Monitoring and Defense of Industrial Cyber-Physical Systems Under Typical Attacks: From a Systems and Control Perspective

Yuchen Jiang, *Senior Member, IEEE*, Shimeng Wu, *Student Member, IEEE*, Renjie Ma, *Member, IEEE*, Ming Liu, *Senior Member, IEEE*, Hao Luo, *Senior Member, IEEE*, and Okyay Kaynak, *Life Fellow, IEEE*

(Review Paper)

***Abstract*—In the new industrial environment, the safe and reliable operation of Industrial Cyber-Physical Systems (ICPSs) is being threatened by new types of attacks: Attackers carefully tamper with the measurement and control data transmitted over the network, causing the controlled systems to behave abnormally. The essence of such threats is operational safety issues induced by information security issues, which need to be studied at the bottom monitoring and control layer of the system. Studying safety and security monitoring, as well as defense strategies against these attacks, is of paramount importance. The primary objective of this article is to offer readers a timely survey that sheds light on the current status of safety and security issues in ICPSs. A comprehensive comparison is conducted with existing approaches and relevant literature, focusing on a systems and control perspective. Specifically, we emphasize the concept of cyber-physical attacks by contrasting them with conventional cyberattacks. A summary of real-world instances of typical cyber-physical attacks is provided to illustrate their significance. In terms of methodology, we conduct a thorough review of attack principles, attack detection, and evaluation approaches, as well as defense schemes. During this process, we carefully compare the pros and cons of different detection methods. It is further elaborated that the information asymmetry between the offensive and defensive parties is the booster of the integrated design of industrial safety and security. Looking ahead, we identify and summarize fourteen open questions that warrant further research.**

***Index Terms*—Attack defense, attack detection, industrial cyber-physical systems, industrial safety and security.**

I. Introduction

CYBER-PHYSICAL systems (CPSs) are important research objects in the process of contemporary industrial transformation. The core elements include communication, computing, control, cognition, and cloud (known as “5 C”) [1]. According to the definition given by the United States National Science Foundation, “CPS is a system controlled or monitored by computer-based algorithms, closely integrated with the Internet and its users. In CPS, each physical or software component operates on different spatial and temporal scales, displays various behavior patterns, and interacts in a variety of ways.” The White Paper [2] defines the essence of CPS as building a closed-loop system between cyberspace and physical space based on automatic data flow, focusing on state perception, real-time analysis, scientific decision-making, and accurate execution, to solve the complexity and uncertainty problems in the process of manufacturing and application services and realize resource optimization. With a large number of field devices connected to the network and working online, the lack of comprehensively protected network communication leads to the expansion of the attack surface and brings huge security risks to the reliable operation of the controlled systems. Compared with traditional industrial control systems, the attack surface in industrial CPSs (ICPSs) is enlarged because more physical devices are connected to open networks (especially wide-area networks). Attackers can launch attacks either during the network transmission process or by targeting the control and monitoring center (see Fig. 1). It is noteworthy that, unlike spontaneous failures such as aging and damaged equipment, data anomalies caused by new types of attacks on cyber-physical systems are more complex. The attacker can manipulate multiple components at the same time, without

Manuscript received 28 April 2023; revised 5 August 2023; accepted 13 September 2023. Date of publication 19 September 2023; date of current version 4 October 2023. This work was supported in part by the National Natural Science Foundation of China under Grants 62203143, U20A20186, and 62073104, in part by China Postdoctoral Science Foundation under Grants 2022M710965 and 2023M730864, and in part by Heilongjiang Province Postdoctoral Foundation under Grants LBH-Z22130 and LBH-Z22150. *(Corresponding authors: Yuchen Jiang; Hao Luo.)*

Yuchen Jiang, Shimeng Wu, Renjie Ma, Ming Liu, and Hao Luo are with the Harbin Institute of Technology, Harbin, 150001, China (e-mail: yc.jiang@hit.edu.cn; hi.sandra.wu@gmail.com; renjiema@hit.edu.cn; mingliu23@hit.edu.cn; hao.luo@hit.edu.cn).

Okyay Kaynak is with the Turkish Academy of Sciences, Bogazici University, 34342 Istanbul, Turkey, and also with the King Abdulaziz University, Jeddah 21589, Saudi Arabia (e-mail: okyay.kaynak@boun.edu.tr).

Digital Object Identifier 10.1109/TICPS.2023.3317237

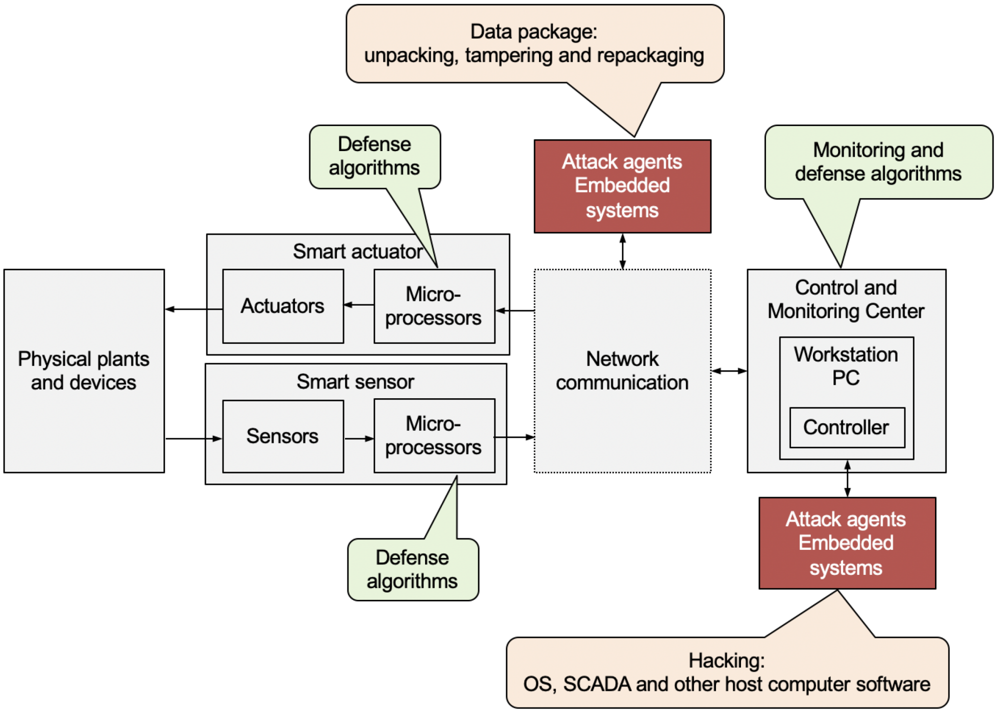

Data package:
unpacking, tampering and repackaging
Defense algorithms
Attack agents Embedded systems
Monitoring and defense algorithms
Smart actuator
Actuators
Micro-processors
Physical plants and devices
Network communication
Control and Monitoring Center
Workstation PC
Controller
Smart sensor
Sensors
Micro-processors
Defense algorithms
Attack agents Embedded systems
Hacking:
OS, SCADA and other host computer software

Fig. 1. Setup of industrial CPSs and locations of attacks: attacks during network transmission & attacks at the terminal devices.

changing the observation results, to make the operating state of the system deviate from the expected track [3]. In the meantime, cyber-physical attacks are more concealed and destructive, making many existing abnormal data detection methods and fault detection methods no longer applicable [4]. Therefore, it is urgent to study novel schemes to ensure dedicated safety and security monitoring and defense at the system control level.

After the occurrence of various infamous attack incidents in industrial control systems, research on attack principles and defenses has fostered emerging frontier research directions in the industrial automation field [5]. For the monitoring and defense of ICPSs, it is necessary not only to consider the nature of the controlled system, but also to have a deep understanding of the attack principles and even the attack intention, so as to make use of the characteristics related to the attack signals and develop countermeasures. As a basis, the major differences between physical attacks, cyber-attacks, and cyber-physical attacks need to be clarified.

Physical attacks take place in the physical world. Such attacks can be realized by de-functioning critical units. The most severe case lies in performing damages to the physical assets. For example, birds or missiles can cause airplanes to lose engines; electromagnetic interference can cause failures in network-based communication infrastructures and hinder digital terminals to complete the designed functions. Alternatively, making malicious structural changes is also a means to perform physical attacks. There were news presses where villagers steal water by privately installing pumps from the reservoir. From the perspective of closed-loop control, what they did is to add extra actuators that are unknown by the system designers and operators. Besides, causing abnormal environmental changes consists of another type of physical attack. The surrounding environment is the working conditions to be perceived by sensors, or sometimes regarded as an external disturbance. No functional changes are made to the sensors. For example, in an air-conditioner, a thermometer used to sense the temperature cannot distinguish whether there is an attacker holding a heat source beside it, and as a result, it will keep refrigerating. Such attacks can be quite covert because the environmental factors are not a part of closed-loop systems. As a result, they are undetectable by the monitoring systems.

Looking back on the aforementioned three types of physical attacks, despite the ability to perform targeted damages, they are usually limited by the scale of implementation and need plenty of informatics and economic inputs. Especially, it is unlikely to launch physical attacks at geographically dispersed areas simultaneously or in an organized manner.

Cyber-attacks take place in cyberspace. As a major threat to the security of computer systems, they have been extensively studied by computer scientists. With the advent of the industrial Internet and the emerging need for services in Industry 4.0, they are becoming a novel form of security risk in networked control systems in industrial facilities. Specifically, cyber-attacks occur between control/monitoring ends and the onsite operating machines. According to the location of security breaches, they can be categorized into 1) attacks during network transmission, and 2) attacks at the terminal devices. The former can be realized by sending fake data packages to multi-hop networks or routers along wireless data links. The latter can be realized by hacking the field control units or computers in the control rooms. Both have been extensively studied by the communication and computer network disciplines. Cyber-attacks pose serious threats to the confidentiality of data and the availability of network resources.

Cyber-physical attack is a terminology proposed by systems and control researchers in recent years. In one aspect, it can be understood as a concept that covers both physical space and cyberspace, and in another aspect, the terminology emphasizes the intention of attackers, which is to cause physical damage to ICPSs through cyber-attacks. Specifically, different from the research focus of traditional IT network security, this new type of safety/security risk will not only destroy the integrity and availability of information in the virtual space, but also cause direct and real damage to the connected and controlled physical entities. The attacks are highly targeted, with clear purposes, dedicatedly organized, and of huge destructive power.

The essence of this new type of threat is an operational safety issue induced by information security, so it is particularly evident in CPSs that integrate core elements of communication, control, and computing. Some existing literature refer to this kind of "novel attacks that illegally eavesdrops and tampers with the data transmitted over the network, thereby causing operation safety problems in CPSs" as **cyber-physical attacks** [6], [7], [8]. In this article, we also refer to the new type of attacks in CPSs as "cyber-physical attacks" for short. It has become insufficient to protect CPSs in such a context by excessively relying on secure communication protocols, network firewalls, and other border defense technologies and methods such as identity authentication and access control. Especially, after the malicious attackers manage to break in, it is critical to add on bottom level (i.e., control level) strategies and solutions to avoid the control system being compromised, leading to unstable operation or drastic performance degradation. Considering many cyber-physical attack events that took place in real world (e.g., Stuxnet virus, Black Energy, and so forth), it is reasonable to make efforts to design countermeasures and to take action when the traditional cybersecurity issues, as aforementioned, fail to protect the control systems. Therefore, this article emphasizes the importance of a comprehensive study on both attack principles and detection/defense schemes implemented alongside the control loops. The change in system dynamics due to external attacks must be identified, timely detection, and properly dealt with. To this end, it is urgent to study a set of safety/security monitoring and defense schemes at the operation and control layer by making full use of the relationship between the observation data, the underlying physical process, and the characteristics of typical attacks, so as to resist the well-designed attacks and establish a new bottom line of defense for the safe and reliable operation of the systems [9], [10], [11], [12].

This article consolidates empirical evidence and synthesizes observations from a systems and control viewpoint, primarily focusing on deepening our understanding of the existing safety and security climate in industrial cyber-physical systems (ICPSs). Its goal is to offer valuable insights to researchers seeking guidance by conducting a contemporary survey and review. The pivotal contributions of this study are as follows:

1) It presents an extensive exploration of the core reasons that have escalated the necessity to investigate new forms of attacks on ICPSs.
2) It sharpens essential technical approaches for studying attack principles, attack detection schemes, and defense strategies, providing an exhaustive review of existing methodologies in this field.
3) It performs a comparative analysis between current methods and relevant literature, seen through the lens of systems and control perspective. This helps facilitate a wider comprehension of the merits and limitations of various methods.

Addressing these facets, the study contributes towards progressing the knowledge base in the sphere of safety and security in ICPSs. It offers significant guidance for future research and development in this vital domain.

The remaining sections of the article are organized as follows: The subsequent section discusses the industrial background and the importance of the subject. Section III delineates how our approach differs from existing work, and showcases examples of typical ICPSs. Section IV provides a systematic review of the current research status, and building upon this, Section V suggests five key unresolved questions and proposes corresponding future directions to combat cyber-physical attacks. Lastly, Section VI concludes the article.

II. Background and Observations

ICPSs are dedicated to realizing real-time perception, dynamic control, and information services [1]. From the perspective of systems and control, by building a bridge between the physical entity space and the network information space, the controlled objects, the equipment, and the environment can be organically integrated from multiple dimensions, achieving in-depth collaboration. It plays an important role in improving the automation level of complex industrial control systems. Typical application fields of CPS include smart grids, smart manufacturing, unmanned autonomous systems, transportation networks, environmental monitoring, etc. [4], [13], [14], [15]. As an emerging and key multidisciplinary research field, CPS involves core elements of communication, computing, control, and cognition, which reflects the power of modern sensor technology, Internet and Internet of Things technology, and digitalization to jointly empower industrial development and transformation [16]. However, while new technology empowers the industry, it also poses novel challenges to the safety and security protection of the systems. Technological innovation and the large-scale construction of information infrastructure have brought unprecedented connectivity. A large number of smart devices are connected to the network and operate online. However, in the context of the expanded attack surface of malicious attacks, the lack of comprehensively protected network communication, and imperfect security protection mechanisms have brought huge risks to the safe and reliable operation of the controlled systems [3], [17].

Cyber-physical attacks have become a practical problem that needs to be dealt with urgently. There have been many notorious cases around the world, which have attracted widespread attention. For example, a water service system in Maroochy in Australia Queensland was attacked by a former employee in 2000. By invading the SCADA network and tampering with control signals, the attackers ultimately affected 150 sewage pumping stations, leading to a large amount of untreated sewage being evacuated to local waterways within three months. The second example is the Stuxnet virus attack discovered in 2011 [7]. Attackers first used social engineering methods such as phishing software and emails to induce enterprise employees to introduce viruses from the external network to the industrial control intranet with physical isolation. The viruses entering the internal network automatically run, eavesdrop and tamper with the data on the industrial control intranet, causing the process to run abnormally. The third example is that in 2007, the Idaho National Laboratory of the United States conducted an "Aurora Generator Test" [7], simulating the attack on a widely used turbo generator. It is considered the first case of cyber-physical attacks with practical significance. After the malicious code was implanted, the device vibrated abnormally, parts were ejected, and then thick smoke billowed, causing serious damage to the device. Another example is the power outage in Ukraine in 2015. The attackers carried out an organized, synchronous, and coordinated attack on three regional power distribution stations, resulting in a large-scale power outage for several hours, and the number of affected users reached 225,000 [7]. In addition, in early May 2021, the largest fuel pipeline company in the United States, Colonial Pipeline, was attacked. Its key fuel supply network had to shut down, thereby seriously affecting the supply of gasoline and diesel on the East Coast regions. The United States declared a state of national emergency on the 9th of the same month. It can be seen that it is of great significance to study the mechanism of cyber-physical attacks in depth and to propose safety/security monitoring and defense techniques suitable for CPSs. We dive into more details from the following three aspects.

1) Academic Significance: It is of great academic significance to study the monitoring and defense methods of cyber-physical systems under typical attacks. First, it is necessary to clarify the applicable conditions of the existing anomaly monitoring and attack detection methods through theoretical analysis. It needs to reveal the hidden theoretical flaws in the application and provide targeted improvement and optimization solutions to make up for theoretical loopholes. Moreover, it has attracted research focus to carry out quantitative analysis of various evaluation indicators of attack detection, and to systematically explain the monitoring results and decisions with a guaranteed theoretical foundation. Second, it can lay a theoretical foundation and provide useful information for related research fields. For example, the online operating status information analyzed by the safety monitoring system can be used in research directions such as active fault-tolerant control, adaptive system identification, online real-time optimization, and plug-and-play control. Third, different theoretical assumptions can guide the selection of the most suitable system monitoring solution and guide engineering practice according to actual needs.

2) Economic Significance: It is of great economic significance to study the monitoring and defense methods of cyber-physical systems under typical attacks. In one aspect, in the early stage of malicious attacks, key information leaks can be identified promptly, and an early warning can be issued before the equipment is physically damaged or completely out of control. It can effectively reduce downtime by online assessment of abnormal working conditions and timely switching of control strategies, isolation of attacks, and determination of urgency and priority of maintenance. In the meantime, the maintenance and repair costs can be reduced, and huge economic losses caused by the escalation of the situation can be avoided. In another aspect, the new research results will be applied to more challenging tasks such as building a key CPS system with a high safety/security level, optimizing control and management oriented to economic indicators, and so on. In the long run, the research will lay the foundation for the integration of upstream and downstream industrial chains, helping optimize the economic structure and resource allocation.

3) Social Significance: It is of great social significance to study the monitoring and defense methods of CPSs under typical attacks. The research direction is in line with the development needs. The *White Paper on Cyber-Physical Systems* [2] pointed out that "safety and security of industrial control system should be regarded as an important content in the promotion and deployment of current cyber-physical systems...to improve the capabilities of industrial information security situation awareness, risk warning, emergency response, and security protection". In addition, the research direction meets the development needs of multiple industries. For example, the *White Paper on Metallurgical Industrial Control System Active Defense Technology Systems* [18] emphasizes that "a mature integrated industrial control security active defense technology system is required...with risk identification, reinforcement, detection, and response capabilities." In key dynamic processes and systems such as smart energy, smart manufacturing, and aerospace, it has become a strategic requirement to study how to improve the reliability of monitoring systems and enhance the ability to resist cyber-physical attacks from the perspective of system and control.

III. Contrasting Current Reviews and Illustrations of Typical ICPSs

Primarily, two distinctive facets set this article apart from existing literature regarding the security of Industrial Cyber-Physical Systems (ICPSs).

In terms of the scope of the study, this work is limited to the countermeasures and solutions provided by the bottom control level, from the standpoints of both attackers and defenders. More specifically, we focus on monitoring attacks on networked close-loop control systems in which the sensor measurement data and control demands (actuation signals) are compromised. This is different from most cybersecurity issues studied extensively by computer scientists, or those focusing on secure/resilient control problems [19]. Nevertheless, in recent years, there are a few very well-written papers by control experts discussing cybersecurity issues. For example, we followed [5], [20], [21] in which a comprehensive discussion has been made about attack prevention, resilient control, detection/isolation, and threat assessment. However, the methods therein are mostly model-based, requiring known model structures and parameters. On top of it, we propose in this article more generic control diagrams for attack principles (represented by Fig. 3) and elaborate systematically on how information asymmetry can be used for the detection of concealed attacks, leveraging data-driven techniques. This is especially useful and distinguished from other anomaly detection

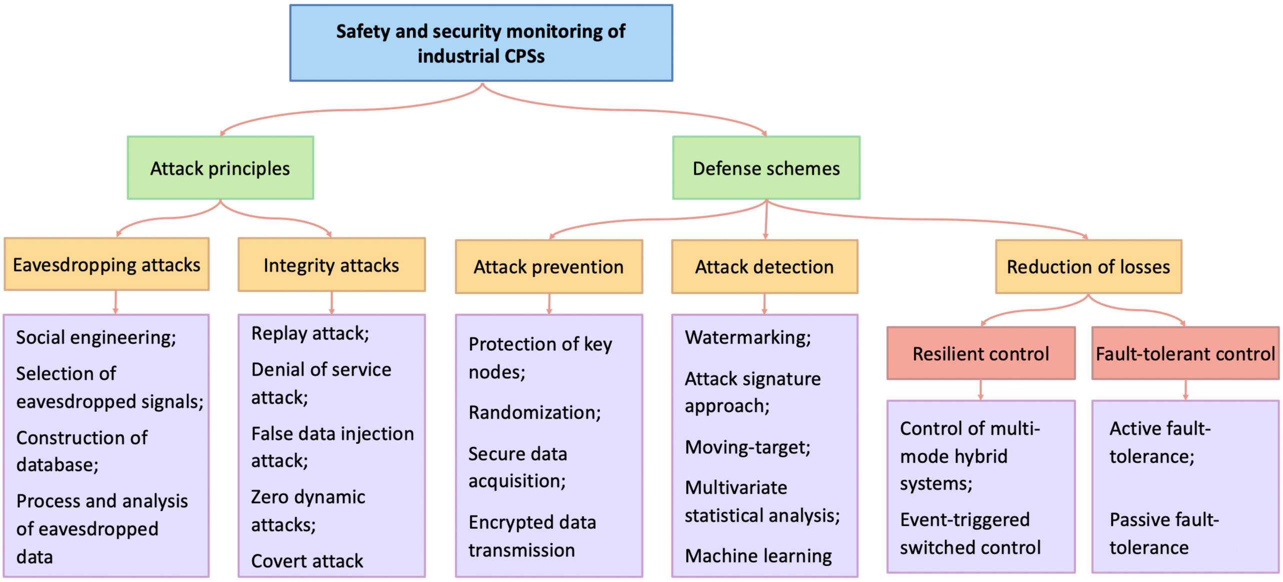
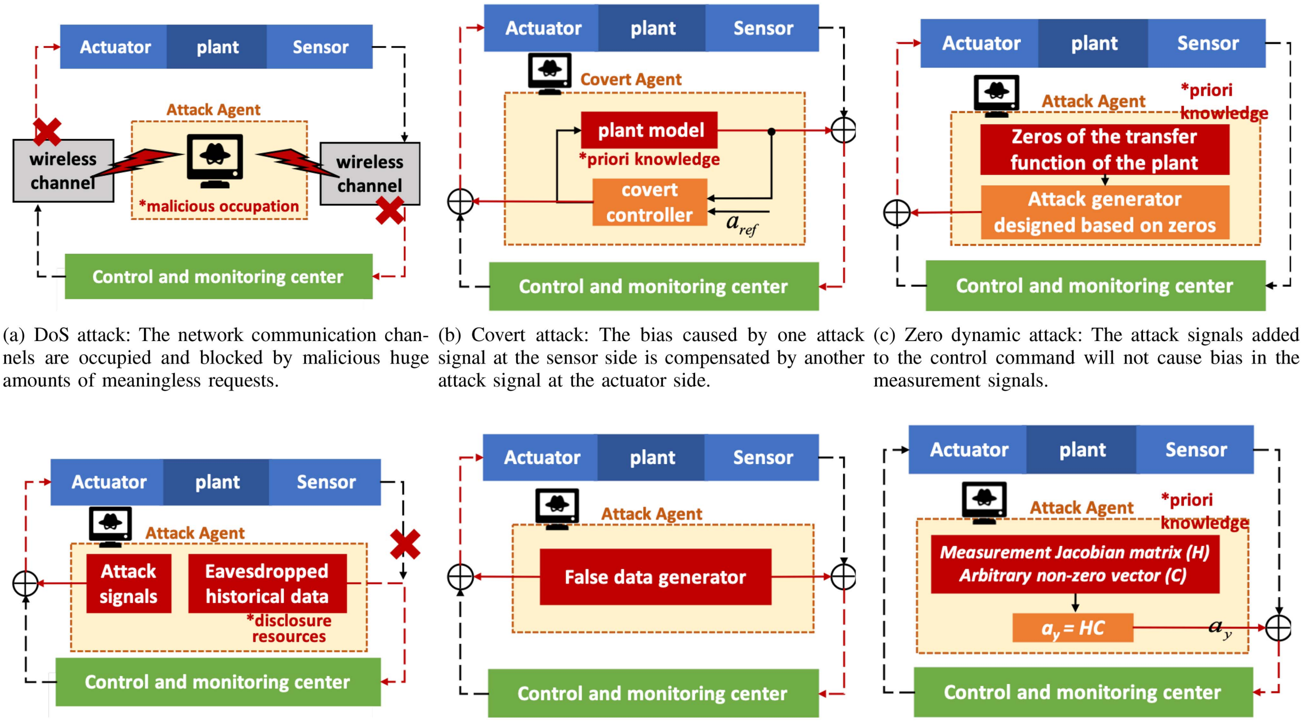


Fig. 2. Research directions and mainstream schemes of safety and security monitoring under cyber-physical attacks (from a systems and control perspective).



(a) DoS attack: The network communication channels are occupied and blocked by malicious huge amounts of meaningless requests.

(b) Covert attack: The bias caused by one attack signal at the sensor side is compensated by another attack signal at the actuator side.

(c) Zero dynamic attack: The attack signals added to the control command will not cause bias in the measurement signals.

(d) Replay attack: Recorded historical measurement data replace real measurements and are sent to the control and monitoring center.

(e) False data injection attack: Artificially designed false data are injected to the sensor and/or actuator communication networks.

(f) Stealthy FDIA: With careful design, false data injection attacks may remain stealthy to the monitoring and attack detection systems [28].

Fig. 3. Principles of typical attacks in industrial cyber-physical systems.

techniques such as fault diagnosis/isolation techniques, because the malicious attackers will become smarter and more equipped with control theoretic knowledge to design undetectable attacks if adopting the existing detectors which were designed for fault detection. In line with this point, we emphasize in Section V (Open Questions and Future Directions) that *distinguishing between external malicious attacks and internal system faults* is a scientific question worth studying.

In terms of the time span of the investigation, a more recent range of research outcomes have been analyzed. This is supported by the fact that over 50% of the references used are within the last five years (2019 to present). Therefore, compared to the existing publications like [5], [21], [22], this review-style paper incorporates more recent developments, embracing innovative concepts, fresh implementations, and the most recent experimental results. With these attributes, this article aims to provide a timely survey to enhance readers' understanding of the current state of intertwined safety and security issues, scientific dilemmas, and technical paths in industrial CPSs. It also intends to provide direction to researchers exploring this area. While differences are highlighted, it is crucial to acknowledge that many foundational concepts, ideas, and typical examples are drawn from these exceptional, trailblazing works.

As emphasized earlier, ICPSs are regarded as the core subject of study in the era of Industry 4.0. It is not limited to a single industry or a certain industrial scenario, but rather, a new pattern that bridges the industrial plants/devices/facilities in the real physical world and the digital replicas, services, and platforms in cyberspace. As such, there are many newly built and upgraded existing industrial systems that can be treated as ICPSs.

The demonstration of ICPSs can be found in most of the references herein. Nevertheless, considering that the scope of the article is about the systems and control perspective of ICPS and attacks on ICPSs, we direct interested readers to the demonstration of the most typical ICPSs with potential threats from cyber-physical attacks.

One of the most extensively studied ICPSs is the (smart) power grids. From a macroscopic point of view, the revolution of the energy and power industry constitutes a global challenge, due to the inevitable contradiction between population demand and energy supply. The traditional power grids meet bottlenecks when dealing with supply-demand balance. Therefore, smart grids, digital substations, and energy storage control systems have been found critical. They can take advantage of ICT technology to help the highly dynamic, uncertain, and distributed physical resources/assets significantly improve efficiency and reduce waste. The enabling effect of ICPSs provides promising solutions to connecting new energy to the grid (which has high uncertainty, e.g., wind energy), connecting massive energy-consuming units to the grids (also with high uncertainty, e.g., electric vehicles), and achieving robust and adaptive scheduling and coordination of multi-area grids. While we have learned the serious consequences of cyberattacks on the Ukraine power grid, other typical demonstrations to lab-scale attack/defense of power grids can be found in many existing publications, such as [6], [9], [10], [17], [19]. In addition to the energy industry, there are also many demonstrations in the process industry, such as water distribution and control systems [13], [21] and metallurgical process control systems [18], and the robot industry [20], [23].

IV. Current Research Status

It is pointed out in the *Guidelines for the Construction of Cyber-Physical Systems (2020)* [24] that CPS safety and security should include five links: risk identification, protection, detection, response, and recovery. It is worth noting that the CPS risks considered in this article all come from the "cyber-physical attacks" defined in the previous section, and the discussion will focus on the monitoring and control level of systems. In view of this, Fig. 2 shows the sub-directions and key technologies of safety and security monitoring research for cyber-physical attacks. It can be learned that the research on safety and security monitoring/defense of ICPSs needs to be carried out from two main angles. Attack principles are studied from the standpoint of attackers. The primary purpose is to learn about the weak points of the ICPSs and the concealedness (undetectability) of advanced attacks and to achieve targeted design and development of new defending systems and methodologies. From the standpoint of defenders, attack defense schemes include attack prevention (before attacking), attack detection (during attacking), and reduction of losses (during and after attacks).

In the following, we will analyze and summarize the current research status from three aspects: the principle of cyber-physical attacks in closed-loop control systems, the methods for attack detection and quantitative evaluation, and the security control strategies for defending against attacks. In the context of this article, the controlled CPS model is a standard feedback control system. Mathematical descriptions (formulas) of the referencing system models and attack models are referred to [20]. The controlled plant and the associated sensors and actuators are on one side. The monitoring and control center is on the other side. In terms of hardware implementation, as shown in Fig. 1, smart sensors and smart actuators in ICPSs are usually equipped with microprocessors that have computing power. Therefore, the defense algorithms can be deployed on these hardware. Meanwhile, on the other side of network communication, the control and monitoring center have more abundant computing power, so attack detection, prevention, and other monitoring and defense algorithms can be deployed here.

A. Research Status of Cyber-Physical Attack Principles

The principles of cyber-physical attacks are closely related to the types of systems under study and the information available. As shown in Fig. 1, the digital intelligent sensors, actuators, controllers, and other key components involved in the CPS closed-loop control and the networked real-time data transmission channels required for information interaction jointly constitute the attack surface. The existing data analyzers, verification mechanisms, and observers in attack detection and anomaly monitoring systems are all the targets of attack [25].

According to whether it will affect the operation of physical entities or practical processes, attacks can be divided into eavesdropping attacks and destructive attacks. Eavesdropping attacks will only destroy the confidentiality of data and illegally collect process operation data as the main source of information for the attacker to grasp the system's operating laws and gain a deep understanding of process dynamics. Destructive attacks will cause the transmission service to be unavailable or the data to be tampered with, which will affect the network receiving end and the subsequent processes to perform timely and correct calculations based on these data. Such a scenario can lead to a substantial drop in anticipated performance or even cause the

TABLE I
TAXONOMY OF CYBER-ATTACKS AGAINST CYBER-PHYSICAL SYSTEMS

Information or capabilities possessed by the attackers		**Priori knowledge** (System type, parameters)	**Disclosure resources** (Can be monitored)	**Disruptive resources** (Can be tampered with)
Denial of service attack		Unneeded	Unneeded	Needed
Integrity attack	Replay attack	Unneeded	Needed	Needed
	False data injection	Partly needed	Partly needed	Partly needed
	Zero dynamic	Needed	Unneeded	Needed
	Covert attack	Needed	Needed	Needed
Eavesdropping attack		Unneeded	Needed	Unneeded

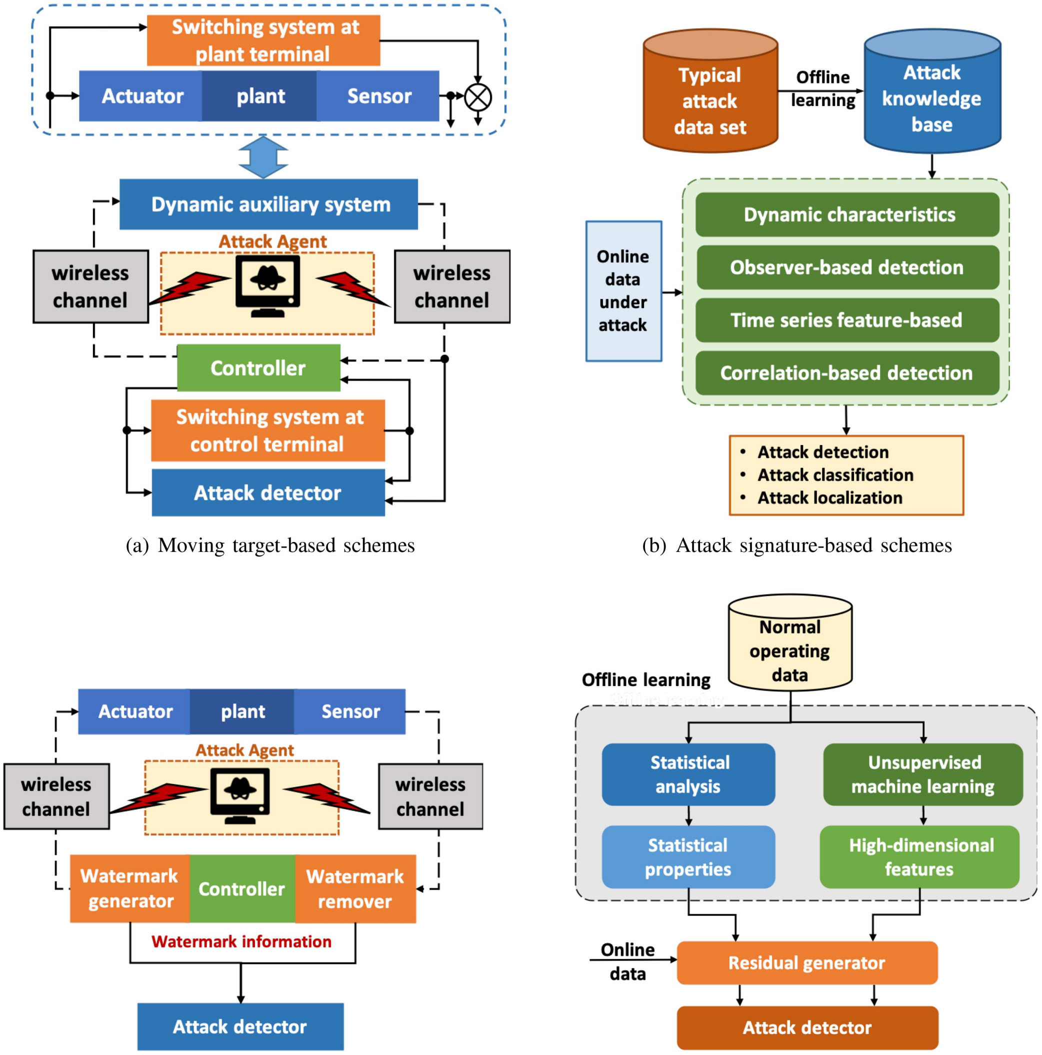
closed-loop system to become uncontrollable. This primarily includes instances of Denial of Service (DoS) attacks [26] and integrity attacks [27]. The principle of DoS attacks is sketched in Fig. 3(a). DoS attack destroys the availability of data through large-scale and continuous malicious occupation of network communication bandwidth. Although how to implement DoS attacks is mainly studied by computer network-related disciplines, it remains to be studied in the research on the cyber-physical attack theory which devices (or variables, ports) need to be attacked and how to collaborate with other types of attacks [26].

Typical integrity attacks can be divided into replay attacks [29], false data injection attacks [28], [30], [31], zero dynamic attacks [32], and covert attacks [13]. The principles of these attacks and the difference between them are shown in Fig. 3 and Table I. It can be learned that the key lies in the difference in the prior knowledge of the system and the degree of dependence on online real-time data. Specifically, it is divided according to the triplets of the attacker's mastery of system knowledge, the data that can be monitored, and the data that can be tampered with online [5]. Among them, the conditions for launching a covert attack are the most stringent. In order to manipulate the state trajectory while making the monitoring system judge that "everything is normal", the attacker needs to have sufficient prior knowledge and be able to collect all the controller commands and sensor data online and conduct a synchronous intervention on the two. A dedicatedly designed dynamic system is required for such a purpose, which is the so-called covert agent (or covert controller). Zero dynamic attacks are aimed at the observers, using the model knowledge of the controlled object to maintain the output value of the observer near the expected value, but the actual value has already deviated and is divergent. The concept of the false data injection attacks emphasizes the replacement of a part of the data during the transmission process, mainly targeting the defects of bad data detection (BDD) systems [28]. In contrast, the principle of replay attack is the simplest, which intends to bypass the inspection of the monitoring system by sending a set of eavesdropped historical data. It is worth noting that due to the different emphasis of the above concepts, the categories are not strictly mutually exclusive. For example, false data injection attacks can be designed to achieve the goal of covert attacks. There are also other commonly used terminologies such as deception attacks, which is to disguise true signals generated by the plant with artificial signals that are faked by the attackers. Deception attacks typically include replay attacks and false data injection attacks.

Today, there is no unified and widely-adopted classification standard to characterize cyber-physical attacks in academia and industry. The existing work makes strict assumptions for specific conditions, so it is still an open question how to define a general attack model for attack detection and resilient control under various types of CPSs and constraints.

According to [33], a general attack model that is based on data injection is shown in Fig. 5. It can be used to describe a range of attack types, including false data injection attacks, covert attacks, zero dynamic attacks, amplifying attacks, replay attacks, and so forth. In the central block ($a = Gen(K, I)$), $Gen(\cdot)$ denotes an attack generator (or an attack agent so to speak) that takes K and I as the inputs and then outputs an attack vector a. K denotes the attacker's knowledge about the attacked systems that can be used to design the attack strategy, including process knowledge K_p, control knowledge K_c, and detector knowledge K_d. For example, in an attack on two-area micro-grids [34], K_p represents the system topology, the setup of generators, loads, AC/DC communication lines, and energy storage devices; K_c represents the load-frequency controller, hybrid energy storage controller, optimal scheduling strategies, etc. K_d represents the adopted change detectors, anomaly detectors, fault diagnosers, and attack detectors. The other input, I, denotes the information that can be acquired by eavesdropping at the online stage. It is a subset of all sensor variables and control variables and is also referred to as disclosure resource in literature. It defines the data/signals from which communication channels for transmitting measurements and control commands are available for the attackers to drive the attack generator. The last block shows how the attack vector is used to tamper with the online measurements. By contrast to the disclosure resource block, the block of disruptive resource defines which subsets of sensor and control variables can be modified online by the attackers, either by replacing them with false data/replayed data or by adding/multiplying certain values on top of the real-time data. Overall, the normal attack-free data $\{u, y\}$ are changed to $\{u^a, y^a\}$ after the attack.

In the direction of modelling concealed attacks, the work of [35] proposed the concept of "kernel attack" to describe a general form of stealthy integrity attacks and the theoretical results of its detectability were recently published. The main idea is that knowing the system information, attack signals can be designed to perfectly bypass the stable kernel representation (SKR) based observer, thereby free from being detected. Specifically, the article shows that several types of replay attacks and zero dynamic attacks can result in an additional residual signal with zero-mean and small variance when the linear system is precisely modelled, so that the traditional fault diagnosis observers cannot detect them. Furthermore, the work of [36] defined the notions of detectability and identifiability by the degree of impact of attacks on the output measurements and described them from a



(a) Moving target-based schemes

(b) Attack signature-based schemes

(c) Watermarking-based schemes

(d) Statistical analysis and machine learning schemes

Fig. 4. Mainstream attack detection schemes against cyber-physical attacks.

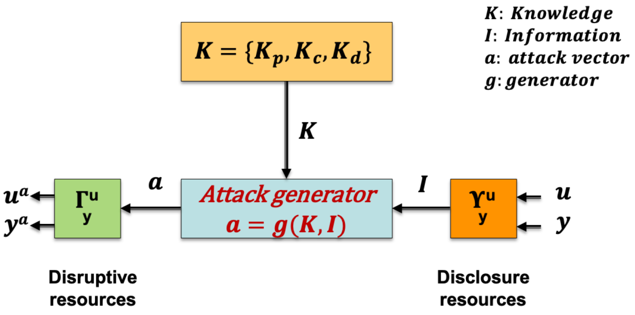


Fig. 5. General attack model based on data injection [33].

graph-theoretic perspective. The team of Professor Thomas Parisini conducted a series of research on large-scale interconnected systems and has put forward propositions on the detectability of distributed covert attacks [37]. However, an important issue that was omitted is that the following scientific questions (a and b) are not equivalent: a) The states of the local sub-system can be arbitrarily manipulated without changing the corresponding outputs. b) The attack is concealed, meaning that the residual signals generated by the local diagnosis systems fail to reflect abnormal changes. Having this clarified, the detectability condition of the false data injection attack was later corrected in [38], and the relationship between the concealment of the attack and the system topology was revealed. More recently, the detectability condition is further investigated for the partition attack and interconnection attack, which provides guidance on how to design a distributed control structure that is robust to attacks [39].

B. Research Status of Attack Detection and Quantitative Evaluation Approaches

There are a large number of existing model-based anomaly detection methods that can be used for attack detection [40], [41]. Such types of methods require model structures and model parameters to be known a priori. Also, the types of the controlled system, the principles of attacks, and the monitoring strategies need to meet a series of assumptions. Thereby, the applicable range is clearly defined yet rather limited [42], [43]. For example, to detect false data injection attacks in DC microgrids, the article [44] proposes a method to find invariant sets and determine the reachability of states online. To construct an approximate linearization model based on the shallow water equation, a residual generation method based on the unknown input observer (UIO) is used to detect abnormal fluctuations caused by attacks or failures [45]. For generalized linear systems with external inputs, the literature [36] proposes a centralized attack detection method based on the geometric control theory and puts forward a distributed attack identification method based on generalized linear filtering. In another aspect, with the wide application of sensor technology, industrial Internet technology, and the rapid development of digital twin technology, there is plenty of valuable information and knowledge contained in the historical data and the online real-time data which are collected during system operation. It can provide a new data-based technical route for the accurate, sensitive, and timely detection and identification of cyber-physical attacks [46]. Conducting research under the framework of data-driven and data-knowledge fusion not only adapts to the development of the Big Data era, but also provides a relatively simple and general method for complex systems that are difficult to accurately model.

Typical attack detection methods that can be implemented through data-driven techniques include watermarking approaches [29], [47], [48], [49], attack signature approaches [5], moving target approaches [35], hypothesis testing approaches (multivariate statistical analysis) [50], and machine learning-based classification methods [51]. The schematic diagrams of these detection schemes are shown in Fig. 4. The main idea of the watermarking approach is to superimpose a group of auxiliary signals (i.e., the watermarking signals) on the original transmission signal. Therefore, the focus of attack detection is to determine whether the known characteristics of the auxiliary signals have changed. By contrast, the attack signature approach is dedicated to detecting the characteristics of the attack signals with known mechanisms, so as to achieve classification and identification. Furthermore, the core idea of the moving target approaches is to introduce a time-varying dynamic auxiliary system that is difficult for the attacker to identify (e.g., a switching system is usually used) so that the augmented system composed of the original CPS and the auxiliary system remains sensitive to external attacks. Then the differences between the actual operation and the nominal system are checked by the residual generation and residual evaluation systems. Apart from the above, there are also other anomaly detection approaches based on statistical analysis and machine learning, which respectively examine the statistical properties and high-dimensional features of the transmitted data [52]. From the systems and control perspective, the research focus has undergone a shift from the analysis of the time-frequency characteristics of a single signal to the analysis of the relationship between multiple variables, and then to the analysis of the controlled system (or other auxiliary systems).

As mentioned in the previous section, cyber-physical attacks are generally clearly targeted and the attacker can manipulate key facilities after preliminary preparations, resulting in huge destructive power. At the same time, attackers sometimes tend to launch distributed attacks by targeting weakly protected nodes which it will cause the failure to propagate along network nodes and control loops. Therefore, after an attack is detected, a quantitative evaluation system[1] is needed to timely assess the potential risk and to identify which subsystems have been/will be affected, thereby modelling the situation of impact propagation [53], [54].

It is important to recognize that quantitative evaluation systems are not uniformly realized but instead represent a collection of evaluators and methodologies. For instance, article [55] introduces an attack detectability quantification approach for stochastic cyber-physical systems and a performance measurement strategy in terms of Kalman state estimation. The stealth (undetectability) of enhanced attacks and the system's quantitative relationship between performance deterioration were examined. Then, considering stochastic linear time-invariant systems, the literature [56] studied under what conditions integrity attacks could remain undetected to χ^2 (chi-square) detectors based on the invariant set theory and reachable set theory. The system's performance degradation is characterized by the maximum perturbation that can be induced. Furthermore, the propagation of influence between nodes is modelled as a mixed-strategy Bayesian attack-defense game problem with incomplete information based on the Petri net model of attacks [57]. An attack path analysis algorithm is proposed by constructing the threat propagation matrix and calculating the Nash equilibrium, which can indicate the possible attack path with a specific attack loss.

Additionally, in the field of process monitoring and fault detection, numerous methods exist for the quantitative analysis of impact propagation, whose central ideas have yet to be expanded to threat propagation issues related to intentional attacks (as opposed to spontaneous faults). For example, to address the problems of layer-based process monitoring and anomaly propagation path identification in complex industrial processes, a data-driven gap metric method and a neural network-based causal analysis method were proposed [58]. System-wide key performance indicators (KPIs) were used as a guide. The evaluation of system safety/security should employ multiple indicators for a comprehensive analysis. Therefore, research progress in performance evaluation indicators, evaluation methods, and subsystem/sub-region block methods under the data-driven framework were discussed in [59]. Literature [3] compiled recent related research on the cyber-physical security of industrial control networks, revealing that there is still insufficient research on evaluating the influence on system performance. As for the online analysis of cyber-physical system safety/security, the research on attack impact propagation and quantitative evaluation is still nascent, particularly for dynamic systems, closed-loop systems, and distributed interconnected systems. Compared with spontaneous faults, human-designed attacks present more variable patterns and are often more covert. However, despite the greatly increased complexity of the problems, ensuring the safety and security of cyber-physical systems remains crucial.

[1] A quantitative evaluation system is a structured framework designed for assessing the performance and potential risks of cyber-physical attacks utilizing quantitative methods, such as mathematical models or simulation tools.

TABLE II
PROS-AND-CONS OF DIFFERENT DETECTION METHODS: COMPARISON BASED ON A FEW SELECTED REALIZATION FORMS

Ref	Method	CPS Scenarios	Cyber-physical attacks	Pros	Cons
[60]	Moving target approach	Smart Grid	DoS attacks, stealthy FDIA, replay attacks	High detection accuracy for both non-stealthy and stealthy FDIA, replay Attacks	• Detailed knowledge of physical plant is needed
[61]	Robust moving target approach	Smart Grid	FDIA attacks	High detection accuracy in a noisy environment	
[62]	Extended moving target approach	Quardruple tank process	FDIA, covert attacks, replay attacks	High detection accuracy for both non-stealthy and covert attacks	
[63]	Moving target and watermarking	Quardruple tank process	Covert, replay, zero-dynamics attacks	High detection accuracy for both non-stealthy and zero-dynamics Attacks	
[64]	Sensor Watermarking	Time-invariant system	Stealthy replay attacks	High detection accuracy for stealthy replay attacks; Fault and attack isolation	• Detailed knowledge of physical plant is needed •Limited to replay attacks
[29]	Dynamic Watermarking	Time-variant vehicle model	Generalized replay attacks	High detection accuracy for stealthy replay attacks	
[65]	Autoencoder	Smart Grid	DoS, FDIA, replay attacks	High detection accuracy only with normal operating data	• Low detection accuracy for stealthy attacks • High false alarm rate caused by faults or disturbance
[34]	Multi-variate statistical analysis	Smart Grid	DoS, FDIA, replay attacks	High detection accuracy without system knowledge and operating data	
[66]	Signature-based method	Two-loop forced flow loop system	DoS, FDIA	Attack classification, fast detection	• Requires labeled normal and cyberattack data • Weak protections against unknown attacks

As noted in Footnote 1, a quantitative evaluation system is a structure designed to assess the performance and potential risks of cyber-physical attacks using quantitative methodologies like mathematical models or simulation tools. This system facilitates a systematic approach to evaluating the impact of various types of attacks on a CPS, and for comparing the efficacy of different defensive strategies. The presence of such a system is essential to assist stakeholders in making informed decisions on resource allocation for mitigating cyber-physical risks.

Remark 1: Table II lists several existing methods for detecting cyber-physical attacks in different application domains, all of which fall within the aforementioned technical routes and categories in this section. Please note that these methods are only a few possible realization forms. It is not intended to be comprehensive but to give demonstrative hints. In practice, attack detectability and the performance of detection and defense are dependent on specific scenarios, especially the characteristics of the controlled plants, the control and monitoring strategies, and the specific means of attacks.

C. Research Status of Attack Defense Schemes

The essence of defending against cyber-physical attacks is to make use of the confidentiality of system design and configuration and the information asymmetry between the attackers and the defenders to enhance the security of online operations. According to the phase of attacks, research on the defense schemes can be divided into attack prevention, robust design against typical attacks, attack identification and isolation, and optimized configuration to reduce losses.

In the monitoring and control layer, the methods to prevent attacks mainly include security acquisition of sensor measurements, encrypted data transmission, and randomization methods [34], [67], [68], [69]. For distributed and decentralized systems, determining the importance of nodes and achieving prioritization (physical) protection of key nodes plays an essential role in controlling the cost of safety-critical systems. For discrete-event ICPSs, Tao et al. proposed a reliable and secure data acquisition scheme before and after transmission [70]. To prevent attackers from using eavesdropping data for system identification and state estimation, many studies have introduced chaotic systems and synchronous control theory [71]. A chaotic system is a deterministic system whose state trajectory is extremely sensitive to the initial conditions. Any tiny offset will lead to significant differences in the evolution process. It has good unpredictable characteristics and is also suitable for large-scale lightweight deployment. For example, a check signal was designed by combining chaotic sequences and Chebyshev maps [65]. Furthermore, He et al. proposed an event-triggered strategy to realize the synchronization of master-slave neural networks, which can be used for encrypted data transmission [72]. The authors proposed a unified design framework in [52] that can defend against eavesdropping attacks and integrity attacks simultaneously. Data-driven encryption and decryption are achieved by constructing an auxiliary masking signal that has a strong correlation with the transmitted signal. It is suitable for static systems. Under this framework, even if the encryption mechanism is known to the attackers, it cannot be cracked due to the lack of auxiliary masking data in the offline training stage, which is necessary to obtain the decryption matrix.

When an attack has occurred, especially when the key transmission data in the SCADA system and field devices has been maliciously tampered with, certain measures need to be taken to locate and isolate the attack and reduce the overall performance loss. If the data received by the observer are erroneous, they will directly affect the estimation of the real-time states and further interfere with the observer-based control and fault diagnosis [73]. To deal with this, research on secure state estimation, resilient control, and attack compensation has received extensive attention and achieved fruitful results [74], [75], [76], [77], [78]. It is worth noting that most of the existing work adopts a model-based approach. For example, in order to study the secure state estimation approach when the CPS is sparsely attacked, it is possible to first analyze the difference of multiple transmission signals and group the attacked and unattacked signals, before which a distributed secure state estimator is dedicatedly designed. Commonly used auxiliary systems include sliding mode observers, multi-mode Luenberger observers, Kalman filters, etc. [73], [79]. The research focus of resilient control lies in the control of multi-mode hybrid systems, event-triggered switched control systems, and optimal control methods integrated with game theory [80].

If the attack has caused damage to components, it is necessary to depend on automatic fault-tolerant control methods to promptly and temporarily maintain the system stability and avoid escalation before human intervention. Active fault-tolerant control methods use residual signals to update the control law online [81] while passive fault-tolerant methods take known fault influence into account at the offline design phase and derive conditions for guaranteed (yet conservative) stability of the closed-loop system [82], [83], [84]. It is still an open issue to integrate data, principle models, and knowledge to improve defense performance and safe and secure control in the condition of cyber-physical attacks. Studying new technologies of digital twins and conducting online deduction will contribute to formulating global optimal response strategies under external threats [46].

D. Academic Organizations and Recent Research Activities

In the context where CPS acts as the pillar technology of Industry 4.0, the research on cyber-physical attacks has gradually developed into a relatively independent topic direction and has become the focus of discussions in mainstream academic journals, academic organizations, and high-level international conferences in recent years. For example, the technical committee of Fault Detection, Supervision, and Safety for Technical Processes with the International Federation of Automatic Control (IFAC SAFEPROCESS) recently included "detection, isolation, estimation, and diagnosis of cyber-physical attacks" among its core objectives and research directions. IEEE Industrial Electronics Society newly established the Technical Committee on Industrial Cyber-Physical Systems in 2015, and has successfully organized five annual International Conferences on Industrial Cyber-Physical Systems (IEEE ICPS). The Chinese Association of Automation has also set up several special committees closely related to cyber-physical safety and security. In the special issue of "Theory and Application of Cyber-Physical Fusion Systems" in the journal of Acta Automatica Sinica, about half of the articles directly study the defense methods against cyber-physical attacks.

At the 2021 IEEE International Conference on Industrial Technology, Professor Peter Palensky gave a keynote speech entitled "Cyber-Physical Security of Electrical Energy Systems", which introduced the real threats faced by cyber-physical power systems. Based on European's digitization process of energy networks, the safety monitoring and control problems of digital substations were discussed. At the 2021 IEEE ICPS, Professor Xinghuo Yu interpreted the security of cyber-physical systems from the perspective of systems engineering and nature inspiration in the plenary speech. Flagship conferences to be held in 2023 will continue focusing on the theory and technology of the safety and security of CPSs at various special sessions.

V. Open Questions and Future Research Directions

In this part, we summarize five key open challenges and the associated future research directions.

1) Attack Detection and Identification Approaches for Nonlinear Systems: Towards the monitoring system design tasks under cyber-physical attacks, existing studies have considered linear static systems and linear dynamic systems. For nonlinear systems, although a few theoretical methods have been proposed by researchers in the field of control, most of them are under the model-based framework and have many limitations for practical use due to strict assumptions and constraints. They are not applicable in the condition of complex working conditions that cannot be modelled. In such circumstances, the practical problem has to be oversimplified by approximate linearization near the operating point, which significantly limits the sensitivity of attack detection and the specificity of attack recognition. At present, data-based methods mostly use the correlation between variables or the time series characteristics of signals for anomaly detection, but do not make full use of the dynamic characteristics of the system and prior knowledge related to attacks. Therefore, how to make full use of the measurable data, the controller information and attack characteristics to detect and identify attacks on nonlinear systems is still a difficult problem that has not yet been solved.

2) Quantitative Evaluation and Analysis of the Propagation of Attack Influence in Closed-Loop Interconnected Systems: As mentioned in the previous section, research on attack influence propagation and quantitative evaluation is still in its infancy. On the one hand, relevant methods in the field of process monitoring still need to be extended to the problem of threat propagation related to external attacks. A series of quantitative analysis and evaluation schemes considering the influence of cyber-physical attacks need to be proposed, including sensitivity analysis of monitoring schemes to different attacks, estimation of key performance indicators, estimation of control performance degradation indexes (such as stability margin, tracking performance), prediction of remaining useful life of key equipment, etc. On the other hand, different from the spontaneous failures in CPSs, intentional attacks are more complicated and organized: many sites (or nodes) are usually attacked at the same time. Since the attackers' objective may lie in manipulating the system's state trajectory, the designed attack agents are with advanced dynamic characteristics and are mutually collaborative. In this case, how to make full use of the attack detection and attack identification results to track and locate the adversarial agents is still an open challenge to be solved.

3) Integrated Design Framework for Attack Defense and Monitoring: When attackers have broken through border defense technologies such as network firewalls and security defense technologies such as software access control, it is necessary to conduct attack defense and monitoring at the bottom control layers of the CPSs. Different from the analysis of network throughput and access frequency, defense and monitoring at the bottom layer can use the relationship between the measurement data and the principles of operation and control of the physical processes, making it possible to resist those intentional and organized attacks. At present, the research on cyber-defense problems (represented by confidential and secure data transmission schemes) and the research on the system monitoring problems (represented by the attack detection and identification schemes) are relatively independent, lacking complementarity and optimization. The existing integrated design framework of attack defense and monitoring is only suitable for linear static systems. The integrated framework and methods that are suitable for nonlinear systems and dynamic systems still need to be studied. In addition, how to achieve lightweight, modular, plug-and-play realization and reduce the impact of the deployment of the defense & monitoring systems on the control performance are meaningful future research directions.

4) Construct Benchmark Datasets for Typical Cyber-Physical Attacks: Different from traditional IT network attacks, cyber-physical attacks are new types of attacks targetting at the online measurement (process) data and the control command data which are transmitted through the network in the control and monitoring layer of the system. Although there have been many actual cases of attacks with huge impacts worldwide, the disclosure of specific attack information and data is extremely limited. For example, the German government work report mentioned that a steel plant was attacked but did not disclose the details of its industrial control system in detail. As such, for academic research, there is a lack of field raw data on cyber-physical attacks. Most studies can only analyze the attack principles and test the defense methods in a virtual environment. Since the publicly available and widely recognized benchmark test datasets for cyber-physical attacks are limited, validations are usually conducted based on generated attack data based on simulations and numerical examples [85]. Moreover, since cyber-physical attacks may cause huge damage to entities such as devices in the controlled system, the use of simulation-generated attacks also helps to conduct safer experiments within a controllable range. Nevertheless, to conduct comprehensive tests on the performance of attack detection and identification methods and quantitative evaluation of attack impacts, it is still necessary to construct datasets of typical integrity attacks as a benchmark for comparison studies.

5) Distinguish Between External Malicious Attacks and Internal System Faults: Distinguishing faults and attacks has important values for decision-making such as determining the maintenance levels and implementing security control strategies. If a fault occurs, it is necessary to repair or replace the failed devices and components. If it is an attack, it is necessary to isolate and block the compromised signals, make up for loopholes, and strengthen information protection. However, although there is an essential difference between unintentional faults and intentional attacks, both attacks and faults will cause data and signal anomalies, and traditional data anomaly detection methods, fault diagnosis methods, and attack detection methods cannot distinguish between the two. Therefore, it is still necessary to dig deeper into the physical nature behind signals and data to distinguish whether it is an internal fault or an external attack and whether the function of the physical entity is damaged or the data transmitted through the network has been invaded. Several criteria are needed. In addition, the knowledge of attacks and faults should be taken into account to accurately identify different data anomaly patterns.

In addition to the above, other commonly concerned open questions from existing literature [5], [19], [21], [59] are listed below. The positioning and relationship of these scientific questions are shown in Fig. 6.

6) Real-time Threat Assessment [5]: This involves the continual analysis of collected contextual data within the system to promptly identify cybersecurity threats as they occur.

7) Scalability for Large-scale ICPSs [19], [59]: This implies the capacity to expand and remove subsystems (with plug-and-play functions), and to design systems that can manage escalating data volumes, computational demands, and network connections without sacrificing performance or efficiency.

8) Metric of Resilience [5]: This pertains to establishing a quantitative measure utilized to evaluate the system's capacity to maintain critical functions, withstand and recover from adverse events, and rebound from disruptions or cyberattacks.

9) Location and Isolation of Attacks [19]: This includes the ability to identify the source and access point of cyber intrusions or attacks, and to contain the impact of the attack to prevent further damage.

10) Data Analysis Subject to Simultaneous Cyberattacks, Communication Scheduling, and Network-induced Phenomena [19]: This deals with the complications arising from the combination of multiple factors related to data transmission via the network.

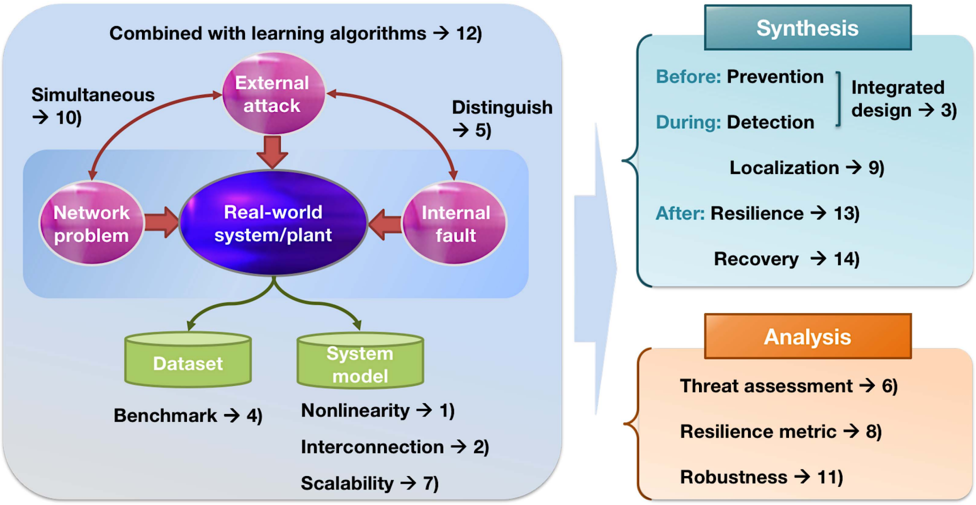


Fig. 6. Positioning and relationship of the open scientific questions.

11) Robustness to Uncertainties and Limited Knowledge about the Adversary [21]: This refers to the system's resilience to uncertainties and adaptability to situations with limited information about the adversary, especially in unpredictable scenarios.

12) Attacks Combined with Learning Algorithms [5], [21]: This involves understanding the mechanisms of how attackers can use machine learning techniques to optimize their attack strategies and evade traditional security measures, and the development of defensive AI systems capable of counteracting these evolving threats.

13) Attack-resilient Control [19], [59]: This refers to the design of control systems capable of maintaining certain performance levels and stability in the face of cyberattacks.

14) System Maintenance and Recovery [59]: As a crucial aspect of ICPS safety-security management, this involves restoring functionality online or within a short downtime and enhancing security measures to prevent future incidents.

VI. Conclusion

Modern ICPSs are developing toward large-scale interconnection and high automation. A growing number of industrial facilities and processes rely heavily on computer-based digital systems connected to industrial control networks. In this context, security and safety are deeply intertwined. This article refers to the novel form of threats as cyber-physical attacks, which belong to a type of security breach-induced operational safety problems. To deal with external attacks that intrude into the networks, the safety and reliability of the ICPSs are facing severe challenges. Security loopholes must be compensated by fail-safe strategies by the monitoring and control units. Meanwhile, the identified safety weak points can reversely guide the reinforcement of security defense strategies. That is the main reason why we emphasize studying from a systems and control perspective.

Because the attack activities are intentional and involve human intelligence, there are evident differences in the design of monitoring and defense schemes from schemes against spontaneous system faults. From the perspective of systems and control, the article discusses the disciplines of typical cyber-physical attacks as well as recent research results on detection, evaluation, and defense schemes. It is imperative to establish a new bottom line for system safety, even after the attackers have managed to penetrate the industrial control networks. In addition to the key questions and open challenges summarized in this article, many other technical issues await multidisciplinary efforts.

References

[1] A. W. Colombo, S. Karnouskos, O. Kaynak, Y. Shi, and S. Yin, "Industrial cyber physical systems: A backbone of the fourth industrial revolution," *IEEE Ind. Electron. Mag.*, vol. 11, no. 1, pp. 6–16, Mar. 2017.

[2] "The white paper on cyber-physical systems, " (in Chinese), China Electronics Standardization Institute, Beijing, China, 2017. [Online]. Available: http://www.cesi.cn/201703/2251.html

[3] M. Cheminod, L. Durante, and A. Valenzano, "Review of security issues in industrial networks," *IEEE Trans. Ind. Informat.*, vol. 9, no. 1, pp. 277–293, Feb. 2013.

[4] F. Farivar, M. S. Haghighi, A. Jolfaei, and S. Wen, "Covert attacks through adversarial learning: Study of lane keeping attacks on the safety of autonomous vehicles," *IEEE/ASME Trans. Mechatron.*, vol. 26, no. 3, pp. 1350–1357, Jun. 2021.

[5] S. Dibaji et al., "A systems and control perspective of CPS security," *Annu. Rev. Control*, vol. 47, pp. 394–411, 2019.

[6] A. Ashok, M. Govindarasu, and J. Wang, "Cyber-physical attack-resilient wide-area monitoring, protection, and control for the power grid," *Proc. IEEE*, vol. 105, no. 7, pp. 1389–1407, Jul. 2017.

[7] A. Humayed, J. Lin, F. Li, and B. Luo, "Cyber-physical systems security–A survey," *IEEE Internet Things J.*, vol. 4, no. 6, pp. 1802–1831, Dec. 2017.

[8] G. Loukas, *Cyber-Physical Attacks: A Growing Invisible Threat*, London, U.K.: Butterworth-Heineman, 2015.

[9] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber physical system security for the electric power grid," *Proc. IEEE*, vol. 100, no. 1, pp. 210–224, Jan. 2012.

[10] F. Pasqualetti, F. Dörfler, and F. Bullo, "Control-theoretic methods for cyberphysical security: Geometric principles for optimal cross-layer resilient control systems," *IEEE Control Syst. Mag.*, vol. 35, no. 1, pp. 110–127, Feb. 2015.

[11] X. Jiang and Z. Ge, "Adversarial attack on data-driven process monitoring systems: Subspace transfer networks," *IEEE Trans. Artif. Intell.*, vol. 3, no. 3, pp. 470–484, Jun. 2022.

[12] M. S. Rahman, M. A. Mahmud, A. M. T. Oo, and H. R. Pota, “Multi-agent approach for enhancing security of protection schemes in cyber- physical energy systems,” *IEEE Trans. Ind. Informat.*, vol. 13, no. 2, pp. 436–447, Apr. 2017.

[13] R. S. Smith, “Covert misappropriation of networked control systems: Presenting a feedback structure,” *IEEE Control Syst. Mag.*, vol. 35, no. 1, pp. 82–92, Feb. 2015.

[14] S. Rath, D. Pal, P. S. Sharma, and B. K. Panigrahi, “A cyber-secure distributed control architecture for autonomous AC microgrid,” *IEEE Syst. J.*, vol. 15, no. 3, pp. 3324–3335, Sep. 2021.

[15] S. Karnouskos and F. Kerschbaum, “Privacy and integrity considerations in hyperconnected autonomous vehicles,” *Proc. IEEE*, vol. 106, no. 1, pp. 160–170, Jan. 2018.

[16] S. Yin, J. J. Rodriguez, and Y. Jiang, “Real-time monitoring and control of industrial cyberphysical systems with integrated plant-wide monitoring and control framework,” *IEEE Ind. Electron. Mag.*, vol. 13, no. 4, pp. 38–47, Dec. 2019.

[17] K. Pan, J. Dong, E. Rakhshani, and P. Palensky, “Effects of cyber attacks on AC and high-voltage DC interconnected power systems with emulated inertia,” *Energies*, vol. 13, no. 5583, pp. 1–24, 2020.

[18] P. Cheng et al., “White paper on metallurgical industrial control system active defense technology systems,” (in Chinese), The key research and development project of the Ministry of Science and Technology “Research on the active defense mechanism and system for industrial control system security,” 2021. [Online]. Available: http://nesc.zju.edu.cn/#/res/whitebook

[19] D. Ding, Q. L. Han, X. Ge, and J. Wang, “Secure state estimation and control of cyber-physical systems: A survey,” *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 51, no. 1, pp. 176–190, Jan. 2021.

[20] M. Kordestani and M. Saif, “Observer-based attack detection and mitigation for cyberphysical systems: A review,” *IEEE Syst., Man, Cybern. Mag.*, vol. 7, no. 2, pp. 35–60, Apr. 2021.

[21] H. Sandberg, V. Gupta, and K. H. Johansson, “Secure networked control systems,” *Annu. Rev. Control Robot. Auton. Syst.*, vol. 5, pp. 445–464, 2022.

[22] S. Yong, M. Zhu, and E. Frazzoli, “Switching and data injection attacks on stochastic cyber-physical systems: Modeling, resilient estimation, and attack mitigation,” *ACM Trans. Cyber- Phys. Syst.*, vol. 2, no. 3, pp. 1–26, 2018.

[23] H. Alemzadeh, D. Chen, X. Li, T. Kesavadas, Z. T. Kalbarczyk, and R. K. Iyer, “Targeted attacks on teleoperated surgical robots: Dynamic model-based detection and mitigation,” in *Proc. IEEE/IFIP 46th Annu. Int. Conf. Dependable Syst. Netw.*, 2016, pp. 395–406.

[24] “Guidelines for the construction of cyber-physical systems,” 2020. [Online]. Available: http://www.cesi.cn/202008/6753.html

[25] K. Paridari, N. O’Mahony, A. El-Din Mady, R. Chabukswar, M. Boubekeur, and H. Sandberg, “A framework for attack-resilient industrial control systems: Attack detection and controller reconfiguration,” *Proc. IEEE*, vol. 106, no. 1, pp. 113–128, Jan. 2018.

[26] L. An and G.-H. Yang, “Decentralized adaptive fuzzy secure control for nonlinear uncertain interconnected systems against intermittent DoS attacks,” *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 827–83, Mar. 2019.

[27] J. Qin, M. Zhong, Y. Liu, X. Wang, and D. Zhou, “Convert attack detection based on Hi/Hinf optimization for cyber-physical systems,” *IFAC PapersOnLine*, vol. 53, no. 2, pp. 4487–4492, 2020.

[28] R. Deng, G. Xiao, R. Lu, H. Liang, and A. V. Vasilakos, “False data injection on state estimation in power systems–attacks, impacts, and defense: A survey,” *IEEE Trans. Ind. Informat.*, vol. 13, no. 2, pp. 411–423, Apr. 2017.

[29] M. Porter, P. Hespanhol, A. Aswani, M. Johnson-Roberson, and R. Vasudevan, “Detecting generalized replay attacks via time-varying dynamic watermarking,” *IEEE Trans. Autom. Control*, vol. 66, no. 8, pp. 3502–3517, Aug. 2021.

[30] R. Ma et al., “Adversarial FDI attack monitoring: Toward secure defense of industrial electronics,” *IEEE Ind. Electron. Mag.*, early access, Jul. 31, 2023, 2021, doi: 10.1109/MIE.2023.3292988.

[31] R. Ma, P. Shi, and L. Wu, “Sparse false injection attacks reconstruction via descriptor sliding mode observers,” *IEEE Trans. Autom. Control*, vol. 66, no. 11, pp. 5369–5376, Nov. 2021.

[32] Y. Chen, S. Kar, and J. M. F. Moura, “Dynamic attack detection in cyberphysical systems with side initial state information,” *IEEE Trans. Autom. Control*, vol. 62, no. 9, pp. 4618–4624, Sep. 2017.

[33] A. Teixeira, D. Pérez, H. Sandberg, and K. Johansson, “Attack model and scenarios for networked control systems,” in *Proc. Int. Conf. High Confidence Networked Syst.*, 2012, pp. 55–64.

[34] Y. Jiang et al., “Secure data transmission and trustworthiness judgement approaches against cyber-physical attacks in an integrated data-driven framework,” *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 52, no. 12, pp. 7799–7809, Dec. 2022.

[35] S. X. Ding, L. Li, D. Zhao, C. Louen, and T. Liu, “Application of the unified control and detection framework to detecting stealthy integrity cyber-attacks on feedback control systems,” *Automatica*, vol. 142, 2022, Art. no. 110352.

[36] F. Pasqualetti, F. Dörfler, and F. Bullo, “Attack detection and identification in cyber-physical systems,” *IEEE Trans. Autom. Control*, vol. 58, no. 11, pp. 2715–2729, Nov. 2013.

[37] A. Barboni, A. J. Gallo, F. Boem, and T. Parisini, “A distributed approach for the detection of covert attacks in interconnected systems with stochastic uncertainties,” in *Proc. IEEE 58th Conf. Decis. Control*, 2019, pp. 5623–5628.

[38] Y. Jiang, J. Dong, and S. Yin, “Improving the safety of distributed cyber-physical systems against false data injection attack by establishing interconnections,” in *Proc. IEEE 46th Annu. Conf. Ind. Electron. Soc.*, 2020, pp. 2623–2628.

[39] A. Gallo, A. Barboni, and T. Parisini, “On detectability of cyber-attacks for large-scale interconnected systems,” *IFAC PapersOnLine*, vol. 53, no. 2, pp. 3521–3526, 2020.

[40] A. Lu and G. Yang, “Secure luenberger-like observers for cyber–physical systems under sparse actuator and sensor attacks,” *Automatica*, vol. 98, pp. 124–129, 2018.

[41] Z. Tang, M. Kuijper, M. S. Chong, I. Mareels, and C. Leckie, “Linear system security–detection and correction of adversarial sensor attacks in the noise-free case,” *Automatica*, vol. 101, pp. 53–59, 2019.

[42] A. Barboni, F. Boem, and T. Parisini, “Model-based detection of cyber-attacks in networked MPC-based control systems,” *IFAC PapersOnLine*, vol. 51, no. 24, pp. 963–968, 2018.

[43] V. Palletia, T. Chong, and S. Lakshminarayanan, “A mechanistic fault detection and isolation approach using Kalman filter to improve the security of cyber physical systems,” *J. Process Control*, vol. 68, pp. 160–170, 2018.

[44] O. A. Beg, T. T. Johnson, and A. Davoudi, “Detection of false-data injection attacks in cyber-physical dc microgrids,” *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2693–2703, Oct. 2017.

[45] S. Amin, X. Litrico, S. Sastry, and A. Bayen, “Cyber security of water scada systems–part II: Attack detection using enhanced hydrodynamic models,” *IEEE Trans. Control Syst. Technol.*, vol. 21, no. 5, pp. 1679–1693, May 2013.

[46] C. Gehrmann and M. Gunnarsson, “A digital twin based industrial automation and control system security architecture,” *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 669–680, Jan. 2020.

[47] Y. Mo, S. Weerakkody, and B. Sinopoli, “Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs,” *IEEE Control Syst. Mag.*, vol. 35, no. 1, pp. 93–109, Feb. 2015.

[48] D. Wang, J. Huang, Y. Tang, and F. Li, “Watermarking strategy against linear deception attacks on remote state estimation under K-L divergence,” *IEEE Trans. Ind. Informat.*, vol. 17, no. 5, pp. 3273–3281, May 2021.

[49] B. Satchidanandan and P. Kumar, “Dynamic watermarking: Active defense of networked cyber–physical systems,” *Proc. IEEE*, vol. 105, no. 2, pp. 219–240, Feb. 2017.

[50] X. Li and K. W. Hedman, “Enhancing power system cyber-security with systematic two-stage detection strategy,” *IEEE Trans. Power Syst.*, vol. 35, no. 2, pp. 1549–1561, Mar. 2020.

[51] J. Zhang, L. Pan, Q. L. Han, C. Chen, S. Wen, and Y. Xiang, “Deep learning based attack detection for cyber-physical system cybersecurity: A survey,” *IEEE/CAA J. Automatica Sinica*, vol. 9, no. 3, pp. 377–391, Mar. 2022.

[52] Y. Jiang, S. Wu, X. Zhao, H. Luo, X. Li, and Y. Wu, “A lightweight defense scheme for industrial data transmission against eavesdropping attacks and integrity attacks,” in *Proc. IEEE 4th Int. Conf. Ind. Cyber- Phys. Syst.*, 2021, pp. 461–466.

[53] A. Stefanov and C. Liu, “Cyber-physical system security and impact analysis,” in *Proc. 19th IFAC World Congr.*, 2014, pp. 11238–11243.

[54] P. Lima, L. Carvalho, and M. Moreira, “Detectable and undetectable network attack security of cyber-physical systems,” *IFAC PapersOnLine*, vol. 51, no. 7, pp. 179–185, 2018.

[55] C. Bai, F. Pasqualetti, and V. Gupta, “Data-injection attacks in stochastic control systems: Detectability and performance tradeoffs,” *Automatica*, vol. 82, pp. 251–260, 2017.

[56] Y. Mo and B. Sinopoli, “On the performance degradation of cyber-physical systems under stealthy integrity attacks,” *IEEE Trans. Autom. Control*, vol. 61, no. 9, pp. 2618–2624, Sep. 2016.

[57] X. Liu, J. Zhang, P. Zhu, Q. Tan, and W. Yin, "Quantitative cyber-physical security analysis methodology for industrial control systems based on incomplete information Bayesian game," in *Proc. 19th IFAC World Congr.*, 2021, pp. 1021–38.

[58] L. Ma, J. Dong, and K. Peng, "A novel key performance indicator oriented hierarchical monitoring and propagation path identification framework for complex industrial processes," in *Proc. 19th IFAC World Congr.*, 2020, pp. 1–13.

[59] Y. Jiang, S. Yin, and O. Kaynak, "Performance supervised plant-wide process monitoring in industry 4.0: A roadmap," *IEEE Open J. Ind. Electron. Soc.*, vol. 2, pp. 21–35, 2020.

[60] J. Tian, R. Tan, X. Guan, Z. Xu, and T. Liu, "Moving target defense approach to detecting StuxNet-like attacks," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 291–300, Jan. 2020.

[61] W. Xu, I. Jaimoukha, and F. Teng, "Robust moving target defence against false data injection attacks in power grids," *IEEE Trans. Inf. Forensics Secur.*, vol. 18, pp. 29–40, 2023.

[62] N. Babadi and A. Doustmohammadi, "A moving target defence approach for detecting deception attacks on cyber-physical systems," *Comput. Elect. Eng.*, vol. 100, 2022, Art. no. 107931.

[63] M. Ghaderi, K. Gheitasi, and W. Lucia, "A blended active detection strategy for false data injection attacks in cyber-physical systems," *IEEE Trans. Control Netw. Syst.*, vol. 8, no. 1, pp. 168–176, Mar. 2021.

[64] R. Ferrari and A. Teixeira, "Detection and isolation of replay attacks through sensor watermarking," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 7363–7368, 2017.

[65] S. Wu, Y. Jiang, H. Luo, J. Zhang, S. Yin, and O. Kaynak, "An integrated data-driven scheme for the defense of typical cyber-physical attacks," *Rel. Eng. Syst. Saf.*, vol. 220, no. 108257, pp. 1–9, 2022.

[66] F. Zhang, H. A. D. E. Kodituwakku, J. W. Hines, and J. Coble, "Multilayer data-driven cyber-attack detection system for industrial control systems based on network, system, and process data," *IEEE Trans. Ind. Informat.*, vol. 15, no. 7, pp. 4362–4369, Jul. 2019.

[67] M. Usman, M. A. Jan, A. Jolfaei, M. Xu, X. He, and J. Chen, "A distributed and anonymous data collection framework based on multi-level edge computing architecture," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 6114–6123, Sep. 2020.

[68] H. Pearce, S. Pinisetty, P. Roop, M. M. Y. Kuo, and A. Ukil, "Smart I/O modules for mitigating cyber-physical attacks on industrial control systems," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4659–4669, Jul. 2020.

[69] B. Wang, L. Liu, C. Deng, M. Zhu, S. Yin, and S. Wei, "Against double fault attacks: Injection effort model, space and time randomization-based countermeasures for reconfigurable array architecture," *IEEE Trans. Inf. Forensics Secur.*, vol. 11, no. 6, pp. 1151–1164, Jun. 2016.

[70] H. Tao et al., "TrustData: Trustworthy and secured data collection for event detection in industrial cyber-physical system," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3311–3321, May 2020.

[71] D. Abbasinezhad, A. Ostad, S. Mazinani, and M. Nikooghadam, "Provably secure escrow-less Chebyshev chaotic map-based key agreement protocol for vehicle to grid connections with privacy protection," *IEEE Trans. Ind. Informat.*, vol. 16, no. 12, pp. 7287–7294, Dec. 2020.

[72] W. He, T. Luo, Y. Tang, W. Du, Y.-C. Tian, and F. Qian, "Secure communication based on quantized synchronization of chaotic neural networks under an event-triggered strategy," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 9, pp. 3334–3345, Sep. 2020.

[73] H. Yang, Y. Jiang, and S. Yin, "Adaptive control for cyber-physical systems against actuator attacks," in *Proc. IEEE Conf. Ind. Cyberphysical Syst.*, 2020, pp. 73–78.

[74] L. Hu, Z. Wang, Q. Han, and X. Liu, "State estimation under false data injection attacks: Security analysis and system protection," *Automatica*, vol. 87, pp. 176–183, 2018.

[75] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Trans. Autom. Control*, vol. 59, no. 6, pp. 1454–1467, Jun. 2014.

[76] Y. Gao, G. Sun, J. Liu, Y. Shi, and L. Wu, "State estimation and self-triggered control of CPSS against joint sensor and actuator attacks," *Automatica*, vol. 113, 2020, Art. no. 108687.

[77] M. Showkatbakhsh et al., "Securing state reconstruction under sensor and actuator attacks: Theory and design," *Automatica*, vol. 116, 2020, Art. no. 108920.

[78] J. Huang, D. Ho, F. Li, W. Yang, and Y. Tang, "Secure remote state estimation against linear man-in-the-middle attacks using watermarking," *Automatica*, vol. 121, 2020, Art. no. 109182.

[79] R. Ma, P. Shi, and L. Wu, "Dissipativity-based sliding-mode control of cyber-physical systems under denial-of-service attacks," *IEEE Trans. Cybern.*, vol. 51, no. 5, pp. 2306–2318, May 2021.

[80] Y. Li, L. Shi, P. Cheng, J. Chen, and D. E. Quevedo, "Jamming attacks on remote state estimation in cyber-physical systems: A game-theoretic approach," *IEEE Trans. Autom. Control*, vol. 60, no. 10, pp. 2831–2836, Oct. 2015.

[81] S. Yin, H. Luo, and S. Ding, "Real-time implementation of fault-tolerant control systems with performance optimization," *IEEE Trans. Ind. Electron.*, vol. 61, no. 5, pp. 2402–2411, May 2014.

[82] Z. Gao, S. X. Ding, and C. Cecati, "Real-time fault diagnosis and fault-tolerant control," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3752–3756, Jun. 2015.

[83] S. X. Ding, *Data-Driven Design of Fault Diagnosis and Fault-Tolerant Control Systems*, Berlin, Germany: Springer, 2014.

[84] S. X. Ding, *Advanced Methods for Fault Diagnosis and Fault-Tolerant Control*, Berlin, Germany: Springer, 2021.

[85] Y. Zhang, W. Deng, K. Huang, and C. Yang, "False data injection attack testbed of industrial cyber-physical systems of process industry and a detection application," in *Proc. IEEE Int. Conf. Recent Adv. Syst. Sci. Eng.*, 2021, pp. 1–7.

Yuchen Jiang (Senior Member, IEEE) received the B.E. degree in automation and the Ph.D. degree in control science and engineering from the Harbin Institute of Technology, Harbin, China, in 2016 and 2021, respectively. From 2019 to 2020, he was a Visiting Student Researcher with Technische Universität München with Lehrstuhl für Steuerungs- und Regelungstechnik, Munich, Germany. He is currently an Assistant Professor of control science and engineering with the School of Astronautics, Harbin Institute of Technology. His research interests include data-driven process monitoring, fault diagnosis and prognosis, industrial cyber-physical systems, and artificial intelligence.

Shimeng Wu (Student Member, IEEE) received the B.E. degree in automation from Harbin Engineering University, Harbin, China, in 2020, and the M.S. degree in control science and engineering in 2022 from the Harbin Institute of Technology (HIT), Harbin, where she is currently working toward the Ph.D. degree. Her research interests include fault diagnosis and prognosis, security of cyber-physical systems, and artificial intelligence.

Renjie Ma (Member, IEEE) was born in Harbin, China in 1994. He received the B. Eng degree in automation from the Hefei University of Technology, Hefei, China, in 2016, and the Ph.D. degree in control science and engineering from the Harbin Institute of Technology, Harbin, in 2022. He is currently an Assistant Professor of mechanical engineering with the State Key Laboratory of Robotics and Systems, Harbin Institute of Technology. His research interests include cyber-physical systems security, machine learning and optimization, robust control, and autonomous robotic systems.

Ming Liu (Senior Member, IEEE) received the B.S. degree in information and computing science and the M.S. degree in operational research and cybernetics from the Northwestern University, Xi'an, China, in 2003 and 2006, respectively, and the Ph.D. degree in mathematics from the City University of Hong Kong, China, in 2009. In 2010, he joined the Harbin Institute of Technology, where he is currently a Professor. His research interests include spacecraft control theory, intelligent fault diagnosis and health management technology. Dr. Ming Liu was selected as the New Century Excellent Talents in University of the Ministry of Education of China in 2013 and the Top-Notch Young Talents of Ten-Thousands Talents Program in 2018.

Hao Luo (Senior Member, IEEE) received the B.E. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2007, and the M.Sc. and Ph.D. degrees in electrical engineering and information technology from the University of Duisburg-Essen, Duisburg, Germany, in 2012 and 2016, respectively. He is currently a Professor with the School of Astronautics, Harbin Institute of Technology, Harbin, China. His research interests include model-based and data-driven fault diagnosis, fault-tolerant systems, and their plug-and-play application on industrial systems.

Okyay Kaynak (Life Fellow, IEEE) received the B.Sc. (first-class Hons.) degree and the Ph.D. degree in electronic and electrical engineering from the University of Birmingham, Birmingham, U.K., in 1969 and 1972, respectively. From 1972 to 1979, he held various positions within the industry. In 1979, he joined Bogazici University, Istanbul, Turkey, where he is currently a Professor Emeritus, holding the UNESCO Chair on Mechatronics. He has held long-term (near to or more than a year) Visiting Professor/Scholar positions with various institutions in Japan, Germany, USA, Singapore, and China. He has authored three books and edited five and authored or coauthored more than 450 papers that have appeared in various journals, books, and conference proceedings. His research interests include the fields of intelligent control and CPS. Dr. Kaynak is active in international organizations, has served on many committees of IEEE and was the president of IEEE Industrial Electronics Society during 2002–2003. He was elevated to IEEE Fellow status in 2003. In 2016, he was the recipient of the Chinese Government Friendship Award and Humboldt Research Prize, and the International Research Prize of the Turkish Academy of Sciences in 2020. He is a Member of this Academy.