Z. Yu, H. Gao, X. Cong, N. Wu and H. H. Song, "A Survey on Cyber-Physical Systems Security," in IEEE Internet of Things Journal, doi: 10.1109/JIOT.2023.3289625.

https://doi.org/10.1109/JIOT.2023.3289625

A Survey on Cyber-Physical Systems Security

Zhenhua Yu, *Member, IEEE*, Hongxia Gao, Xuya Cong, *Member, IEEE*, Naiqi Wu, *Fellow, IEEE*, and Houbing Herbert Song, *Fellow, IEEE*

***Abstract*—Cyber-Physical Systems (CPSs) are new types of intelligent systems that integrate computing, control, and communication 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 vulnerabilities such that their security has become a critical issue. Therefore, it is important to classify and comprehensively investigate this issue. Most of the existing surveys on it are conducted from a single perspective. In this paper, we present a comprehensive view of the security of CPSs from three perspectives: the physical domain, the cyber domain, and 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 survey 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 eventually 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*—Cyber-physical systems, security, cyber-attack, cyber-physical attack, vulnerability, defense.**

I. Introduction

The rapid development of information technology has put forward higher requirements on the physical world, which entails investigations into cyber-physical systems (CPSs). CPSs are intelligent systems that integrate computing, communication and control. They form an important part of the Industrial Internet of Things and play an important role in Industry 4.0 [1]. They can sense world around them and have the ability to adapt to and control the physical world [2]. They closely integrate cyber and physical processes, and exchange data and information in real time. Physical processes are usually carried out by several tiny devices with sensing, computing, or communication capabilities. 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 different 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, Venezuelas Gury Hydropower Station that provides 80% of its country's electricity, was destroyed, causing power outages in 18 of the countrys 23 states. Large-scale blackouts paralyse traffic, interrupt communications, and prevent fighter jets from taking off and landing [3]. Therefore, it is important to establish robust security measures.

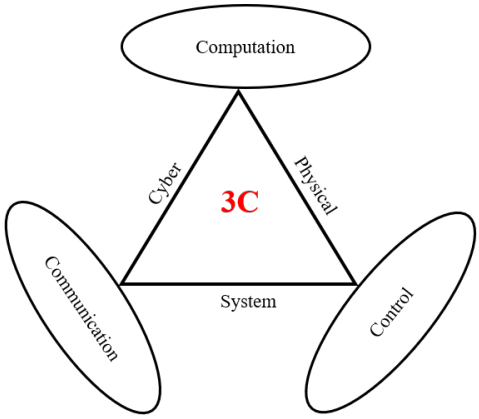


Fig. 1: 3C technology

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 collective name of communication, computation, and control technologies. The main purpose of CPSs is to use the combined 3C technology to achieve feedback control of a computing system [6]. The 3C technology is shown in Fig.1. Since the advent of CPSs, many researchers have attempted to define them. Baheti et al. [7] propose that a CPS is a highly reliable system that closely integrates various computational and physical elements in a system and coordinates them with each other under dynamic uncertain events. Sastry [8] believes

This work was supported by the National Natural Science Foundation of China under Grants 61873277 and 62273272. (*Corresponding authors: Xuya Cong and Houbing Herbert Song.*)

Zhenhua Yu, Hongxia Gao, and Xuya Cong are with the College of Computer Science and Technology, Xi an University of Science and Technology, Xi'an, 710054 China (e-mail: zhenhuayu@xust.edu.cn; 19208207028@stu.xust.edu.cn; congxuya@xust.edu.cn).

Naiqi Wu is with the Institute of Systems Engineering, Macau University of Science and Technology, Taipa, Macau (e-mail: nqwu@must.edu.mo).

Houbing Herbert Song is with the Security and Optimization for Networked Globe Laboratory (SONG Lab), Department of Information Systems, University of Maryland, Baltimore County, Baltimore, MD 21250 USA (e-mail: h.song@ieee.org).

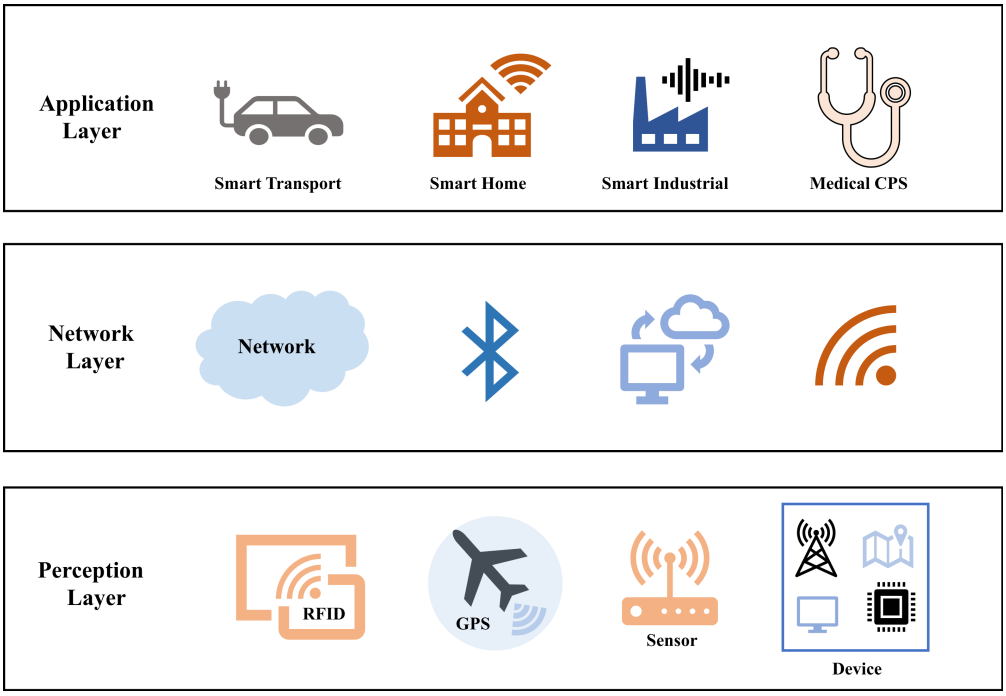


Fig. 2: CPSs layer structure

that CPSs integrate computing, communication, and storage capabilities, and can operate in real-time reliably, safely, stably, and efficiently. They are networked computer systems that can monitor various entities in the physical world. Lee [9] defines CPSs as the tight integration of a series of computing and physical processes. The computing core of CPSs is used to monitor the operations of physical entities, and the physical entities use the network and computing components to realize perception and control of the environment.

CPSs are first proposed as a new technology to integrate the physical world and virtual applications (such as cloud computing) into computing applications [10]. Helen Gill [11] of the National Science Foundation provides a more complete definition: 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 integration of computing, communication, 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 consensus, but there are still many different opinions about the architecture of CPSs. They have several mainstream architectures, such as the prototype architecture [13], "publish and subscribe" architecture [14], service-oriented architecture [15], and cloud-based architecture [16]. In this paper, 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, actuators, Global Positioning Systems (GPS), 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 executes control commands from the application layer. The data transmission uses local area network (LAN) and communication networks, including 4G, 5G, infrared, Wi-Fi, and ZigBee, etc. This layer also uses routing devices, Internet gateways, 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 computing and data mining algorithms are used to manage this layer of data [23]. In addition, this layer requires a robust multi-factor 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 cyber-physical system [25]. By 2025, with the introduction 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 technology 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 architecture of the system and the modeling, system testing and verification, information acquisition and processing, communication modes and protocols, intelligent computing methods, advanced control methods, information security and comprehensive security analysis, and other theories and methods. Researches on CPSs in industrial control systems, intelligent transportation systems, energy systems, and medical treatment have also attracted much attention.

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 identified 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] propose a security framework for CPSs and analyze three aspects of the security objectives. Dibaji et al. [28] review the security of CPSs from the perspective of system and control. Different CPSs security objectives are discussed in [29] [30]. The security issues and challenges faced by CPSs are presented in [31] [32]. 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 security of CPSs. Cyber-physical security is the difference of the security issues between CPSs and other systems or applications. 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 Paper

In this paper, we classify the security threats to CPSs into three domains: physical, cyber, and cyber-physical domains, and review the attack mechanisms as well as detection 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 perspective, 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 artificial 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 paper 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 the paper.

II. Attacks on Cyber-Physical Systems

There are much related work on attack classification in CPSs [33]–[35], and this paper classifies CPSs attacks from three domains. Fig. 3 shows the classification of attacks faced by CPSs in this paper. Attackers can directly damage 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 SQL attacks that may result in data leakage and transmission delays. Attackers can damage 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.

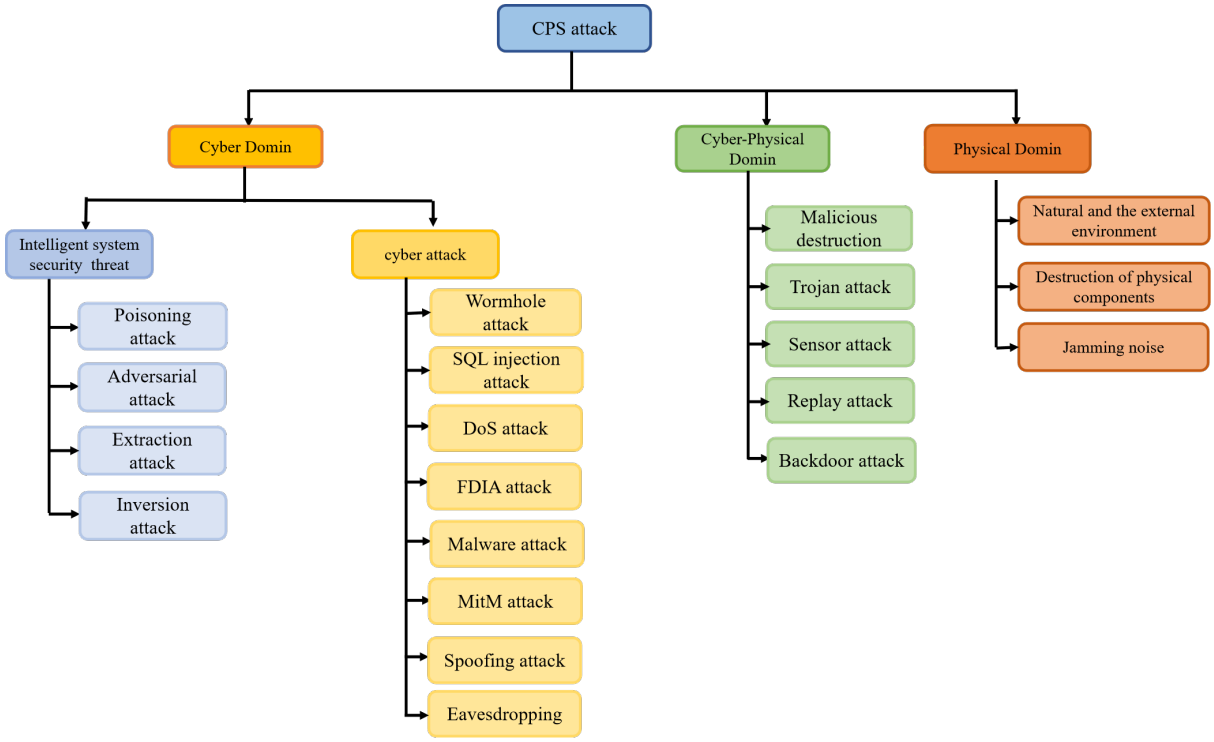


Fig. 3: Overall structure of a CPS attack

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 physical attacks, such as stealing device components or replacing these devices. Common physical failures are equipment failure, line failure, perceived data destruction, node capture, etc. A summary of the physical domain attacks is presented in Table 1. Natural and environmental factors include the impact of tornadoes, wildlife, and plants that may grow wildly. In [36], a hailstorm in Philadelphia that lasted several days make 75000 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 [32]. 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 distributed outdoors without much protection and are thus easily damaged. For example, a smart energy meter, i.e., the Intron centrum [37], 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. In [38], Cardenas et al. mention 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 audio/video system. The system inevitably suffers from noise interference. Maheshwari [39] mentions that by blocking the wireless channel between sensor nodes and remote base stations, noise or signals of the same frequency can be introduced. These attacks may result in DoS by creating intentional network interference [40]. An attacker can transmit interfering signals at the same frequency via a malicious device. If the interference continues in an area, all nodes in the area would be unable to communicate [41].

B. Attacks on the Cyber Domain

1) Cyber attacks:

a) Wormhole attack: According to [42], a wormhole attack makes a node transmit data by masquerading as the shortest channel; it is a malicious node in networks that captures 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 transmitted through a tunnel can arrive faster or with fewer hops than that transmitted through conventional multihop routing [42]. Attackers can make nodes *a* and *b* believe that they are neighbors by forwarding routing messages, and then selectively discard the

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.

data messages to destroy the communication between nodes *a* and *b* [43].

Wormhole attacks are common in wireless sensor networks. Attackers can create wormhole tunnels between two endpoints to replay messages observed in different regions [43], [44]. For cars in Internet of Vehicles, two malicious vehicles in a network can cooperate and transmit packets from their dedicated tunnel. In addition, the first malicious node would generate a higher signal strength intensity to persuade legitimate nodes to believe that they are close to the destination [45].

Teng et al. [46] present a wormhole attack detection algorithm related to the node trust optimization model against wormholes in wireless sensor networks (WSNs). The proposed method owns a high detection 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.

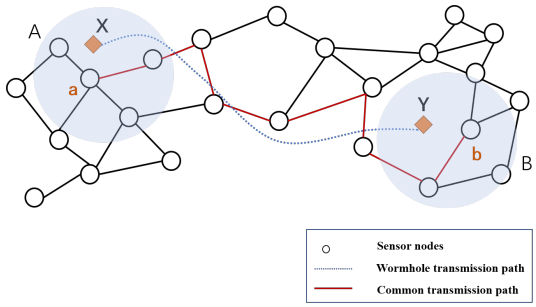


Fig. 4: Wormhole attack

b) SQL injection attack: Many CPSs still rely on databases for data management. Structured query language (SQL) injection attacks are commonly used by hackers to attack databases, and attackers can access data records without authorization. SQL comes from many different dialects, but most are based on the SQL-92 ANSI standard [34]. 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 flexible. 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 [47].

Halfond et al. [48] mention 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

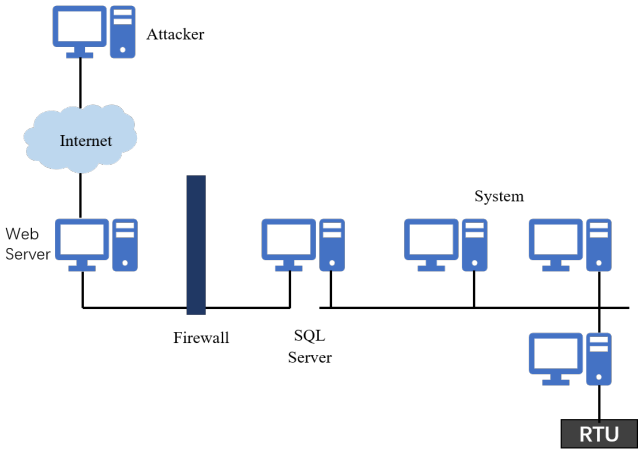


Fig. 5: SQL Attack

significance to the security of a SCADA system [34]. In [49], 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 (SQLIA), Gowtham and Pramod [50] propose SQLIA-prediction systems by using semantic features combined with the highly robust computing environment. Moreover, to alleviate computational burden, the authors introduce two feature selection algorithms called Mann-Whitney significance predictor test and principal component analysis.

This work in [51] 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 intelligence.

c) DoS (Denial of Service) attack: A denial-of-service attack [52], [53] is a kind of resource exhaustion attack that makes communication networks or servers unable to provide services by using the defects of software or communication protocol or by sending a large number of useless requests to exhaust the server's resources [54]. In [55], some examples of DoS attacks that occur in CPSs are described. The work in [56] presents different types of DoS attacks.

A more serious DoS attack is a distributed DoS (DDoS) attack. In 2016, US attackers launched the largest distributed DoS attack on the Dyn server through small CPSs and Internet of Things devices, causing downtime for Twitter, Cable News

Network (CNN), and the Guardian [35].

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 actuators, thus causing CPSs to be out of control [57]. Similarly, Koscher et al. [58] propose a DoS attack applied to intelligent 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.

The article in [59] provides a structured and comprehensive survey of the existing application layer DoS attacks and defense mechanisms. The article classifies the existing attacks and defense mechanisms into different categories, describes how they work, and compares them based on relevant parameters.

d) False data injection attack: Another potential threat to CPSs is false data injection attack (FDIA) [60], [61]. This type of attack mainly involves an attacker injecting false sensor data into a sensor or transmitting false data to trigger a malicious event [35]. Fig. 6 shows the process. The FDIA was originally

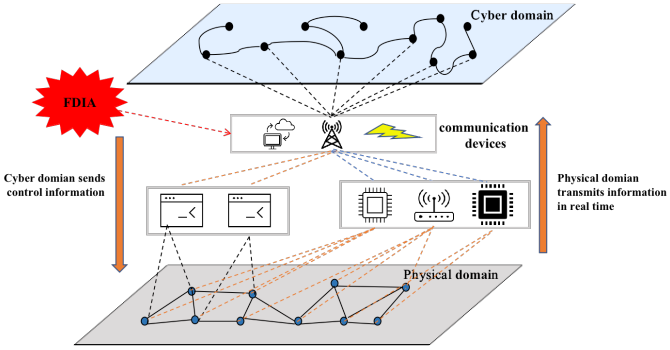


Fig. 6: False data injection attack

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 successful FDIA may cause a state estimator to send an error message to system operators, resulting in a series of impacts on power systems [62]. 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, in [64], Hichem et al. mention that drones located in the same neighborhood should report the same phenomenon. However, malicious drones may disrupt sensor readings and cause erroneous physical phenomena. This attack is usually directed against CPSs with wireless sensor networks [65]. Injecting false data into smart grid traffic can lead to different consequences, such as service interruption and financial losses [66]. Some researchers have put forward other FDIA attacks against data integrity in CPSs [67]–[69].

Lu and Yang [70] study the stealthy false data injection attack design problem for CPSs that has state estimators and attack detectors. The work obtains a necessary and sufficient condition for the existence 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.

e) 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 [71] proposes 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 industrial control system (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 traffic [72]. Yu et al. [73] present a malware propagation model in CPSs, namely SEI^{2}RS, which considers two infectious rates. The equilibria are calculated, and the stability, bifurcation 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 [66]. The work in [74] 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.

f) 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 [35]. 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 consequences [75]. The attack process is shown in Fig. 7. In [76], Melamed discusses 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.

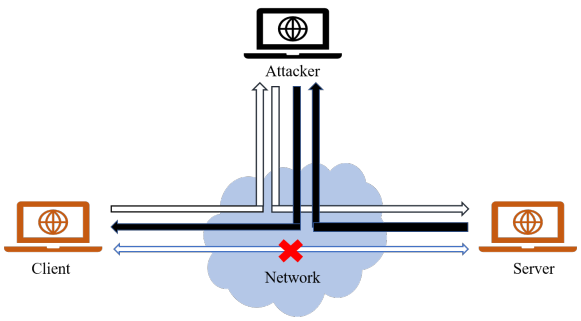


Fig. 7: Man-in-the-middle attack

The commonly used techniques for the man-in-the-middle

attacks are packet injection, session hijacking, and SSL stripping [35]. Akter et al. [77] establish MITM attacks in near field communication (NFC) between a passive tag and an active terminal, illustrate the possibility that the designed attack can compromise the process of a contactless payment via a malicious MITM card, and also show the impacts of the MITM attacks on attack/victim scenarios.

g) 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 [35]. The work [78] tackles three problems in GPS spoofing attack: multiattack detection on different phasor measurement unit (PMU), attack detection about the dynamic model of power systems, and measurement correction. The results are illustrated for the detection method in the PMU and supervisory control and data acquisition systems.

h) Eavesdropping: Eavesdropping refers to an attack in which an adversary can intercept information communicated by a system [79]. In CPSs, control information may be monitored during the transmission from a sensor to a server [3]. In addition, it is possible to intercept the monitoring data transmitted by sensor networks collected by monitoring via traffic analysis.

In [80], Balakrishnan et al. introduce two new types of eavesdropping attacks based on a next-generation wireless communication network, i.e., opportunistic stationary attacks and active nomadic attacks, and study the success probability of these two attacks.

Wang et al. [81] study the security issues to a CPS under eavesdropping attacks. For a network system that is attacked by eavesdropping, the researchers establish necessary and sufficient conditions for an eavesdropper to carry out observations in CPSs.

Wu et al. [82] study eavesdropping and anti-eavesdropping relations between a UAV-enabled eaves-dropper (UAV-E) and a UAV-enabled base station (UAV-BS) in a downlink wiretap 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.

2) Intelligent system security threat: In recent years, the rapid development of artificial intelligence technology has made CPSs more intelligent, which brings many new security 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 [83]. The work in [84] 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, adversarial attacks, extraction attacks, and inversion attacks, which are shown in Fig. 8.

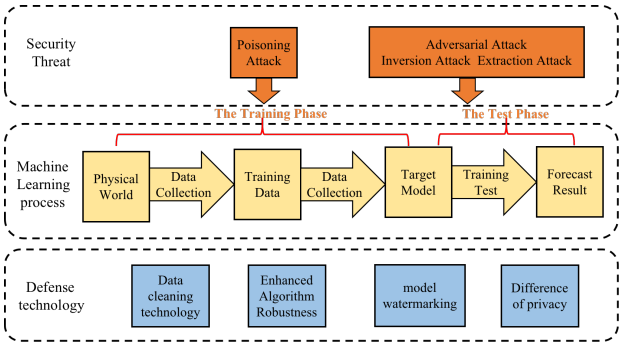


Fig. 8: Intelligent system security threat

a) Poisoning attacks: In poisoning attacks, an attacker modifies data and distribution to affect training results of an artifical intelligence model in CPSs [85].

Generally, using various methods to gain unauthorized access to data, attackers can mark enough data points to tamper with training data to obtain desired effects. Yang et al. [86] contaminate 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 instigation of some racist netizens, Microsoft's chat robot eventually turns into a racist and foul-mouthed robot [87].

In [88], 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 [89] reviews existing poisoning attacks and countermeasures in intelligent networks, compares the principles of different types of poisoning attacks, and analyzes the advantages and drawbacks of defense methods.

b) Adversarial attacks: Most of 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 [85].

By constructing an adversarial sample, an attacker can interfere with reasoning process of artificial intelligence services to achieve attacks such as evasion detection. In [90], researchers design 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 [91] and nontarget attacks [92] according to the attack effect; based on an attacker's ability,

attacks can be classified as white-box attacks [93] and black-box attacks [94]. Kumar et al. [95] conduct an empirical study on speech error interpretation attacks in speech systems.

This article in [96] carefully discusses different types of adversarial attacks and corresponding defense strategies, concluding that adversarial learning is the real threat to machine learning in applications.

c) 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 [85].

Lowd and Meek [97] propose an algorithm to steal the parameters of a linear classifier model. Based on the principle that the parameters learned by the machine learning model can minimize the cost function, Wang et al. [98] present a hyperparameter estimation method for the machine learning model. Yan et al. [98] introduce an model extraction attack that are used for stealing confidential information of the learning models through public queries, and optimize the attack behaviour 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.

d) 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 [85]. Attackers can pry into the privacy of the training data based on the difference in fitting between the training data and non-training data.

The attribute reasoning attacks [85] mainly obtain attribute information such as age distribution, prevalence, and income distribution of the data set. For instance, Fredrikson et al. [100] elaborate on the inversion attack through the issue of privacy in medical machine learning. Specifically, attackers try to infer the patient's genotype based on the warfarin dosage information.

The member reasoning attack [85] mainly infers whether a specific record appears in the data set. Truex et al. [45] propose 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 [101], [102], probability information calculation [103], and similar sample generation [104].

Alufaisan et al. [105] introduce a novel technique that complements 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.

C. Cyber-physical attack

In [106], Valise and Miller refer to cyber-physical attacks as cyber-attacks "that result in physical control of various aspects" of a cyber-physical system. However, Yam et al define them more generally as cyber-attacks with "physical effect propagation". A more general definition is put forward in [107]. Researchers consider that a cyber-physical attack as a security 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 [108] that exploits software vulnerabilities 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.

1) Malicious destruction attack: Malicious damage can occur through malware injection. In smart cars, malware injection through an OBD-II port requires physical access to a car. Hoppe et al. [109] show 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. [110] conduct 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 connections and then sends a malicious payload to the Transmission 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, Woo et al. [111] 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 [110]. 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 [107].

2) Trojan attack: A Trojan virus refers to a piece of malicious code with special functions hidden in normal programs. In [38], 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 [112], the explosion 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 [113] provides attack methodologies to neural-architecture-search (NAS) enabled edge devices for identifying NAS's vulnerability to trojaning attacks and interpret the backdoor attack, and it illustrates that the occurrence of high impact nodes decreases the robustness of the systems.

3) 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 generated 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 [114].

4) 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 [115]. As shown in Fig. 9, an attacker captures the authentication of one or more sessions.

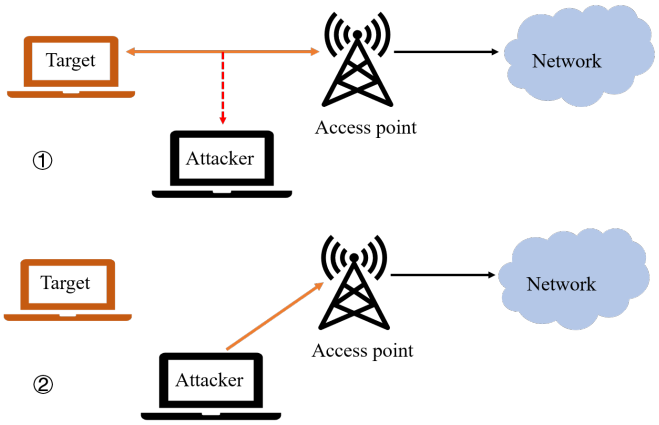


Fig. 9: Replay attack

Attackers replay an authenticated session, or use multiple sessions to synthesize the authentication portion of the session. Since the session is valid, the attacker can establish an authenticated session. Koscher et al. [58] successfully disable 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 [116]. In addition, replay attacks may lead to incorrect decisions about insulin injection [117]. 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. [118] tackle the problem of replay attack detection 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.

5) 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 [119]. A backdoor may be one of the main security problems of CPSs. Backdoors can be created by programmers during software development stages. Backdoors can also be created by attackers. A common way to create an application backdoor is to use a Trojan horse [35].

The work in [120] 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 paper 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 [66]. 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 [121].

B. Cyber Domain Attack Defense

1) Cyber attacks defense: Defense is very important for the security in the cyber domain. The paper in [122] provides 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 control approaches in real-world CPSs, namely, water distribution systems and wide-area power systems, to provide a preliminary analytical framework for modern infrastructure security. For attacks in the cyber domain mentioned above, we propose corresponding defense and detection methods as follows.

a) 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. [44] propose 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 [43], Hu and Evans propose a strategy to use directional antennas. In this proposed cooperation agreement, nodes share direction information to prevent wormhole endpoints from

TABLE II: Cyber attack defense summary

Attack Type	Defense	References
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.	[40-43]
SQL injection attacks	Defense coding, SQLIV detection, and SQLIA runtime prevention	[48],[113],[114]
DoS attacks	An elastic model predictive control (MPC) framework, a queuing model, Bernoulli model and Markov model	[115],[116]
False data injection attack	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]

being disguised as fake neighbors. The defense proposed in the article greatly reduces the threat of wormhole attacks and does not require location information or clock synchronization.

Tun et al. [123] propose wormhole detection mechanisms 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 introducing 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. [124] propose a local neighborhood information detection method based on a transmission range. Simulation results show that this method can effectively detect wormhole attacks.

b) SQL injection attack defense: SQL injection defense methods can be roughly divided into three categories: defense coding, SQLIV detection, and SQLIA runtime prevention [125]. 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 [69]. Halfond et al. propose a series of techniques to prevent SQL injection attacks, such as new query development paradigms, proxy filters and instruction set randomization [48]. In [126], Musaab et al. present a heuristic algorithm based on machine learning to prevent SQL injection attacks. The article uses a dataset containing a large number of statements to train different machine learning classifiers, 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.

c) DoS attack defense: In [127], Agah et al. 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 significantly increases the success rate of wireless sensor and actor networks in defense strategies.

In [128], Sun et al. propose an elastic model predictive control (MPC) framework. This system can mitigate the adverse effects of DoS attacks on CPSs by modeling linear time-invariant systems. Ding et al. [129] investigates 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 filtering 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 paper in [130] has investigated the maximum correntropy 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 deception 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.

d) False data injection attack defense: It is difficult to defend against FDIAs due to their concealment [131]. Two FDIA detection methods are proposed in [132]: 1) state-estimation based detection, and 2) machine-learning based detection.

In [133], Wang et al. propose two methods to defend against false data injection attacks. One is a false data filtering scheme based on geographic 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.

e) Man-in-the-middle attack defense: In [35], Ahmad et al. use a private network in CPSs to prevent man-in-the-middle attacks. Lima et al. study a man-in-the-middle attack in [134]. 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 man-in-the-middle attacks. To realize this model, the paper develops a plant model under sensor attack and a supervisor model under actuator attack.

f) Malware attack defense: The rapid growth of malware has caused very large economic losses for various organizations. 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. Joseph and Errin [135] introduce 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 segmentation to isolate machines to maximize the availability of related machines while preventing attacks.

In [136], unsupervised learning and supervised learning are used to classify malware, and machine learning algorithms and deep learning models are used to analyze and detect malware. The article uses methods such as cross-validation and fixing class imbalance problem to build models that ultimately increase the accuracy rate significantly.

g) Spoofing attack defense: Spoofing could be avoided by packet filtering or by using a secure encryption protocol. The prevention of these attacks includes DVCerts and DAPS [35].

In [137], Zeng et al. propose a malicious host detection algorithm based on 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.

In [138], Gao et al. use 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.

2) 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.

a) Poisoning attack defense: For data poisoning attacks, current defense methods are mainly divided into two types: the data cleaning technology and 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 [139]. An attack detection strategy is proposed to detect potential contamination by isolating a special holdout set. In [88], Baracaldo et al. use source information as part of a filtering algorithm to detect poison attacks. They use the source of training data points and transform 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 poisoning attacks. For partially trusted and completely untrusted datasets, the authors propose two variants of source defense.

A learning algorithm always has to make a trade-off 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. propose improving a PCA algorithm and reduce the influence of malicious training data by combining the PCA with the Laplacian truncation threshold [87].

In [140], Jagielski et al. propose 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.

b) Adversarial attack defense: The defense methods of adversarial attacks mainly focus on preventing the generation of confrontation samples and the detection of confrontation samples [141].

In [141], 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.

In [142], McDaniel et al. use 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 [92].

c) Extraction attacks defense: The defense strategy for model extraction attacks is mainly to approximate model parameters [143] or output results [144]. In addition, to avoid the model from being stolen to protect intellectual property rights, some researchers have proposed the concept of model watermarking [145].

The researchers in [146], [147] add a watermark to the neural network by adding a new regularization term to the loss function. Merrer et al. [148] combine adversarial examples and adversarial training methods to inject watermarks into neural networks. Adi et al. [149] study a black-box deep neural network watermarking technology, which proved through experiments that this method does not affect the performance of the original model.

In [101], researchers 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

TABLE III: Intelligent system attack defense summary

Type	Defense	References
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],[137-148]
Inversion attack	Use of machine learning algorithms with privacy protection functions.	[85],[141-144]

output trusted value can reduce the success rate of model extraction [150].

d) Inversion attacks defense: A typical method of defending against inversion attacks is the use of machine learning algorithms with privacy protection functions. Currently, homomorphic encryption [151] and differential privacy technologies [152] 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. [153] propose a defense method that uses homomorphic encryption 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 [154]. Papernot et al. [91] put forward a universal PATE framework to protect training data in machine learning.

C. Cyber-physical Domain Attack Defense

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

1) Trojan attack defense: There are also many Trojans in integrated circuits, 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 integrated circuit (IC) and protect hardware Trojan. In [155], we elaborate an integrated circuit market model to illustrate the potential Trojan threat participation model faced by both parties.

2) 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 concealment, such attacks may have disastrous consequences [156]. 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.

3) Replay attack defense: Mo and Bruno [157] assume 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 [158], 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 industry, 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 challenging 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

- 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 profoundly changed the way in which humans interact with the physical world by integrating the physical environment 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 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 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.

TABLE IV: Cyber-physical attack defense summary

Types	Defense	References
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]

- 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 multi-objective optimization problem arises.

B. Future Research Directions

The size of the CPSs become large and complex, and enormous 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, security 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 carrying out more targeted and destructive attacks based on CPSs operating mechanisms and defense strategies. Although some defense mechanisms have been proposed, new defense strategies for identifying threats and vulnerabilities for specific systems still need to be updated.

With the deep integration of cyber systems and physical 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. Conclusions

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 security. In this paper, we first review the architecture and security issues of CPSs. Then, possible attacks on CPSs are classified in three aspects, i.e., physical domain, cyber domain, and cyber-physical domain. As CPSs inevitably use some intelligent algorithms, they are vulnerable to artificial intelligence attacks. Therefore, artificial intelligence attacks are added to the classification and the corresponding defenses are given. Next, for each of the above classified attacks, we give the corresponding 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 paper 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

[1] 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," *Microprocessors and microsystems*, vol. 77, 2020.

[2] S. Gries, M. Hesenius, and V. Gruhn, "Cascading data corruption: About dependencies in cyber-physical systems: Poster," in *Proceedings of the 11th ACM International Conference on Distributed and Event-based Systems*, Barcelona Spain, 2017, pp. 345–346.

[3] Y. Ashibani, and Q. H. Mahmoud, "Cyber physical systems security: Analysis, challenges and solutions," *Computers & Security*, vol. 68, pp. 81–97, 2017.

[4] U. Sendler, "Industrie 4.0–Beherrschung der industriellen Komplexit?t mit SysLM (Systems lifecycle management)," Springer Vieweg, Berlin, Heidelberg, 4rd, 2013.

[5] T. Haidegger, G. S. Virk, C. Herman, R. Bostelman, P. Galambos, G. Györök, and I. J. Rudas, "Industrial and medical cyber-physical systems: Tackling user requirements and challenges in robotics," *Topics in Intelligent Engineering and Informatics*, vol. 14, pp. 253–277, 2019.

[6] G. R. Gonzélez, M. M. Organero, and C. D. Kloos, "Early infrastructure of an internet of things in spaces for learning," in *Proceedings of 2008 eighth IEEE international conference on advanced learning technologies*, Santander, Spain, 2008, pp. 381–383.

[7] R. Baheti, and H. Gill, "Cyber-physical systems," *The impact of Control Technology*, vol. 12, pp. 161–166, 2011.

[8] S. Sastry, "Networked embedded systems: From sensor webs to cyber-physical systems," in *Proceedings of International Workshop on Hybrid Systems: Computation and Control*, Berlin, Heidelberg, 2007, pp. 1–1.

[9] E. A. Lee, "CPS foundations," in *Proceedings of Design automation conference*, Anaheim, CA, USA, 2010, pp. 737–742.

[10] A. Darwish, and A. E. Hassanien, "Cyber physical systems design, methodology, and integration: The current status and future outlook," *Journal of Ambient Intelligence and Humanized Computing*, vol. 9, pp. 1541–1556, 2018.

[11] H. Gill, "From vision to reality: Cyber-physical systems," in *Proceedings of HCSS National Workshop on New Research Directions for High Confidence Transportation CPS: Automotive, Aviation, and Rail*, USA: Austin, 2008, pp. 18–20.

[12] Q. Shafi, "Cyber physical systems security: A brief survey," in *Proceedings of 2012 12th International Conference on Computational Science and Its Applications*, Salvador, Brazil, 2012, pp. 18–21.

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

[14] D. Seifert, and H. Reza, "A security analysis of cyber-physical systems architecture for healthcare," *Computers*, vol. 5, pp. 27, 2016.

[15] Y. F. Li, D. H. Sun, W. N. Liu, and X. B. Zhang, "A service-oriented architecture for the transportation cyber-physical systems," in *Proceedings of the 31st Chinese Control Conference*, Hefei, China, 2012, pp. 7674–7678.

[16] A. Caggiano, T. Segreto, and R. Teti, "Cloud manufacturing framework for smart monitoring of machining," *Procedia Cirp*, vol. 55, pp. 248–253, 2016.

[17] R. Mahmoud, T. Yousuf, F. Aloul, and I. Zualkernan, "Internet of things (IoT) security: Current status, challenges and prospective measures," in *Proceedings of 2015 10th International Conference for Internet Technology and Secured Transactions (ICITST)*, London, UK, 2015, pp. 336–341.

[18] K. Zhao, and L. Ge, "A survey on the internet of things security," in *Proceedings of 2013 Ninth International Conference on Computational Intelligence and Security*, Emeishan, China, 2013, pp. 663–667.

[19] T. B. Lu, J. X. Lin, L. L. Zhao, Y. Li, and Y. Peng, "A security architecture in cyber-physical systems: Security theories, analysis, simulation and application fields," *International Journal of Security and Its Applications*, vol. 9, pp. 1–16, 2015.

[20] B. Zhu, and S. Sastry, "SCADA-specific intrusion detection/prevention systems: A survey and taxonomy," in *Proceedings of the 1st Workshop on Secure Control Systems (SCS)*, Stockholm, Sweden, 2010.

[21] H. H. Gao, Y. Peng, K. B. Jia, Z. H. Dai, T. Wang, "The design of ics testbed based on emulation, physical, and simulation," in *Proceedings of 2013 Ninth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Beijing, China, 2013, pp. 420–423.

[22] R. Khan, S. U. Khan, R. Zahee, and S. Khan, "Future internet: The internet of things architecture, possible applications and key challenges," in *Proceedings of 2012 10th International Conference on Frontiers of Information Technology*, Islamabad, Pakistan, 2012, pp. 257–260.

[23] B. Zhang, X. X. Ma, and Z. G. Qin, "Security architecture on the trusting internet of things," *Journal of Electronic Science and Technology*, vol. 9, pp. 364–367, 2011.

[24] C. Konstantinou, M. Maniatakos, F. Saqib, S. Y. Hu, J. Plusquellic, and Y. E. Jin, "Cyber-physical systems: A security perspective," in *Proceedings of 2015 20th IEEE European Test Symposium*, Cluj-Napoca, Romania, 2015, pp. 1–8.

[25] J. Jamaludin, and J. M. Rohani, "Cyber-physical system (CPS): State of the art," in *International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*, Quetta, Pakistan, 2018, pp. 1–5.

[26] S. Heng, "Industry 4.0: Huge potential for value creation waiting to be tapped," *Deutsche Bank Research*, pp. 8–10, 2014.

[27] T. B. Lu, J. X. Lin, L. L. Zhao, Y. Li, and Y. Peng, "An analysis of cyber physical system security theories," in *Proceedings of 2014 7th International Conference on Security Technology*, Hainan, China, 2014, pp. 19–21.

[28] 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, 2019.

[29] E. B. Harb, "A brief survey of security approaches for cyber-physical systems," in *Proceedings of 2016 8th IFIP International Conference on New Technologies, Mobility and Security*, Larnaca, Cyprus, 2016, pp. 1–5.

[30] N. Sklavos, and I. D. Zaharakis, "Cryptography and security in internet of things: Models, schemes, and implementations," in *Proceedings of 2016 8th IFIP International Conference on New Technologies, Mobility and Security*, Larnaca, Cyprus, 2016, pp. 1–2.

[31] H. Yoo, and T. Shon, "Challenges and research directions for heterogeneous cyber–physical system based on IEC 61850: Vulnerabilitie s, security requirements, and security architecture," *Future Generation Computer Systems*, vol. 61, pp. 128–136, 2016.

[32] R. Alguliyev, Y. Imamverdiyev, and L. Sukhostat, "Cyber-physical systems and their security issues," *Computers in Industry*, vol. 100, pp. 212–223, 2018.

[33] D. R. Ding, Q. L. Han, Y. Xiang, X. H. Ge, and X. M. Zhang, "A survey on security control and attack detection for industrial cyber-physical systems," *Neurocomputing*, vol. 275, pp. 1674–1683, 2018.

[34] B. Zhu, A. Joseph, and S. Sastry, "A taxonomy of cyber attacks on SCADA systems," in *Proceedings of 2011 International Conference on Internet of Things and 4th International Conference on Cyber, Physical and Social Computing*, Dalian, China, 2011, pp. 380–388.

[35] I. Ahmad, M. K. Zarrar, T. Saeed, and S. Rehman, "Security aspects of cyber physical systems," in *Proceedings of 2018 1st International Conference on Computer Applications and Information Security*, Riyadh, Saudi Arabia, 2018, pp. 1–6.

[36] Powering Business Worldwide Eaton. "Power Outage Annual Report: Blackout Tracker," Available at: Retrieved from http://www.eaton.com/blackouttracker, 2014.

[37] J. Wurm, Y. E. Jin, Y. Liu, S. Y. Hu, K. Heffner, F. Rahman, and M. Tehranipoor, "Introduction to cyber-physical system security: A cross-layer perspective," *IEEE Transactions on Multi-Scale Computing Systems*, vol. 3, pp. 215–227, 2016.

[38] A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, and S. Sastry, "Challenges for securing cyber physical systems," in *Proceedings of Future Directions in Cyber-Physical Systems Security*, 2009.

[39] P. Maheshwari, "Security issues of cyber physical system: A review," in *Proceedings on National Conference on Advances*, 2016, pp. 7–11.

[40] M. Rushanan, A. D. Rubin, D. F. Kune, and C. M. Swanson, "SoK: Security and privacy in implantable medical devices and body area networks," in *Proceedings of 2014 IEEE symposium on security and privacy*, Berkeley, CA, USA, 2014, pp. 524–539.

[41] I. Butun, P. Österberg, and H. B. Song, "Security of the internet of things: Vulnerabilities, attacks, and countermeasures," *IEEE Communications Surveys & Tutorials*, vol. 22, pp. 616–644, 2019.

[42] L. George, K. Eirini, P. Emmanouil, S. Panagiotis, B. Anatolij, and T. Vuong, "A taxonomy and survey of cyber-physical intrusion detection approaches for vehicles," *Ad Hoc Networks*, vol. 84, pp. 124–147, 2019.

[43] L. X. Hu, and D. Evans, "Using Directional Antennas to Prevent Wormhole Attacks," in *Proceedings of the Network and Distributed System Security Symposium*, SanDiego, California, USA, 2004, pp. 241–245.

[44] Y. C. Hu, A. Perrig, and D. B. Johnson, "Packet leashes: A defense against wormhole attacks in wireless networks," in *Proceedings of IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies*, San Francisco, CA, 2003, pp. 1976–1986.

[45] S. Truex, L. Liu, M. E. Gursoy, L. Yu, and W. Q. Wei, "Towards demystifying membership inference attacks," Available at: https://arxiv.org/abs/1807.09173, 2018.

[46] Z. J. Teng, C. Q. Du, M. Li, H. J. Zhang, and W. H. Zhu, "A wormhole attack detection algorithm integrated with the node trust optimization model in WSNs," *IEEE Sensors Journal*, vol. 22, no. 1, pp. 7361–7370, 2022.

[47] V. N. Gudivada, S. Ramaswamy, and S. Srinivasan, "Data management issues in cyber-physical systems," *Transportation Cyber-Physical Systems*, pp. 173–200, 2018.

[48] W. G. J. Halfond, J. Viegas, and A. Orso, "A classification of SQL-injection attacks and countermeasures," in *Proceedings of the IEEE International Symposium on Secure Software Engineering*, 2006, pp. 13–15.

[49] T. Paukatong, "SCADA Security: A New Concerning Issue of an In-house EGAT-SCADA," in *Proceedings of 2005 IEEE/PES Transmission & Distribution Conference & Exposition: Asia and Pacific*, Dalian, China, 2005, pp. 1–5.

[50] M. Gowtham and H. B. Pramod, "Semantic query-featured ensemble learning model for SQL-injection attack detection in IoT-ecosystems," *IEEE Transactions on Reliability*, vol. 71, no. 2, 1057-1074, 2022.

[51] M. Alghawazi, D. Alghazzawi, and S. Alarifi, "Detection of SQL injection attack using machine learning techniques: A systematic literature review," *Journal of Cybersecurity and Privacy*, vol. 2, pp. 764-777, 2022.

[52] A. D. Wood, and J. A. Stankovic, "Denial of service in sensor networks," *Computer*, vol. 35, pp. 54–62, 2002.

[53] W. Y. Xu, K. Ma, W. Trappe, and Y. Y. Zhang, "Jamming sensor networks: Attack and defense strategies," *IEEE Network*, vol. 20, pp. 41–47, 2006.

[54] P. Srikantha, and D. Kundur, "Denial of service attacks and mitigation for stability in cyber-enabled power grid," in *Proc. IEEE Power Energy*

Soc. Innov. Smart Grid Technol. Conf., Washington, DC, USA, 2015, pp. 1–5.

[55] M. Long, C. H. Wu, and J. Y. Hung, "Denial of service attacks on network-based control systems. impact and mitigation," *IEEE Transactions on Industrial Informatics*, vol. 1, pp. 85–96, 2005.

[56] A. Arış, S. F. Oktuğ, and S. B. Ö. Yalçın, "Internet-of-Things security: Denial of service attacks," in *Proceedings of 2015 23nd Signal Processing and Communications Applications Conference*, Malatya, Turkey, 2015, pp. 903–906.

[57] L. W. Cao, X. N. Jiang, Y. M. Zhao, S. G. Wang, D. You, and X. L. Xu, "A survey of network attacks on cyber-physical systems," *IEEE Access*, vol. 8, pp. 44219–44227, 2020.

[58] K. Koscher, A. Czeskis, F. Roesner, S. Patel, T. Kohno, S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, and S. Savage, "Experimental security analysis of a modern automobile," in *Proceedings of 2010 IEEE Symposium on Security and Privacy*, Oakland, CA, USA, 2010, pp. 447–462.

[59] T. Nikhil and H. Neminath, "Application Layer Denial-of-Service Attacks and Defense Mechanisms: A Survey," *ACM Computing Surveys*, vol. 54, no. 4, pp. 1–33, 2022.

[60] A. Asmaa, and X. M. Shen, "Efficient prevention technique for false data injection attack in smart grid," in *Proceedings of 2016 IEEE International Conference on Communications*, Kuala Lumpur, Malaysia, 2016, pp. 22–27.

[61] J. Wei, and G. J. Mendis, "A deep learning-based cyber-physical strategy to mitigate false data injection attack in smart grids," in *Proceedings of 2016 Joint Workshop on Cyber- Physical Security and Resilience in Smart Grids*, Vienna, Austria, 2016, pp. 12–12.

[62] G. Q. Liang, J. H. Zhao, F. J. Luo, S. R. Weller, and Z. Y. Dong, "A review of false data injection attacks against modern power systems," *IEEE Transactions on Smart Grid*, vol. 8, pp. 1630–1638, 2016.

[63] Y. L. Mo and B. Sinopoli, "False data injection attacks in control systems," in *Proceedings of Secure Control Systems*, 2010, pp. 1–6.

[64] H. Sedjelmaci, S. M. Senouci, and N. Ansari, "A hierarchical detection and response system to enhance security against lethal cyber-attacks in UAV networks," in *Proceedings of the IEEE International Symposium on Secure Software Engineering*, 2017, pp. 1594–1606.

[65] Y. L. Mo, E. Garone, A. Casavola, and B. Sinopoli, "False data injection attacks against state estimation in wireless sensor networks," in *Proceedings of 49th IEEE Conference on Decision and Control*, Atlanta, GA, USA, 2010, pp. 15–17.

[66] A. Humayed, J. Q. Lin, F. J. Li, and B. Luo, "Cyber-physical systems security–A survey," *IEEE Internet of Things Journal*, vol.4, pp. 1802–1831, 2017.

[67] J. Lin, W. Yu, and X. Y. Yang, "On false data injection attack against multistep electricity price in electricity market in smart grid," in *Proceedings of 2013 IEEE Global Communications Conference*, Atlanta, GA, USA, 2014, pp. 9-13.

[68] Q. Y. Yang, L. G. Chang, and W. Yu, "On false data injection attacks against kalman filtering in power system dynamic state estimation," *Security and Communication Networks*, vol. 9, pp. 833–849, 2013.

[69] Q. Y. Yang, J. Yang, W. Yu, D. An, N. Zhang, and W. Zhao, "On False data-injection attacks against power system state estimation: Modeling and countermeasures," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, pp. 717–729, 2013.

[70] A. Y. Lu and G. H. Yang, "False data injection attacks against state estimation without knowledge of estimators," *IEEE Transactions on Automatic Control*, vol. 67, no. 9, pp. 4529–4540, 2022.

[71] B. Min, and V. Varadharajan, "Design and evaluation of feature distributed malware attacks against the internet of things (IoT)," in *Proceedings of 2015 20th International Conference on Engineering of Complex Computer Systems*, Gold Coast, QLD, Australia, 2016, pp. 9–12.

[72] K. Munro, "Deconstructing Flame: the limitations of traditional defences," *Computer Fraud and Security*, vol. 2012, pp. 8–11, 2012.

[73] Z. Yu, H. Gao, D. Wang, A. A. Alnuaim, M. Firdausi, A. M. Mostafa, "SEI^{2}RS malware propagation model considering two infection rates in cyber–physical systems," *Physica A: Statistical Mechanics and its Applications*, 2022, vol. 597, 127207.

[74] A. A. Elkhail, R. Refat, R. Habre, A. Hafeez, A. Bacha, and H. Malik, "Vehicle security: A survey of security issues and vulnerabilities, malware attacks and defenses," *IEEE Access*, vol. 9, pp. 162401–162437, 2021.

[75] S. Ali, R. W. Anwar, and O. K. Hussain, "Cyber security for cyber physcial systems: A trust-based approach," *Cyber Security for Cyber Physical Systems*, vol. 71, pp. 144–152, 2015.

[76] T. Melamed, "An active man-in-the-middle attack on bluetooth smart devices," *International Journal of Safety and Security Engineering*, vol. 8, pp.200–211, 2018.

[77] S. Akter, S. Chellappan, T. Chakraborty, T. A. Khan, A. Rahman, and A. Islam, "Man-in-the-middle attack on contactless, payment over NFC communications: design, implementation, experiments and detection," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 6, pp. 3012–3023, 2021.

[78] S. Siamak, M. Dehghani, and M. Mohammadi, "Dynamic GPS spoofing attack detection, localization, and measurement correction exploiting PMU and SCADA," *IEEE Systems Journal*, vol. 15, no. 2, pp. 2531–2540, 2021.

[79] J. C. Kao and R. Marculescu, "Eavesdropping minimization via transmission power control in ad-hoc wireless networks," in *Proceedings of 2006 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks*, Reston, VA, USA, 2007, pp. 707–714.

[80] S. Balakrishnan, P. Wang, A. Bhuyan, and Z. Sun, "Modeling and analysis of eavesdropping attack in 802.11ad mmWave wireless networks," *IEEE Access*, vol. 7, pp. 70355–70370, 2019.

[81] W. Yang, Z. Q. Zheng, G. R. Chen, Y. Tang, and X. F. Wang, "Security analysis of a distributed networked system under eavesdropping attacks," *J. IEEE T CIRCUITS-II*, vol. 67, pp. 1254–1258, 2019.

[82] H. C. Wu, M. Li, Q. Y. Gao, Z. Q. Wei, N. Zhang, and X. F. Tao, "Eavesdropping and anti-eavesdropping game in UAV wiretap system: A differential game approach," *IEEE Transactions on Wireless Communications*, vol. 22, no. 11, pp. 9906–9920, 2022.

[83] Ubers self-driverings death [Online]. Available at: https://companies/self-driving-uber- death/index.html

[84] J. Li, Y. Liu, T. Chen, Z. Xiao, Z. J. Li, and J. P. Wang, "Adversarial attacks and defenses on cyber–pysical systems: A survey," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 5103–5115, 2020.

[85] Y. Z. He, X. B. Hu, J. W. He, G. Z. Meng, and K. Chen, "Privacy and security issues in machine learning systems: A survey," *J. Journal of Computer Research and Development*, vol. 56, 2019.

[86] G. L. Yang, N. Z. Gong, and Y. Cai, "Fake co-visitation injection attacks to recommender systems," in *Proceedings of Network and Distributed System Security Symposium*, Iowa State University, Ames, 2017.

[87] B Biggio, I Corona, G Fumera, G Giacinto, and F Roli, "Bagging classifiers for fighting poisoning attacks in adversarial classification tasks," in *Proceedings of the 10th international conference on Multiple classifier systems*, Naples, Italy, 2011, pp. 350–359.

[88] N. Baracaldo, B. Chen, H. Ludwig, and J. A. Safavi, "Mitigating poisoning attacks on machine learning models: A data provenance based approach," in *Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security*, New York, USA, 2017, pp. 103–110.

[89] C. Wang, J. Chen, Y. Yang, X. Q. Ma, and J. C. Liu, "Poisoning attacks and countermeasures in intelligent networks: Status quo and prospects," *Digital Communications and Networks*, vol. 8, pp. 225–234, 2022.

[90] A. Kurakin, I. Goodfellow, and S. Bengio, "Adversarial examples in the physical world," Available at: https://arxiv.org/abs/1607.02533, 2016.

[91] N. Papernot, P. McDaniel, S. Jha, M. Fredrikson, Z. B. Celik, and A. Swami, "The limitations of deep learning in adversarial settings," in *Proceedings of 2016. IEEE European Symposium on Security and Privacy*, Saarbruecken, Germany, 2016, pp. 372–387.

[92] I. J. Goodfellow, J. Shlens ,and C. Szegedy, "Explaining and harnessing adversarial examples," Available at: https://arxiv.org/abs/1412.6572, 2014.

[93] S. M. Moosavi-Dezfooli, A. Fawzi, and P. Frossard, "Deepfool: A simple and accurate method to fool deep neural networks," in *Proceedings of the. IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2017, pp. 2574–2582.

[94] N. Papernot, P. McDaniel, I. Goodfellow, S. Jha, Z. B. Celik, and A. Swami, "Practical black-box attacks against machine learning," in *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*, Abu Dhabi, United Arab Emirates, 2017, pp. 506–519.

[95] D. Kumar, R. Paccagnella, P. Murley, E. Hennenfent, J. Mason, A. Bates, and M. Bailey, "Skill squatting attacks on Amazon Alexa," in *Proceedings of 27th. USENIX conference on Security Symposium*, Baltimore, MD, USA, 2018, pp. 33–47.

[96] A. Chakraborty, M. Alam, V. Dey, A. Chattopadhyay, and D. Mukhopadhyay, "A survey on adversarial attacks and defenses," *CAAI Transactions on Intelligence Technology*, vol. 6, 25-45, 2021.

[97] D. Lowd and C. Meek, "Adversarial learning," in *Proceedings of the 11th. ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, Illinois, Chicago, USA, 2005, pp. 641–647.

[98] B. H. Wang and N. Z. Gong, "Stealing hyperparameters in machine learning," in *Proceedings of 39th. IEEE Symposium on Security and Privacy*, San Francisco, CA, USA, 2018, pp. 36–52.

[99] H. N. Yan, X. G. Li, H. Li, J. M. Li, W. H. Sun, and F. H. Li, "Monitoring-based differential privacy mechanism against query flooding-based model extraction attack," *IEEE Transactions on Dependable and Secure Computing*, vol. 19, no. 4, pp. 2680–1694, 2022.

[100] M. Fredrikson, E. Lantz, S. Jha, and S. M. Lin, "Privacy in pharmacogenetics: An end-to-end case study of personalized warfarin dosing," in *Proceedings of 23rd. USENIX conference on Security Symposium*, San Diego, United States, 2014, pp. 17–32.

[101] R. Shokri, M. Stronati, C. Z. Song, and V. Shmatikov, "Membership inference attacks against machine learning models," in *Proceedings of 2017 IEEE Symposium on Security and Privacy*, San Jose, 2017, pp. 3–18.

[102] A. Pyrgelis, C. Troncoso, and E. De Cristofaro, "Knock knock, who's there? Membership inference on aggregate location data," Available at: https://arxiv.org/abs/1708.06145, 2017.

[103] Y. H. Long, V. Bindschaedler, L. Wang, D. Y. Bu, X. F. Wang, H. X. Tang, and K. Chen, "Understanding membership inferences on well-generalized learning models," Available at: https://arxiv.org/abs/1802.04889, 2018.

[104] K. S. Liu, B. Li ,and J. Gao, "Generative model: Membership attack, generalization and diversity," Available at: https://arxiv.org/abs/1805.09898, 2018.

[105] Y. Alufaisan, M. Kantarcioglu, and Y. Zhou, "Robust transparency against model inversion attacks," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 5, pp. 2061–2073, 2021.

[106] C. Miller and C. Valasek, "A survey of remote automotive attack surfaces," *black hat USA*, vol. 2014, 2014.

[107] A. Ashok, M. Govindarasu ,and J. Wang, "Cyber-Physical Attack-Resilient Wide-Area Monitoring, Protection, and Control for the Power Grid," *Proceedings of the IEEE*, vol. 105, no. 7, pp. 1389–1407, 2017.

[108] R. Langner, "Stuxnet: Dissecting a cyberwarfare weapon," *IEEE Security & Privacy*, vol. 9, no. 3, pp. 49–51, 2011.

[109] T. Hoppe, S. Kiltz, and J. Dittmann, "Security threats to automotive CAN networks–Practical examples and selected short-term countermeasures," in *Proceedings of International Conference on Computer Safety, Reliability, and Security*, United Kingdom, 2008, pp. 235-248.

[110] S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, T. Kohno, and S. Savage, "Comprehensive experimental analyses of automotive attack surfaces," in *Proceedings of USENIX Security Symposium*, San Francisco, 2011, pp. 447–462.

[111] W. Samuel, J. H. Jin, and L. D. Hoon, "A practical wireless attack on the connected car and security protocol for in-vehicle CAN," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, pp. 993C-1006, 2015.

[112] S. Jill and M. Michael, "Lessons learned from the maroochy water breach," in *Proceedings of International conference on critical infrastructure protection*, Hanover, NH, USA, 2007, pp. 73–82.

[113] S. P. Xu, K. Wang, M. Hassan, M. M. Hassan, and C. M. Chen, "An interpretive perspective: adversarial trojaning attack on neural-architecture-search enabled edge AI systems," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 1, pp. 503–510, 2023.

[114] O. Hamed and A. M. Abdollahi, "Evaluating the complexity and impacts of attacks on cyber-physical systems, in *Proceedings of CSI Symposium on Real-Time and Embedded Systems and Technologies*, Tehran, Iran, 2015, pp. 1–8.

[115] H. Mehdi, S. Bruno, and G. Emanuele, "Feasibility and detection of replay attack in networked constrained cyber-physical systems, in *Proceedings of 2019 57th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, 2019, pp. 712–717.

[116] C. X. Li, A. Raghunathan, and N. K. Jha, "Hijacking an insulin pump: Security attacks and defenses for a diabetes therapy system," in *Proceedings of 2011 IEEE 13th International Conference on e-Health Networking, Applications and Services*, Columbia, MO, USA, 2011, pp. 150–156.

[117] J. Radcliffe, "Hacking medical devices for fun and insulin: Breaking the human SCADA system," in *Proceeding of Black Hat Conference presentation slides*, 2011.

[118] A. Naha, A. Teixeira, A. Ahlén, and S. Dey, "Sequential detection of replay attacks," *IEEE Transactions on Automatic Control*, doi: 10.1109/TAC.2022.3174004, 2022.

[119] O. Nasser, S. AlThuhli, M. Mohammed, R. AlMamari, and F. Hajamohideen, "An investigation of backdoors implication to avoid regional security impediment," in *Proceedings of 2015 Global Conference on Communication Technologies*, Thuckalay, India, 2015, pp. 409–412.

[120] B. Y. Hou, J. Q. Gao, X. J. Guo, T. Baker, Y. Zhang, Y. L. Wen, and Z. L. Liu, "Mitigating the backdoor attack by federated filters for industrial IoT applications," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 5, pp. 3562–3571, 2022.

[121] V. Y. Pillitteri, and T. L. Brewer, "Guidelines for Smart Grid Cyber Security," *Privacy & the Smart Grid*, 2010.

[122] D. R. Ding, Q. L. Han, X. H. Ge, and J. Wang, Secure state estimation and control of cyber-physical systems: A survey, IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 51, no. 1, pp. 176-190, 2021.

[123] Z. Tun, and A. H. Maw, "Wormhole Attack Detection in Wireless Sensor Networks," *World Academy of Science, Engineering and Technology*, vol.46, 2008.

[124] G. W. Wu, X. J. Chen, L. Yao, Y. Lee, and K. Yim, "An efficient wormhole attack detection method in wireless sensor networks," *Computer Science and Information Systems*, vol. 11, pp. 1127–1141, 2014.

[125] S. L. Khin and T. H. B. Kuan, "Defeating SQL injection." *Computer*, vol. 46, no. 3, pp. 69–77, 2012.

[126] H. Musaab, B. Zayed, and T. Mohammed, "Detection of SQL injection attacks: A machine learning approach," in *Proceedings of 2019 International Conference on Electrical and Computing Technologies and Applications(ICECTA)*, Ras AI Khaimah, United Arab Emirates, 2019, pp. 1–6.

[127] A. Agah, K. Basu, and S. K.Das, "Preventing DoS attack in sensor networks: A game theoretic approach," in *Proceedings of IEEE International Conference on Communications*, Seoul, Korea(South), 2005, pp. 3218–3222.

[128] Q. Sun, K. W. Zhang, and Y. Shi, "Resilient model predictive control of cyber–physical systems Under DoS attacks," *IEEE Transactions on Industrial Informatics*, vol. 16, pp. 4920–4927, 2019.

[129] W. Chen, D. R. Ding, H. L. Dong, and G. L. Wei, "Distributed resilient filtering for power systems subject to denial-of-service attacks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 8, pp. 1688–1697, 2019.

[130] H. F Song, D. R. Ding, and H. L. Dong, and X. J. Yi, Distributed filtering based on Cauchy-kernel-based maximum correntropy subject to randomly occurring cyber-attacks, Automatica, 135, 110004, 2022.

[131] J. Cao, D. Wang, Z. Y. Qu, M. S. Cui, P. C. Xu, Xue Kai, and K. W. Hu, "A novel false data injection attack detection model of the cyber-physical power system," *IEEE Access*, vol. 8, pp. 95109–95125, 2020.

[132] W. P. Deng, Z. Y. Yang, P. Xun, P. D. Zhu, and B. S. Wang, "Advanced bad data injection attack and its migration in cyber-physical systems," *Electronics*, vol. 8, no. 9, 2019.

[133] J. X. Wang, Z. X. Liu, S. G. Zhang, and X. Zhang, "Defending collaborative false data injection attacks in wireless sensor networks," *Information Sciences*, vol. 254, pp. 39–54, 2014.

[134] P. M. Lima, M. V. Alves, L. K. Carvalho, and M. V. Moreira, "Security against network attacks in supervisory control systems," *IFAC-Papers OnLine*, vol. 50, no. 1, pp. 12333–12338, 2017.

[135] J. V. Antrosiom and E. W. Fulp, "Malware defense using network security authentication," in *Proceedings of Third IEEE International Workshop on Information Assurance (IWIA'05)*, College Park, MD, USA, 2005, pp. 43–54.

[136] H. Rathore, S. Agarwal, S. K. Sahay, and M Sewak, "Malware detection using machine learning and deep learning," in *Proceedings of International Conference on Big Data Analytics. November*, 2018, pp. 402–411.

[137] Y. Zeng and R. Zhang, "Active eavesdropping via spoofing relay attack," in *Proceedings of 2016 IEEE International Conference on Acoustics Speech and Signal Processing*, Shanghai, 2016, pp. 2159–2163.

[138] J. H. Gao and K. J. Xia, "ARP spoofing detection algorithm using ICMP protocol," in *Proceedings of 2013 International Conference on Computer Communication and Informatics*, Coimbatore, Tamil Nadu, India, 2013, pp. 1–6.

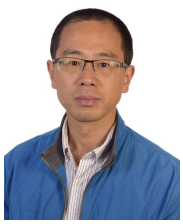
[139] M. Barreno, B. Nelson, R. Sears, A. D. Joseph, and J. D. Tygar, "Can machine learning be secure?" in *Proceedings of the 2006 ACM Symposium on Information, computer and communications security*, New York, USA, 2006, pp. 16–25.

[140] M. Jagielski, A. Oprea, B. Biggio, C. Liu, C. N. Rotaru, and B. Li, "Manipulating machine learning: Poisoning attacks and countermeasures for regression learning," in *Proceedings of 2018. IEEE Symposium on Security and Privacy*, San Francisco, CA, USA, 2018, pp. 19–35.

[141] J. J Lu, T. Issaranon, and D. Forsyth, "Safetynet: Detecting and rejecting adversarial examples robustly," In *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, 2017, pp. 446–454.

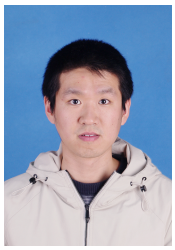
[142] N. Papernot, P. McDaniel, X. Wu, S. Jha, and A. Swami, "Distillation as a defense to adversarial perturbations against deep neural networks,"

in *Proceedings of 2016 IEEE Symposium on Security and Privacy*, San Jose, CA, USA, 2016, pp. 582–597.
[143] B. H. Wang, and N. Z. Gong, "Stealing hyperparameters in machine learning," in *Proceedings of 2018 IEEE Symposium on Security and Privacy*, San Francisco, CA, USA, 2018, pp. 36–52.
[144] Q. X. Xiao, Y. F. Chen, C. Shen, Y. Chen, and K. Li, "Seeing is not believing: Camouflage attacks on image scaling algorithms," in *Proceedings of 28th USENIX Security Symposium*, Santa Clara, CA, 2019, pp. 443–460.
[145] A. Venugopal, J. Uszkoreit, D. Talbot, F. J. Och, and J. Ganitkevitch, "Watermarking the outputs of structured prediction with an application in statistical machine translation," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Edinburgh, United Kingdom, 2011, pp. 1363–1372.
[146] Y. Uchida, Y. Nagai, S. Sakazawa, and S. Satoh, "Embedding watermarks into deep neural networks," in *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, Bucharest, Romania, 2017, pp. 269–277.
[147] H. L. Chen, B. D. Rohani, and F. Koushanfar, "Deepmarks: A digital fingerprinting framework for deep neural networks," Available at: https://arxiv.org/abs/1804.03648, 2018.
[148] E. L. Merrer, P. Perez, and G. Trédan, "Adversarial frontier stitching for remote neural network watermarking," *J. NEURAL COMPUT APPL*, vol. 32, no. 13, pp. 9233–9244, 2020.
[149] Y. Adi, C. Baum, M. Cisse, B. Pinkas, and J. Keshet, "Turning your weakness into a strength: Watermarking deep neural networks by backdooring," in *Proceedings of the 27th USENIX Conference on Security Symposium*, Baltimore, MD, USA, 2018, pp. 1615–1631.
[150] F. Tramèr, F. Zhang, A. Juels, M. K. Reiter, and T. Ristenpart, "Stealing machine learning models via prediction APIs," in *Proceedings of the 25th USENIX Conference on Security Symposium*, Austin, TX, USA, 2016, pp. 601–618.
[151] Z. Erkin, T. Veugen, T. Toft, and R. L. Lagendijk, "Generating private recommendations efficiently using homomorphic encryption and data packing," *J. IEEE T INF FOREN SEC*, vol. 7, no. 3, pp. 1053–1066, 2012.
[152] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, Vienna, Austria, 2016, pp. 308C-318.
[153] P. T. Xie, M. Bilenko, T. Finley, G. Ran, and M. Naehrig, "Crypto-nets: Neural networks over encrypted data," Available at: https://arxiv.org/abs/1412.6181, 2014.
[154] Y. C. Yu, L. Ding, and Z. N. Chen, "Research on attacks and defenses towards machine learning systems," *J. Netinfo Security*, vol. 9, pp. 10–18, 2018.
[155] H. Li, Q. Liu, and J. L. Zhang, "A survey of hardware Trojan threat and defense," *J. VLSI*, vol. 55, pp. 426–437, 2018.
[156] Y. J. Chen, X. L. Gong, Q. Wang, X. Di and H. Y. Huang, "Backdoor Attacks and Defenses for Deep Neural Networks in Outsourced Cloud Environments," *IEEE Network*,vol. 34, no. 5, pp. 141–147, 2020.
[157] Y. L. Mo and B. Sinopoli, "Secure control against replay attacks," in *Proceedings of 47th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, 2009, pp. 911–918.
[158] A. Hoehn and P. Zhang, "Detection of replay attacks in cyber-physical systems," in *Proceeding of 2016 American Control Conference*, Boston, MA, USA, 2016, pp. 290–295.



Zhenhua Yu (IEEE Member) received the B.S. degree and M.S. degree from Xidian University, Xi'an, China, in 1999 and 2003, respectively, and the Ph.D. degree from Xi'an Jiaotong University, Xi'an, China, in 2006. He is currently a Professor with the Institute of System Security and Control, College of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an, China. He has authored more than 30 technical papers for conferences and journals, and holds two invention patents. His research mainly focuses on cyber-physical systems, unmanned aerial vehicles, multi-agent systems, artificial intelligence system security, and nonlinear dynamical systems.

Hongxia Gao received the M.S. degree in computer science from Xi'an University of Science and Technology in 2022. Her research interests include system trustworthiness modeling and analysis. Currently, she is a front-end R&D engineer of Shanxi Unicom Industrial Internet Co., LTD.



Xuya Cong (IEEE Member) received the B.S. degree in automation from Xidian University, Xi'an, China, in 2014, and the Ph.D. degrees in control theory and control engineering and in electrical and information engineering from Xidian University and from the Polytechnic University of Bari, Bari, Italy, in 2020, respectively. He is currently a Lecturer with the College of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an. His current research interests include critical observability, opacity, supervisory control, and state estimation of discrete event systems and cyber-physical systems.

Naiqi Wu (IEEE Fellow) Naiqi Wu received his B.S. Degree in Electrical Engineering from Anhui University of Technology, Huainan, China, in 1982, the M.S. and Ph.D. Degrees in Systems Engineering both from Xi'an Jiaotong University, Xi'an, China in 1985 and 1988, respectively. From 1988 to 1995, he was with Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China, and from 1995 to 1998, with Shantou University, Shantou, China. He moved to Guangdong University of Technology, Guangzhou, China in 1998. He joined Macau University of Science and Technology, Taipa, Macau in 2013. He is currently a Professor at the Institute of Systems Engineering, Macau University of Science and Technology, Taipa, Macau. His research interests include production planning and scheduling, manufacturing system modeling and control, discrete event systems, Petri net theory and applications, intelligent transportation systems, and energy systems. He is the author or coauthor of one book, five book chapters, and 130+ peer-reviewed journal papers. Dr. Wu was an associate editor of the IEEE Transactions on Systems, Man, and Cybernetics, Part C, IEEE Transactions on Automation Science and Engineering, IEEE Transactions on Systems, Man, and Cybernetics: Systems, and Editor in Chief of International Journal of Information Sciences and IEEE/CAA Journal of Automatica Sinica.

Houbing Herbert Song (M12SM14-F23) received the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, VA, in August 2012. He is currently a Tenured Associate Professor, the Director of the NSF Center for Aviation Big Data Analytics (Planning), the Associate Director for Leadership of the DOT Transportation Cybersecurity Center for Advanced Research and Education (Tier 1 Center), and the Director of the Security and Optimization for Networked Globe Laboratory (SONG Lab, www.SONGLab.us), University of Maryland, Baltimore County (UMBC), Baltimore, MD. Prior to joining UMBC, he was a Tenured Associate Professor of Electrical Engineering and Computer Science at Embry-Riddle Aeronautical University, Daytona Beach, FL. He serves as an Associate Editor for IEEE Transactions on Artificial Intelligence (TAI) (2023-present), IEEE Internet of Things Journal (2020-present), IEEE Transactions on Intelligent Transportation Systems (2021-present), and IEEE Journal on Miniaturization for Air and Space Systems (J-MASS) (2020-present). He was an Associate Technical Editor for IEEE Communications Magazine (2017-2020). He is the editor of eight books, the author of more than 100 articles and the inventor of 2 patents. His research interests include cyber-physical systems/internet of things, cybersecurity and privacy, and AI/machine learning/big data analytics. His research has been sponsored by federal agencies (including National Science Foundation, National Aeronautics and Space Administration, US Department of Transportation, and Federal Aviation Administration, among others) and industry. His research has been featured by popular news media outlets, including IEEE GlobalSpec's Engineering360, Association for Uncrewed Vehicle Systems International (AUVSI), Security Magazine, CXOTech Magazine, Fox News, U.S. News & World Report, The Washington Times, and New Atlas. Dr. Song is an IEEE Fellow (for contributions to big data analytics and integration of AI with Internet of Things), and an ACM Distinguished Member (for outstanding scientific contributions to computing). He is an ACM Distinguished Speaker (2020-present), an IEEE Vehicular Technology Society (VTS) Distinguished Lecturer (2023-present) and an IEEE Systems Council Distinguished Lecturer (2023-present). Dr. Song has been a Highly Cited Researcher identified by Clarivate (2021, 2022). Dr. Song received Research.com Rising Star of Science Award in 2022, 2021 Harry Rowe Mimno Award bestowed by IEEE Aerospace and Electronic Systems Society, and 10+ Best Paper Awards from major international conferences, including IEEE CPSCom-2019, IEEE ICII 2019, IEEE/AIAA ICNS 2019, IEEE CBDCom 2020, WASA 2020, AIAA/ IEEE DASC 2021, IEEE GLOBECOM 2021 and IEEE INFOCOM 2022.