacker destroys sensors and sends damaged sensor readings to state estimators to mislead controllers [63]. For example, Sedjelmaci et al. [64] mentioned that drones located in the same neighborhood should report the same phenomenon. However, malicious drones may dis- rupt sensor readings and cause erroneous physical phenomena. This attack is usually directed against CPSs with WSNs [6]. Injecting false data into smart grid traffic can lead to differ- ent consequences, such as service interruption and financial losses [6]. Some researchers have put forward other FDIA attacks against data integrity in CPSs [7], [8], [9].

Lu and Yang [7] studied the stealthy FDIA design problem for CPSs that has state estimators and attack detectors. The work obtains a necessary and sufficient condition for the exis- tence of perfect and nonperfect attacks. The advantage of the proposed method is that attacks have no knowledge of estimator and can be injected at any time.

*Malware attack:* Malware is used to damage CPSs devices to steal data or bypass control systems [1] and is one of the potential threats to CPSs. It can result in abnormal system behavior, including stealing important system data.

Min and Varadharajan [1] proposed an attack method called feature distributed malware (FDM), which can be used to attack CPSs supported by the Internet. This attack mainly targets low-cost devices such as sensors because they are less secure.

Malware attacks may be able to see a user's system activities without the user's authorization. Flame is a typical malware that targets ICS with spying purposes. Flame monitored the ICS networks in the Middle East and was discovered in 2012. The main goal of this malware is to collect private data related to companies, such as emails, keyboard keys, and network traf- fic [2]. Yu et al. [3] presented a malware propagation model in CPSs, namely, $SEI^2RS$, which considers two infectious rates. The equilibria are calculated, and the stability, bifur- cation of the equilibria are analyzed and proved. Simulation results show the impact of malware spread on CPSs.

There are also some malware targeting specific systems to intercept traffic or interrupt operations [6]. The work in [7] presents an overview of different malware types and the vectors of attacks subjected to modern vehicles injection. Moreover, the work also have an in-depth survey of available defenses against such attacks, and show how the defense can be used for secure intelligent vehicles against malware threats.

*Man-in-the-middle attack:* In CPSs, when an attacker tries to eavesdrop on communication between a system and a server, a man-in-the-middle (MITM) attack may occur [5]. The attacker sends forged information to the server, and the server performs unnecessary operations based on the received information, which may lead to some undesirable conse- quences [12]. The attack process is shown in Fig. 7. In [2], Melamed discussed an MITM attack between a Bluetooth smart device and its designated mobile application. This case study proves that when a Bluetooth device is connected to a mobile device, an attacker can control even a mobile device once the attack succeeds.

The commonly used techniques for the MITM attacks are packet injection, session hijacking, and SSL stripping [3]. Akter et al. [15] established MITM attacks in near field communication (NFC) between a passive tag and an active terminal, illustrate the possibility that the designed attack can

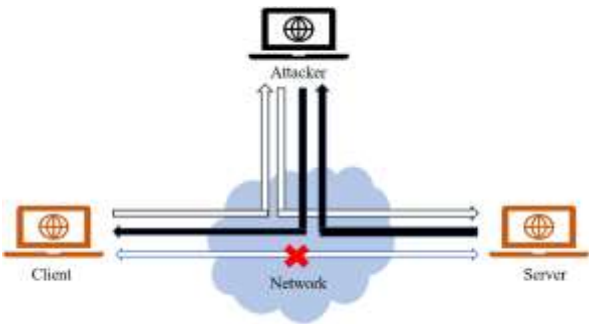


Fig. 7. MITM attack.

compromise the process of a contactless payment via a mali- cious MITM card, and also show the impacts of the MITM attacks on attack/victim scenarios.

*Spoofing attack:* A spoofing attack occurs when an attacker pretends to be a part of a CPS to participate in its legal activities [12]. After successful installation, in addition to introducing incorrect information, attackers can not only access information from CPSs but also modify or delete it [3]. Common network spoofing attacks include IP spoofing, address resolution protocol (ARP) spoofing, domain name server (DNS) spoofing, email spoofing, and routing spoofing. These attacks are usually set up and initiated on a network to obtain confidential system information [5]. The work [16] tackles three problems in GPS spoofing attack: 1) multiat- tack detection on different phasor measurement unit (PMU);

2) attack detection about the dynamic model of power systems; and 3) measurement correction. The results are illustrated for the detection method in the PMU and SCADA systems.

*Eavesdropping:* Eavesdropping refers to an attack in which an adversary can intercept information communicated by a system [26]. In CPSs, control information may be moni- tored during the transmission from a sensor to a server [3]. In addition, it is possible to intercept the monitoring data trans- mitted by sensor networks collected by monitoring via traffic analysis.

Balakrishnan et al. [21] introduced two new types of eavesdropping attacks based on a next-generation wireless communication network, i.e., opportunistic stationary attacks and active nomadic attacks, and studied the success probability of these two attacks.

Yang et al. [28] studied the security issues to a CPS under eavesdropping attacks. For a network system that is attacked by eavesdropping, the researchers establish necessary and suf- ficient conditions for an eavesdropper to carry out observations in CPSs.

Wu et al. [28] studied eavesdropping and anti-eavesdropping relations between a UAV-enabled eavesdropper (UAV-E) and a UAV-enabled base station (UAV-BS) in a downlink wire- tap system. In particular, they provide definition and existence of Nash equilibrium, and a Gauss-Seidel-like implicit finite- difference method. Finally, numerical results illustrate the effectiveness of the proposed game model.

*Intelligent System Security Threat:* In recent years, the rapid development of artificial intelligence technology has made CPSs more intelligent, which brings many new security

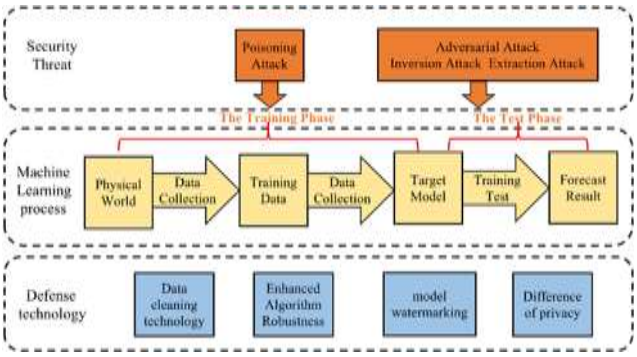


Fig. 8. Intelligent system security threat.

threats to CPSs. For example, in Uber Autonomous Driving accident in Arizona in March 2018, an autonomous vehicle failed to detect pedestrians and killed them. The work in [15] systematically discusses the existing research and summarizes the adversarial attacks and defenses for CPSs by using several kinds of their senor data. With the development of society, we have put forward higher requirements for the security of artificial intelligence systems. The main attacks on artificial intelligence systems in CPSs are poisoning attacks, adversar- ial attacks, extraction attacks, and inversion attacks, which are shown in Fig. 8.

*Poisoning attacks:* In poisoning attacks, an attacker modifies data and distribution to affect training results of an artifical intelligence model in CPSs [16].

Generally, using various methods to gain unauthorized access to data, attackers can mark enough data points to tam- per with training data to obtain desired effects. Yang et al. [17] contaminated a training data set by injecting constructed false association data into a recommendation system and realize human intervention, thus affecting the results of a recommendation system.

Attackers can also confuse a model by changing enough data. For example, through the continuous training and instiga- tion of some racist netizens, Microsoft's chat robot eventually turns into a racist and foul-mouthed robot [17].

In [17], a poisoning attack with a target is executed in a deep learning system. An attacker only needs to know that a small amount of contaminated data is inserted into the training data sets, and a backdoor can be inserted into the training model to make the model classify and judge according to the attacker's purpose.

This article in [29] reviews existing poisoning attacks and countermeasures in intelligent networks, compares the princi- ples of different types of poisoning attacks, and analyzes the advantages and drawbacks of defense methods.

*Adversarial attacks:* Most of the traditional machine learning models are based on a stability assumption: training data and test data follow approximately the same distribution. When rare samples or even maliciously constructed abnormal samples are input into a machine learning model, the machine learning model may output abnormal results [23].

By constructing an adversarial sample, an attacker can interfere with reasoning process of artificial intelligence services to achieve attacks such as evasion detection. Kurakin et al. [21] designed an anti-sample attack against an unmanned driving system. By overlaying a disturbance sign on a road sign, the authors show that the Youdao unmanned driving system recognizes “parking” as a “speed limit.”

In the field of machine vision, adversarial samples are divided into target attacks [21] and nontarget attacks [25] according to the attack effect; based on an attacker’s ability, attacks can be classified as white-box attacks [24] and black- box attacks [22]. Kumar et al. [22] conducted an empirical study on speech error interpretation attacks in speech systems. This article in [11] carefully discusses different types of adversarial attacks and corresponding defense strategies, con- cluding that adversarial learning is the real threat to machine learning in applications.

*Extraction attacks:* In an extraction attack, an attacker can send polling data and view the corresponding response results to infer parameters or functions of a machine learning model and copy a machine learning model with similar or even identical functions [21].

Lowd and Meek [21] proposed an algorithm to steal the parameters of a linear classifier model. Based on the princi- ple that the parameters learned by the machine learning model can minimize the cost function, Wang and Gong [97] present a hyperparameter estimation method for the machine learning model. Wang and Gong [97] introduced an model extraction attack that are used for stealing confidential information of the learning models through public queries, and optimized the attack behavior by sending the data based on the real- time feedback. Then, a defense strategy based on differential privacy is proposed for mitigating this kind of attack.

*Inversion attacks:* An inversion attack refers to inverse extraction of training data set information from the model, which mainly includes member reasoning attack and attribute reasoning attack [22]. Attackers can pry into the privacy of the training data based on the difference in fitting between the training data and nontraining data.

The attribute reasoning attacks [4] mainly obtain attribute information such as age distribution, prevalence, and income distribution of the data set. For instance, Fredrikson et al. [9] elaborate on the inversion attack through the issue of pri- vacy in medical machine learning. Specifically, attackers try to infer the patient’s genotype based on the warfarin dosage information.

The member reasoning attack [84] mainly infers whether a specific record appears in the data set. Truex et al. [5] proposed a general system scheme for member reasoning attacks in the MlaaS platform. At present, member reasoning attacks can be implemented through three methods, namely, the training data model [10], [11], probability information calculation [12], and similar sample generation [10].

Alufaisan et al. [14] introduced a novel technique that com- plements differential privacy to ensure model transparency and accuracy, which are robust against model inversion attacks. In fact, the proposed method with differential privacy has high transparency and preserves privacy against model inversion attacks.

*Cyber–Physical Attack*

In [15], Miller and Valasek referred to cyber–physical attacks as cyberattacks “that result in physical control of var- ious aspects” of a CPS. However, Yam et al. defined them more generally as cyberattacks with “physical effect propa- gation.” A more general definition is put forward in [16]. Researchers consider that a cyber–physical attack as a secu- rity vulnerability in cyberspace, which has have a negative impact on the physical space. For example, some attackers may damage network components by injecting malware. A noted example is Stuxnet [17] that exploits software vulner- abilities to damage centrifuges used for uranium enrichment, causing very serious consequences.

A physical device here refers to any device that collects information about a physical environment such as a sensor, e.g., sensing movement, measuring temperature, and sensing sound. An actuator is a device that can be turned on or off. Actions that occur through the cyber domain include turning on a medical device, disabling an air bag, and turning a light on or off.

*Malicious Destruction Attack:* Malicious damage can occur through malware injection. In smart cars, malware injec- tion through an OBD-II port requires physical access to a car. Hoppe et al. [18] showed how an injected malware can launch a number of malicious destruction attacks, such as preventing passengers from opening and closing windows and preventing a car from displaying missing airbag warning lights.

Checkoway et al. [19] conducted an attack, launched by a compromised device connected to the car via Bluetooth. This is realized by installing hidden malware, a Trojan Horse, on the connected smartphone. The malware captures Bluetooth con- nections and then sends a malicious payload to the transmis- sion control unit (TCU). Then, once the TCU is compromised, the attacker can communicate with safety-critical electronic control unit (ECUs), such as antilock brake system (ABS). In addition, Samuel et al. [10] show a wireless attack that exploits a malicious diagnostic mobile APP connected to the OBD-II port via Bluetooth. Since the APP runs on a mobile device, the attack can be launched through cellular networks. The cellular channel in TCU is exploitable and vulnerable to malware injection attacks. An attack is realized by calling a target car and injecting a payload by playing an MP3 file [19]. In 2003, the Slammer worm, which had infected thousands of personal computers worldwide, injected the network of the Davis–Beth nuclear power plant in Ohio and disabled its display [06].

*Trojan Attack:* A Trojan virus refers to a piece of mali- cious code with special functions hidden in normal programs. In [8], an attacker cooperated with a hacker and used a Trojan virus to control the central switch responsible for controlling the flow of natural gas through a pipeline, thus breaking into the largest natural gas company in Russia. In [11], the explo- sion of the Siberian natural gas pipeline is due to a Trojan virus implanted in SCADA systems that regulates the gas pipeline. The malicious program changes the coordination of the pump, turbine and valve, which changes the pressure in the pipeline and doubles the power of the explosion. The article in [12]

TABLE II

Cyber Attack Defense Summary

| Type | | |
|---|---|---|
| Wormhole attacks | A new mechanism called packet traction, which is a detection mechanism based on the round-trip time (RTT) and the number of neighborhoods. | |
| | Defense coding, SQLIV detection, and SQLIA runtime prevention | [48],[113],[114] |
| | An elastic model predictive control (MPC) framework, a queuing model, Bernoulli model and Markov model | [115],[116] |
| | a false data filtering scheme based on adjacent information (NFFS) and a false data filtering scheme based on geographic information (GFFs) | [117-119] |
| Man-in-the-middle attack | A computing device called an intrusion detection module to detect attacks | [54],[120-121] |
| Malware attack | A new strategy of malware defense using security authentication, an antivirus software, firewalls and web security gateways, based on the cooperation of trace-routing and trusted neighbor nodes | [54],[122-125] |
| Spoofing attack | A malicious host detection algorithm based on the ICMP. | [54],[124-125] |

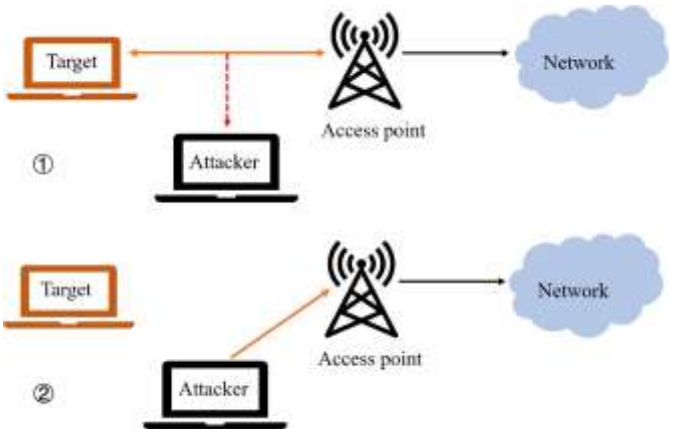


Fig. 9. Replay attack.

provides attack methodologies to neural-architecture-search (NAS) enabled edge devices for identifying NAS's vulnerabil- ity to trojaning attacks and interpret the backdoor attack, and it illustrates that the occurrence of high-impact nodes decreases the robustness of the systems.

*Sensor Attack:* Communication network security plays a very important role in CPSs. The information measured by a sensor from a physical environment or the commands gen- erated by a controller are some of the main attack targets. By sending wrong data to a sensor, an attacker causes a controller to make decisions based on incorrect measurement results and issue incorrect commands, which may make CPSs enter an unsafe state [11].

*Replay Attacks:* A replay attack occurs when an attacker sends a packet that has been received by a destination host to cheat CPSs. It is mainly used in the identity authentication process to destroy the correctness of authentication [14]. As shown in Fig. 9, an attacker captures the authentication of one or more sessions.

Attackers replay an authenticated session, or use multiple sessions to synthesize the authentication portion of the ses- sion. Since the session is valid, the attacker can establish an authenticated session. Koscher et al. [58] successfully dis- abled a cars interior and exterior lights by sending previously eavesdropped packets.

In a medical CPS, by using the loophole of an insulin pump, we can replay eavesdropping packets by replaying the pin of a previous intercepting device [15]. In addition, replay attacks may lead to incorrect decisions about insulin injection [16]. For example, by replaying an old continuous glucose monitor packet to an insulin pump, a patient would receive dishonest glucose level readings and therefore would mistakenly decide to inject a wrong amount of insulin. This incorrect decision can lead to serious health conditions.

Naha et al. [17] tackled the problem of replay attack detec- tion by watermarking the control inputs and execute resilient detection via cumulative sum test on the innovation signal and the watermarking signal. Compared with the related work, the simulation results shows that the presented methodology has smaller detection delay.

*Backdoor Attack:* A backdoor is a computing program that allows an attacker to access a CPS without authorization. With access, an attacker can launch any attack on CPSs [18]. A backdoor may be one of the main security problems of CPSs. Backdoors can be created by programmers during soft- ware development stages. Backdoors can also be created by attackers. A common way to create an application backdoor is to use a Trojan horse [5].

The work in [19] proposes a federated backdoor filter defense that can identify backdoor inputs and save the data to availability by the blur-label-flipping strategy. The proposed method exploits AI, and the accuracy of detecting the backdoor recognition is up to 99%.

## III. Defense Measures on Cyber–Physical Systems

This article reviews the various security threats that CPSs may face in Section II. For the mentioned security threats, this section summarizes the corresponding defense measures and detection methods, as shown in Table II.

### *A. Physical Domain Attack Defense*

As most nodes in the physical domain are distributed in an unsupervised environment, they are vulnerable to intrusion. Attacks mainly focus on exposed physical components, such as sensors and actuators. In addition to being easily affected by harsh natural environments, such as lightning and hurricanes, they can also be deliberately destroyed by human beings.

Since most physical components are exposed, we need to physically protect them [24]. For example, exposed wires should be protected and smart meters should be sealed within moisture-proof devices as smart meters are usually exposed. According to NIST standard, in addition to physical protection, smart meters must have encryption modules. The standard also emphasize that smart meters need to be sealed within tamper- proof units to prevent them from being physically tampered with by unauthorized personnel [12].

### *B. Cyber Domain Attack Defense*

*Cyber Attacks Defense:* Defense is very important for the security in the cyber domain. This article in [11] pro- vides a review of the latest research results on secure state estimation of CPSs for different performance indicators and defense strategies. Then the recent secure control results have been reviewed and classified, and it also provides examples of two representative applications of secure estimation and con- trol approaches in real-world CPSs, namely, water distribution systems and wide-area power systems, to provide a prelimi- nary analytical framework for modern infrastructure security. For attacks in the cyber domain mentioned above, we propose corresponding defense and detection methods as follows.

*Wormhole attack defense:* In CPSs, especially against many ad hoc network routing protocols and location-based wireless security systems, wormhole attacks can pose serious threats. Therefore, the detection and defense against wormhole attacks are particularly important. A summary of the wormhole attack detection approaches is shown in Table II. Hu et al. [4] proposed a new mechanism called packet traction to detect and defend against wormhole attacks using a special protocol called TIK to achieve traction. In particular, TIK requires just $n$ public keys in a network with $n$ nodes, and has relatively modest storage, per packet, and computation overheads.

In [4], Hu and Evans propose a strategy to use directional antennas. In this proposed cooperation agreement, nodes share direction information to prevent wormhole endpoints from being disguised as fake neighbors. The defense proposed in this article greatly reduces the threat of wormhole attacks and does not require location information or clock synchronization. Tun and Maw [12] proposed wormhole detection mech- anisms based on round-trip time (RTT) and the number of neighbors. The first mechanism considers RTT between two consecutive nodes and the number of neighbors of these nodes. This requires comparing the values of other consecutive nodes. The second mechanism is based on the fact that, by intro- ducing new links in a network, the adversary increases the number of neighbors of nodes within its radius. The system does not require any specific hardware, has good performance, low overhead, and consumes no additional energy.

Some existing methods for detecting wormhole attacks require strict clock synchronization or long processing times. Wu et al. [13] proposed a local neighborhood information detection method based on a transmission range. Simulation results show that this method can effectively detect wormhole attacks.

*SQL injection attack defense:* SQL injection defense methods can be roughly divided into three categories:

1) defense coding; 2) SQLIV detection; and 3) SQLIA runtime prevention [14]. The root cause of SQL attacks is insufficient input verification.

It is not sufficient to use professional vulnerability scanning tools to prevent SQL injection attacks.

The latest vulnerability scanner can find newly discovered vulnerabilities [11]. Halfond et al. [8] proposed a series of techniques to prevent SQL injection attacks, such as new query development paradigms, proxy filters and instruction set randomization. Musaab et al. [15] presented a heuristic algorithm based on machine learning to prevent SQL injec- tion attacks. This article uses a data set containing a large number of statements to train different machine learning clas- sifiers, and selects the five classifiers with the highest accuracy to develop the program. The training results show that the algorithm can accurately detect SQL injection attacks.

*DoS attack defense:* Agah et al. [16] put forward two new schemes to prevent DoS attacks. The first one is called utility-based dynamic source routing (UDSR), which combines the total utility of each route in the packet. Utility is the value that we try to maximize in the game theory. The second one is based on a watch list, where each node obtains a score from its neighbors based on its previous cooperation in networks. The results show that the proposed game framework signifi- cantly increases the success rate of wireless sensor and actor networks in defense strategies.

Sun et al. [17] proposed an elastic model predictive con- trol (MPC) framework. This system can mitigate the adverse effects of DoS attacks on CPSs by modeling linear time- invariant systems. Chen et al. [18] investigated the resilient filtering issue for power systems with DoS attacks and gain perturbations. By utilizing elementary inequalities and the fashionable mathematical induction, an upper bound of fil- tering error covariance has been

derived and then minimized via selecting suitable filter gains relying on two Riccati-like difference equations. Finally, a benchmark simulation test is exploited to check the usefulness of the designed filter.

This article in [19] has investigated the maximum cor- rentropy filtering issue for a class of large-scale systems consisting of a set of spatially distributed subsystems subject to randomly occurring cyber attacks and non-Gaussian noises. A hybrid attack model composed of DoS attacks and decep- tion attacks is used to describe the complex attack behavior in practical engineering. With the help of fixed-point iteration rules, a distributed algorithm of MCC-KF has been proposed, and the desired filter gains depend on the local information and the received one-step prediction.

*False data injection attack defense:* It is difficult to defend against FDIAs due to their concealment [10]. Two FDIA detection methods are proposed in [11]:

TABLE III

Intelligent System Attack Defense Summary

| Type | | |
|---|---|---|
| Poisoning attack | Use data cleaning technology and improve thealgorithm robustness. | [130-133] |
| Adversarial attack | Check adversarial examples after building the machine learning system and make the machine learning system more robust before the attacker generates adversarial examples. | [134-136] |
| Extraction attack | To approximate model parameters or output results. | [92],[137-147] |
| | Use of machine learning algorithms with privacy protection functions. | [85],[141-143] |

1) state-estimation-based detection and 2) machine-learning- based detection.

Wang et al. [12] proposed two methods to defend against FDIAs. One is a false data filtering scheme based on geo- graphic information, which makes full use of the absolute position of a sensor; and the other is a false data filtering scheme based on adjacent information, which makes use of the relative position of a sensor when the absolute position is not obtained.

*Man-in-the-middle attack defense:* Ahmad et al. [5] used a private network in CPSs to prevent MITM attacks. Lima et al. study a MITM attack in [13]. They set up a system deterministic model under attacks on the sensor and actuator channels and put forward a defense strategy, which can detect the intrusion and protect CPSs from damage caused by MITM attacks. To realize this model, this article develops a plant model under sensor attack and a supervisor model under actuator attack.

*Malware attack defense:* The rapid growth of malware has caused very large economic losses for various organiza- tions. The continuous progress and development of malware put forward higher requirements for its defense and detection. Previous malware defenses are largely based on fingerprint or signature technology. Antrosiom and Fulp [13] introduced a new strategy that uses security certification to defend against malware. This strategy focuses on malware vulnerabilities rather than attacks. The system uses remote security scanners to check for vulnerabilities and uses logical network segmenta- tion to isolate machines to maximize the availability of related machines while preventing attacks.

In [1], unsupervised learning and supervised learning are used to classify malware, and machine learning algorithms and deep learning models are used to analyze and detect mal- ware. This article uses methods such as cross-validation and fixing class imbalance problem to build models that ultimately increase the accuracy rate significantly.

*Spoofing attack defense:* Spoofing could be avoided by packet filtering or by using a secure encryption proto- col. The prevention of these attacks includes DVCerts and DAPS [3].

Zeng and Zhang [16] proposed a malicious host detec- tion algorithm based onthe Internet control message protocol (ICMP). This technology involves collecting and analyzing ARP packets and then injecting ICMP echo request packets according to their response packets to detect malicious hosts. It does not interfere with host activity on networks. It can also detect real address mappings during an attack.

Gao and Xia [17] used an effective method to prevent IP spoofing attacks based on the cooperation of trace routing and trusted neighbor nodes. This method can effectively detect IP spoofing attacks, thus effectively preventing IP spoofing attacks.

*Defense Against Intelligent System Attacks:* For the security threats to the intelligent system mentioned above, we propose the corresponding defense methods and make a brief summary as shown in Table III.

*Poisoning attack defense:* For data poisoning attacks, current defense methods are mainly divided into two types:

1) the data cleaning technology and 2) algorithm robustness improvement to resist malicious training data.

The data cleaning technology mainly filters and removes malicious training data directly to protect collected data from tampering and rewriting attacks [18]. An attack detection strategy is proposed to detect potential contamination by iso- lating a special holdout set. Baracaldo et al. [11] used source information as part of a filtering algorithm to detect poison attacks. They use the source of training data points and trans- form context to identify harmful data, which is implemented on partially trusted and completely untrusted data sets. This is the first defense strategy that uses data sources to prevent poi-

soning attacks. For partially trusted and completely untrusted data sets, the authors propose two variants of source defense. A learning algorithm always has to make a tradeoff between preventing regularization and reducing loss function, which may lead to vulnerability of the learning algorithm; thus, it is necessary to improve the robustness of the algorithm against malicious training data. Biggio et al. [12] proposed improving a PCA algorithm and reduce the influence of malicious train- ing data by combining the PCA with the Laplacian truncation threshold.

Jagielski et al. [19] proposed a new defense algorithm called TRIM to train a regression model with toxic data. It trains the subset of the smallest residual points in each iteration by trimming iterative regression parameters. In adversarial situations, regularized linear regression is applied, and the algorithm is proved to be more effective than other defenses on a series of models and real data sets.

*Adversarial attack defense:* The defense methods of adversarial attacks mainly focus on preventing the generation of confrontation samples and the detection of confrontation samples [20].

In [15], a SafetyNet detector is designed, and an output binary threshold of each ReLU layer is extracted as the feature of a counter detector. This method can better resist adversarial attacks because it is difficult for attackers to find an optimal value for confrontation samples and the SafetyNet detector.

Papernot et al. [23] used network purification as a defense mechanism to resist the disturbance of deep neural networks. Although there have been many studies on adversarial sample methods, there is still a lack of an effective defense strategy against adversarial attacks. Most current methods measure the lower bound of the ability to resist adversarial attacks [14].

*Extraction attacks defense:* The defense strategy for model extraction attacks is mainly to approximate model parameters [14] or output results [12]. In addition, to avoid the model from being stolen to protect intellectual property rights, some researchers have proposed the concept of model watermarking [29].

Uchida et al. [14] and Chen et al. [13] add a watermark to the neural network by adding a new regularization term to the loss function. Merrer et al. [12] combined adversarial exam- ples and adversarial training methods to inject watermarks into neural networks. Adi et al. [21] studied a black-box deep neu- ral network watermarking technology, which proved through experiments that this method does not affect the performance of the original model.

Shokri et al. [19] inject noise into the parameters, and models such as deep neural networks could be trained by multiparty computation to resist model extraction attacks. Making models no longer output a trusted value or, in some cases where the trusted value must be output, rounding the output trusted value can reduce the success rate of model extraction [26].

*Inversion attacks defense:* A typical method of defend- ing against inversion attacks is the use of machine learning algorithms with privacy protection functions. Currently, homo- morphic encryption [20] and differential privacy technolo- gies [29] are widely used.

Homomorphic encryption allows users to directly perform specific algebraic operations on the ciphertext, and the data obtained is still the result of encryption. Xie et al. [26] proposed a defense method that uses homomorphic encryp- tion technology to encrypt data, so that the neural network does not decrypt the data while processing the data, thereby protecting the confidentiality of a single input.

Differential privacy protects the information in the data by adding interference noise to the data. The greater the noise added, the better the data protection effect [21]. Papernot et al. [25] put forward a universal PATE framework to protect training data in machine learning.

*Cyber–Physical Domain Attack Defense*

We review defense and detection methods of cyber–physical domain attacks mentioned (see Table IV).

*Trojan Attack Defense:* There are also many Trojans in integrated circuits (ICs), and Trojans can be implanted in a variety of ways to weaken the security links of a chip, steal internal sensitive data or modify the original functions, which may cause severe economic losses for society. Therefore, we analyze the entire life cycle of IC and protect hardware Trojan. In [23], we elaborate an IC market model to illustrate the potential Trojan threat participation model faced by both parties.

*Backdoor Attack Defense:* Backdoor attacks have attracted widespread attention. An attacker's goal is to build a malicious deep neural network and use backdoor trigger to incorrectly classify special inputs. Because of their conceal- ment, such attacks may have disastrous consequences [14]. According to the resources owned by the enemy and whether detection is being carried out, we divide the attack and defense methods into several categories. We have made a detailed overview of each kind of attacks compared with these methods, and evaluated some attack schemes through experiments.

*Replay Attack Defense:* Mo and Sinopoli [15] assumed that the control system is a discrete-time linear time invariant (LTI) Gaussian system using an infinite level linear quadratic Gaussian (LQG) controller, which improves the probability of detecting replay attacks.

In the study of Hoehn and Zhang [14], a new method based on an irregular time interval jamming system to detect replay attacks is proposed. The advantage of this method is its robustness, and it can be easily implemented in existing control systems.

## IV. Security Challenges and Future Research Directions

The development of CPSs has made great changes in indus- try, medical care, transportation, and people's daily life, and higher requirements are put forward for quality, security and privacy. In future research, we will pay more attention to the limitations of some existing results and propose several

chal- lenging issues on this topic, shedding insightful light on further research. Through the research on CPSs, we found that the existing research on CPSs still has some problems, presented as follows.

### *A. Security Challenges*

This subsection provides several security challenges with the development of CPSs.

With the development of CPSs, CPSs will inevitably face multiple attacks at the same time instead of a single attack. Existing research has done research on multiple attacks of CPSs, but its security solutions have not been studied in depth. Therefore, designing a comprehensive detection and defense strategy is an important goal for our future research.

CPSs are a key part of Industry 4.0. They have pro- foundly changed the way in which humans interact with the physical world by integrating the physical environ- ment with the network world. Therefore, it is particularly important to study the reliability and availability of the system. Existing works generally use automata to model when studying CPSs attacks. We can use stochastic Petri nets [21], [11], [22] to model system attacks to analyze system availability and reliability.

There is nonlinear dynamic behaviors such as time- varying nodes and time-varying topologies in CPS

TABLE IV

Cyber–Physical Attack Defense Summary

| **Types** | | |
|---|---|---|
| Trojan attack | Use a personal firewall, check registry and startup group or install anti-black master | [129] |
| Backdoor attack | Use professional tools to kill, change ports, and disable services | [128] |
| Replay attack | Use a challenge-response mechanism and a one-time password mechanism or add a random number, add a time stamp, and add serial number prevention. | [126-127] |

systems. From the perspective of cybernetics, the existing analysis methods for reducing attacks cannot analyze complex system dynamics.

When the factors such as communication protocols and network attacks are considered, the complexity of systems will be greatly increased, and the conditions required by typical detection techniques may not be guaranteed. Therefore, the development of new detection strategies is important.

With the continuous development of CPSs, higher requirements are put forward for the security, reliability, availability and stability of CPSs. Therefore, in a real CPS, a multiobjective optimization problemnonumber arises.

### *B. Future Research Directions*

The size of the CPSs becomes large and complex, and enor- mous amount of data also generated by CPSs. In order to handle security issue of large and complex CPSs, security detection of CPS associated with some modern approaches like bid data and clouding computation technique is a promising research aspect in the future.

Due to the distributed nature of some CPSs such as smart grid and intelligent electronic devices, several kinds of attacks can happen simultaneously in a large scale of distributed systems. In this sense, how to identify, locate and detect these attacks in a distributed way is a important research topic in the future.

For guaranteeing the security of CPSs under attacks, secu- rity control approaches becomes a possible way. That is to say, the control policy should satisfy general requirements if there is no attack in a CPS, and it can still hold validation for malicious attacks. Consequently, designing a security resilient controller needs to be studied, which is an encouraging topic in the future.

With the continuous improvement of CPSs functions and the maturity of security defense programs, CPSs will be more widely used in various key system areas. Attacks on CPSs in recent years have shown that attackers are constantly carry- ing out more targeted and destructive attacks based on CPSs operating mechanisms and defense strategies. Although some defense mechanisms have been proposed, new defense strate- gies for identifying threats and vulnerabilities for specific systems still need to be updated.

With the deep integration of cyber systems and physi- cal systems, CPSs may face cyber attacks, physical attacks, and cyber–physical attacks. Developers construct a security framework for certain types of attacks, and according to the framework, effective control strategies can be developed to defend attacks.

Privacy is another primary consideration in defense strategy. Context-aware access control can prevent unauthorized access, and context-aware key management can prevent key leakage and provide key management mechanism.

CPSs may have an impact on the environment when they are applied to future smart cities and smart homes. Researchers need to focus on the environmental impact of CPSs and the study of green CPSs. It will also be an important issue to integrate renewable energy in CPSs to make CPSs coexist with environment friendly.

## V. Conclusion

CPSs are an important part of Industry 4.0. By combining the physical world with the cyber environment, they change the way in which that people interact with the physical world. However, CPSs suffer from many security threats and attacks that can significantly reduce their reliability, stability and secu- rity. In this article, we first review the architecture and security issues of CPSs. Then, possible attacks on CPSs are classi- fied in three aspects, i.e., physical domain, cyber domain, and cyber–physical domain. As CPSs inevitably use some intelli- gent algorithms, they are vulnerable to artificial intelligence attacks. Therefore, artificial intelligence attacks are added to the classification [11] and the corresponding defenses are given. Next, for each of the above classified attacks, we give the cor- responding detection methods and defense measures. Finally, we present the challenges of the current research and the future research directions. Compared with the existing surveys on the security of CPSs that review the security of CPSs from a single perspective, this article provides a comprehensive survey of the security of CPSs, especially from the cyber–physical domain. Finally, we highlight the challenges facing CPSs and point out future research directions, which we hope to stimulate more researchers to be interested in this field.

**References**

J. P. A. Yaacoub, O. Salma, H. N. Noura, N. Kaaniche, A. Chehab, and M. Malli, "Cyber-physical systems security: Limitations, issues and future trends," *Microprocess. Microsyst.*, vol. 77, Sep. 2020, Art. no. 103201.

S. Gries, M. Hesenius, and V. Gruhn, "Cascading data corruption: About dependencies in cyber-physical systems: Poster," in *Proc. 11th ACM Int. Conf. Distrib. Event Syst.*, Barcelona, Spain, 2017, pp. 345–346.

Y. Ashibani and Q. H. Mahmoud, "Cyber physical systems secu- rity: Analysis, challenges and solutions," *Comput. Security*, vol. 68, pp. 81–97, Jul. 2017.

U. Sendler, *Industrie 4.0—Beherrschung der Industriellen Komplexitit Mit SysLM (Systems Lifecycle Management)*, 4rd ed. Berlin, Germany: Springer Vieweg, 2013.

T. Haidegger et al., *Industrial and Medical Cyber-Physical Systems: Tackling User Requirements and Challenges in Robotics* (Topics in Intelligent Engineering and Informatics), vol. 14. Cham, Switzerland: Springer, 2019, pp. 253–277.

G. R. Gonzélez, M. M. Organero, and C. D. Kloos, "Early infrastructure of an Internet of Things in spaces for learning," in *Proc. 8th IEEE Int. Conf. Adv. Learn. Technol.*, Santander, Spain, 2008, pp. 381–383.

R. Baheti and H. Gill, "Cyber-physical systems," *Impact Control Technol.*, vol. 12, pp. 161–166, Jun. 2011.

S. Sastry, "Networked embedded systems: From sensor Webs to cyber-physical systems," in *Proc. Int. Workshop Hybrid Syst. Comput. Control*, 2007, p. 1.

E. A. Lee, "CPS foundations," in *Proc. Design Autom. Conf.*, Anaheim, CA, USA, 2010, pp. 737–742.

A. Darwish and A. E. Hassanien, "Cyber physical systems design, methodology, and integration: The current status and future outlook," *J. Ambient Intell. Humanized Comput.*, vol. 9, pp. 1541–1556, Oct. 2018.

H. Gill, "From vision to reality: Cyber-physical systems," in *Proc. HCSS Nat. Workshop New Res. Directions High Confidence Transp. CPS Autom. Aviation Rail*, 2008, pp. 18–20.

Q. Shafi, "Cyber physical systems security: A brief survey," in *Proc. 12th Int. Conf. Comput. Sci. Appl.*, Salvador, Brazil, 2012, pp. 18–21.

Y. Tan, S. Goddard, and L. C. Prez, "A prototype architecture for cyber- physical systems," *ACM SIGBED Rev.*, vol. 5, no. 26, pp. 1–2, 2008.

D. Seifert and H. Reza, "A security analysis of cyber-physical systems architecture for healthcare," *Computers*, vol. 5, no. 4, p. 27, 2016.

Y. F. Li, D. H. Sun, W. N. Liu, and X. B. Zhang, "A service-oriented architecture for the transportation cyber-physical systems," in *Proc. 31st Chin. Control Conf.*, Hefei, China, 2012, pp. 7674–7678.

A. Caggiano, T. Segreto, and R. Teti, "Cloud manufacturing frame- work for smart monitoring of machining," *Procedia CIRP*, vol. 55, pp. 248–253, Nov. 2016.

R. Mahmoud, T. Yousuf, F. Aloul, and I. Zualkernan, "Internet of Things (IoT) security: Current status, challenges and prospective mea- sures," in *Proc. 10th Int. Conf. Internet Technol. Secured Trans. (ICITST)*, London, U.K., 2015, pp. 336–341.

K. Zhao and L. Ge, "A survey on the Internet of Things security," in *Proc. 9th Int. Conf. Comput. Intell. Security*, Emeishan, China, 2013, pp. 663–667.

T. B. Lu, J. X. Lin, L. L. Zhao, Y. Li, and Y. Peng, "A security architec- ture in cyber-physical systems: Security theories, analysis, simulation and application fields," *Int. J. Security Appl.*, vol. 9, no. 7, pp. 1–16, Jul. 2015.

B. Zhu and S. Sastry, "SCADA-specific intrusion detection/prevention systems: A survey and taxonomy," in *Proc. 1st Workshop Secure Control Syst. (SCS)*, Stockholm, Sweden, 2010, pp. 1–8.

H. H. Gao, Y. Peng, K. B. Jia, Z. H. Dai, and T. Wang, "The design of ICS testbed based on emulation, physical, and simulation," in *Proc. 9th Int. Conf. Intell. Inf. Hiding Multimedia Signal Process.*, Beijing, China, 2013, pp. 420–423.

R. Khan, S. U. Khan, R. Zahee, and S. Khan, "Future Internet: The Internet of Things architecture, possible applications and key challenges," in *Proc. 10th Int. Conf. Front. Inf. Technol.*, Islamabad, Pakistan, 2012, pp. 257–260.

B. Zhang, X.-X. Ma, and Z.-G. Qin, "Security architecture on the trust- ing Internet of Things," *J. Electron. Sci. Technol.*, vol. 9, pp. 364–367, Jan. 2011.

C. Konstantinou, M. Maniatakos, F. Saqib, S. Y. Hu, J. Plusquellic, and Y. E. Jin, "Cyber-physical systems: A security perspective," in *Proc. 20th IEEE Eur. Test Symp.*, Cluj-Napoca, Romania, 2015, pp. 1–8.

J. Jamaludin and J. M. Rohani, "Cyber-physical system (CPS): State of the art," in *Proc. Int. Conf. Comput. Electron. Elect. Eng. (ICE Cube)*, Quetta, Pakistan, 2018, pp. 1–5.

S. Heng, *Industry 4.0: Huge Potential for Value Creation Waiting to Be Tapped*, Deutsche Bank Res., Frankfurt, Germany, 2014, pp. 8–10.

T. B. Lu, J. X. Lin, L. L. Zhao, Y. Li, and Y. Peng, "An analysis of cyber physical system security theories," in *Proc. 7th Int. Conf. Security Technology*, Hainan, China, 2014, pp. 19–21.

S. M. Dibaji, M. Pirani, D. B. Flamholz, A. M. Annaswamy, K. H. Johansson, and A. Chakrabortty, "A systems and control per- spective of CPS security," *Annu. Rev. Control*, vol. 47, pp. 394–411, Jun. 2019.

E. B. Harb, "A brief survey of security approaches for cyber-physical systems," in *Proc. 8th IFIP Int. Conf. New Technol. Mobility Security*, Larnaca, Cyprus, 2016, pp. 1–5