Brief Survey on Attack Detection Methods for Cyber-Physical Systems

Tan, Sen; Guerrero, Josep M.; Xie, Peilin; Han, Renke; Vasquez, Juan C.

Published in:
I E E E Systems Journal

DOI (link to publication from Publisher):
10.1109/JSYST.2020.2991258

Publication date:
2020

Document Version
Early version, also known as pre-print

Brief Survey on Attack Detection Method for Cyber-Physical Systems

Sen Tan, *Student Member, IEEE,* Josep M. Guerrero, *Fellow, IEEE,* Peilin Xie, *Student Member, IEEE,* Renke Han *Member, IEEE* and Juan C. Vasquez, *Senior Member, IEEE*

***Abstract*—In recent years, Cyber-Physical Systems (CPSs) have attracted intense attention due to their potential applications in many areas. However, the strong reliance on communication networks makes CPSs vulnerable to intentional cyber-attacks. Therefore, a great number of attack detection methods have been proposed to enforce security of CPSs. In this paper, various false data injection attack detection methods presented for CPSs are investigated and reviewed. According to the knowledge of control information, the controllers of CPSs are categorized as centralized and distributed controllers. Existing centralized attack detection approaches are discussed in terms of (i) linear time-invariant systems, (ii) actuator and sensor attacks, (iii) nonlinear systems and (iv) systems with noise. Furthermore, the development of distributed attack detection is reviewed according to different decoupling methods. Some challenges and future research directions in the context of attack detection approaches are provided.**

***Index Terms*—Centralized detection, cyber-attacks, cyber-physical systems, distributed detection, false data injection attack.**

I. Introduction

THANKS to the rapid development of technology in communication networks, computer science and control theory, Cyber-Physical Systems (CPSs) have been extensively studied from both academia and industry. CPSs are systems that are controlled or monitored by computer-based algorithms, tightly integrated with networks and users [1], [2]. Examples of CPSs include smart grids, intelligent transportation networks, 5G cellular networks, sustainable developments, medical systems, process control systems, robotics systems and automatic pilot avionics [3]–[8].

A CPS typically consists of a network of interacting units with physical devices and computational elements [9]. The strong dependence on communication networks makes system vulnerable to cyber-attacks [10]–[12], such as Denial of Service (DoS) attacks and deception attacks [10], [13], [14]. Those attacks can be injected into systems both in cyber layer and physical layer [15]. Moreover, some malicious attackers would focus on attacks between cyber and physical layer, which can potentially induce significant damage on physical devices.

It should be noted that an attacker can either arbitrarily disturb the system dynamics or induce any perturbations to CPSs without enough security protections of hardware or software strategies, and thus leads to significant societal losses or the loss of human lives [16]–[22]. Examples include Iranian nuclear facility struck by the Stuxnet malware [18], blackout accident in nuclear plant [19], power blackouts in Brazil [20], etc.

These examples indicate an urgent need for reliable attack detection schemes to deal with malicious attacks and also maintain the performance of CPSs. If cyber-attacks could be detected and located in a short time period, the damage to overall systems would be controlled within a tolerable limit.

Most of the available literature on attack detections are based on centralized architectures [23]–[26]. As highlighted in [27], attack detection schemes can be often divided into knowledge-based and data-driven approaches. In most knowledge-based methods, one representative detection strategy is residual generation method [28]–[30]. Normally, a residual is designed by comparing the measurements of the sensors with an analytical model of the system. This residual is then compared with a fixed or time various threshold in order to determine if there is an attack. It should be mentioned that the residual generation approaches are always combined with the observer-based methods or statistics analysis methods. As for data-driven methods, deep learning and heuristic algorithms [31]–[33] are often used to build a model or map a relation of CPS. If system measurement data does not conform to some of the relationships, then an attack is assumed.

Aside from the centralized systems, more and more distributed systems appears in modern life. A typical example is microgrid [34]–[37]. A microgrid system consists of multiple energy sources, such as photovoltaic, wind turbines and batteries, which are interconnected via transmissions lines among each unit. Although these units are connected with each other, normally they are often operated independently. As a result, the distributed controllers may have limited information of the overall system dynamics. It is hard for a detector to monitor a CPS without enough information. This is the main challenge in the design of a distributed attack detection method.

In this paper, an overview of false data injection attack detection approaches for different CPS structures, methodologies and future trends is provided. A novel classification method based on the knowledge of various types of systems is provided. In this perspective, controllers for CPSs can be cat-

Manuscript received August 16, 2019; revised January 14, 2020 and March 2, 2020; accepted April 22, 2020.

This work was supported by VILLUM FONDEN under the VILLUM Investigator Grant (no. 25920): Center for Research on Microgrids (CROM).

S. Tan, P. Xie and J. C. Vasquez are with the center for research on Microgrids, Department of Energy, Aalborg University, Denmark.

J. M. Guerrero is with the Center for Research on Microgrids (CROM), Department of Energy Technology, Aalborg University, 9220 Aalborg East, Denmark (Tel: +45 2037 8262; Fax: +45 9815 1411; e-mail: joz@et.aau.dk).

R. Han is with the Department of Engineering Science, University of Oxford, OX1 3PJ, Oxford, United Kingdom. (e-mail: renke.han@eng.ox.ac.uk)

egorized as centralized controllers and distributed controllers. Then, different attack detection methods related to these two kinds of controllers are reviewed respectively. The rest of paper is organized as follows. In section II, the structure of cyber-physical control systems are surveyed and categorized. In section III, centralized attack detection methodologies in the existing literature are classified and reviewed. In section IV, distributed attack detection approaches are presented. In section V, potential research trends and conclusions are provided.

II. System Architecture

Several aspects need to be considered before solving an attack detection problem for a certain system. The architecture of a system and a communication link on which a potential attack is located are two main aspects to be considered first.

As seen from Fig. 1, ways implementing the control for most system can be centralized, decentralized, distributed, or in a hierarchical fashion [38], [39]. Based on the knowledge of system, the controllers can be divided into two categories: centralized controllers and distributed controllers. Table I shows the summary of two types of controllers.

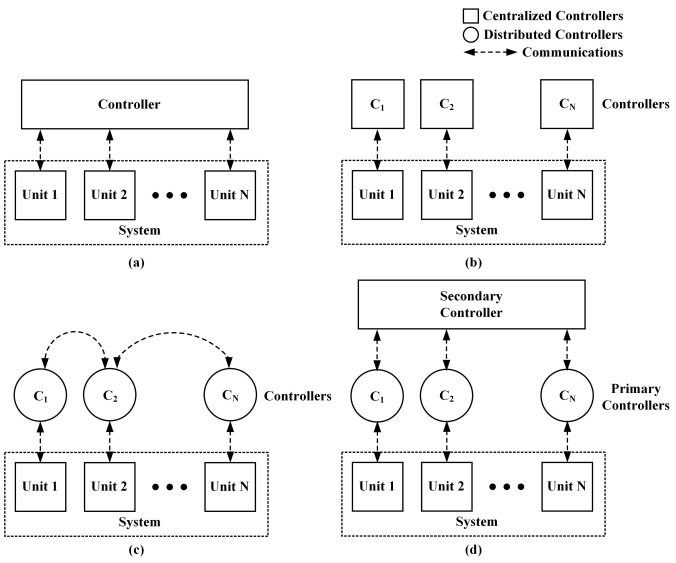


Fig. 1. Typical control structures. (a) Centralized. (b) Decentralized. (c) Distributed. (d) Hierarchical.

TABLE I
Structures of Controllers

Classification	Structure of System	Attack Location
Centralized Controller	1. Centralized 2. Decentralized 3. Hierarchical (Secondary layer)	Signals from 1. Local system
Distributed Controller	1. Distributed 2. Hierarchical (Primary layer)	Signals from 1. Local system 2. Neighbor system

A. Centralized controller

A centralized controller sets global information knowledge from its control unit measurements. Such controllers can be seen in centralized systems, decentralized systems and secondary layer of hierarchical control systems, which are shown in Fig. 1(a), (b) and (d), respectively.

As shown in Fig. 1(a), controller in a centralized control system requires data collection from all the essential components. Based on the gathered information, decisions are made in the controller to achieve proper and efficient operations. The case is definitely the same in the decentralized structure, which is shown in Fig. 1(b). Although it does not require information from other parts of the system, the information gathered is enough for the controller in the decision-making process. In order to implement advanced control or management, the secondary controller of hierarchical system shown in Fig. 1(d) is conventionally realized in a centralized manner as they require global information from all the essential units.

Therefore, controllers under these three systems can be categorized as a centralized controller, which have global measurements of control systems. In such systems, the potential attack occurs only on the two-way communication between the controllers and the physical components. Thus, attack detection algorithms need only be concerned with local states.

B. Distributed controller

Distributed controllers have only knowledge from local measurements and neighbor controllers. Such controllers can be found in a distributed control system (Fig. 1(c)) and the primary layer of a hierarchical system (Fig. 1(d)).

In a distributed control system, controllers only share information with neighbors who have a physical connection with them. As shown in Fig. 1(c), for instance, controller 1 only knows the running states of Unit 2 due to the communication between controller 1 and controller 2. Similarly, it can be seen from Fig. 1(d) that primary controllers of a hierarchical system can also be placed into this category. Although there are some communications between primary controllers and secondary controller, only system demands or references are transmitted into the primary controllers. Thus, the primary controllers still have limited information of the entire system.

Therefore, the controllers in these systems can be seen as distributed controllers. In this structure, the attack can occur not only on the communication between controllers and its units, but also on the communication between controllers and neighbor units.

III. Centralized Controllers

A. System Statement

The basic model of centralized systems can be written as Linear Time-Invariant (LTI) description:

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) + Ed(t) + Rf(t) \\ y(t) = Cx(t) \end{cases} \tag{1}$$

where $x(t)$ are system states; $u(t)$ are control actions applied to the process; $y(t)$ are measurements from the sensors; $d(t)$ are external disturbances and $f(t)$ are unknown signals representing the effects of anomalies or attacks along with

communication lines; A, B, C, E, and R are system matrices with proper dimensions.

In this paper, the false data injection attack is mainly considered, where an attacker might modify the control actions and/or sensor measurements from their calculated or real values to the corrupted signals, respectively. The signals are flowing from one node (controller, sensor or actuator) to another node through communication lines, which satisfies the following equation:

$$\hat{s}(t) = s(t) + a(t) \tag{2}$$

where $s(t)$ are the output signals from the node; $\hat{s}(t)$ are related signals received by another node; $a(t)$ are attacks on communication lines.

B. Attack Detection Design

The centralized attack detection methods can be divided into knowledge-based and data-driven based approaches. The main differences between these two methods are the ways to build system model. Knowledge-based methods build the the system model analytically, while data-driven methods infer a model directly from data.

1) State estimation: The state estimation method is used in the system to estimate system states through the analysis of measurements and system models [29], [30], [40]–[46]. Weighted Least Square (WLS) is most adopted method among state estimation methods, which can be presented as [40]–[46]:

$$\mathbf{z} = h(\mathbf{x}) + \epsilon \tag{3}$$

where $\mathbf{z}$ are vectors of measurements; $\mathbf{x}$ are system states; h is a function establishing dependencies between measured values and state variables; ϵ are measurement errors. The residuals are usually defined as:

$$\mathbf{r} = \mathbf{z} - h(\mathbf{x}) \tag{4}$$

Then, the WLS problem can be presented as:

$$\min F(\mathbf{x}) = (\mathbf{z} - h(\mathbf{x}))^T \cdot W \cdot (\mathbf{z} - h(\mathbf{x})) \tag{5}$$

where W is the weighing matrix whose elements correspond to the inverse of the accuracy of the individual measurements. The attack can be detected by comparing residuals (4) with a given threshold.

Although such methods can detect basic attacks, they may fail in the presence of more intelligent attackers that wish to stay undetected, where the false data could be introduced in a coordinated manner so that it looks consistent with the detection mechanism, thus bypassing it.

To cope with this problem, observer-based estimation methods are widely adopted in the literature [29], [30], [47], [48]. A Luenberger observer was adopted in [29] and [30] to monitor the power system and microgrid, respectively. In [47], an event-triggered state scheme was developed based on Luenberger like observer. In [48], an adaptive Slide Mode Observer (SMO) was designed in the detection of cyber-attack on power systems.

In addition, Kalman Filter (KF) is shown to be effective in detecting various attacks, including short-term and long-term random attacks along with powerful deception attacks [49].

2) χ^2 detector: χ^2 detector [50]–[56] is also widely used to detect anomalies in control systems. It has the capability to detect against bad data (erroneous measurements based on sensor/meter failures or malicious attacks [57]), by capturing the statistical behaviors of states.

Given system (1), a χ^2 detector computes the following quantity:

$$g_k = r_k^T Q^{-1} r_k \tag{6}$$

where r_k are the residuals given by $r_k = y_k - C(A\hat{x}_{k-1} + Bu_k)$; u_k and y_k are the vectors of system input and output at time k; $\hat{x}_k$ are the estimated states at time k; Q is the covariance matrix of r_k. Since r_k is Gaussian distributed, g_k is χ^2 distributed. The χ^2 detector will compare g_k with a certain threshold. If g_k is greater than the threshold, an alarm will be triggered.

The χ^2 detector can detect system attacks, such as DoS attacks, short-term and long-term random attacks. However, some studies show that the χ^2 detector is unable to detect the attacks, when the injected measurement data fit the distribution of historical data [55]. To overcome this limitation, a Euclidean-based detector was proposed by Kebina in [55] to detect attacks in the smart grid, where the distance between measured and estimated variables is defined as:

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2} \tag{7}$$

where $p = \{p_i | i \in 1, 2, \cdots, n\}$ and $q = \{q_i | i \in 1, 2, \cdots, n\}$ are the measurements and estimations of system states respectively. If the difference between the two is greater than the threshold, as in the case of the χ^2 detector, an alarm is triggered.

Similarly, in [25], Rawat and Bajracharya proposed cosine similarity matching-based approach to detect the deviation between measured data and estimated data, which is given as:

$$sim(\hat{\mathbf{x}}, \mathbf{x}) = \frac{\hat{\mathbf{x}} \cdot \mathbf{x}}{\|\hat{\mathbf{x}}\| \|\mathbf{x}\|} \tag{8}$$

where $\hat{\mathbf{x}}$ are vectors of estimated states. If both data vectors are similar, the value of cosine similarity will equal to one, which represents no attack in the system. If the value is less than a given threshold, then an alarm will be triggered.

3) Fault detection and identification method: Thanks to the full development of Fault Detection and Identification (FDI) algorithms, the model-based fault diagnosis technique is nowadays accepted as a powerful tool to solve attack detection problems [30], [58]–[65].

Residual generation is the subject in the application of FDI, which can be divided into two frameworks: observer-based method [30], [58]–[62] and parity space-based method [63]–[65].

The schematic of the observer-based attack detection is to design a residual which is influenced by attacks and system disturbances. The frequency-domain residual signals can be described by:

$$r(s) = G_f f(s) + G_d d(s) \tag{9}$$

where G_f and G_d are transfer functions from attacks and disturbances to residuals.

A residual generation is called Perfect Unknown Input Decoupled (PUID) from the disturbances if $G_d = 0$, which means the residual will only be influenced by attacks. Unknown Input Observers (UIOs) [54], [58]–[61], [66] are most adopted in literature in this case.

It should be recognized that the design of PUID residues requires the assumption of an existing condition and can be only achieved when enough number of sensors are available, which may be too strong for a realistic dissemination of this technique in practice. A reasonable extension of the PUID residual design is to make a compromise between the robustness against the unknown input and the sensitivity to the attacks, which means on the one hand, maximum the effect of attacks on residuals $\|G_f\|$, on the other hand, minimum $\|G_d\|$, where $\|\cdot\|$ denoted matrix norm.

In order to utilize parity space-based residual generation method, the system (1) need to be written into following discrete form:

$$y_s(k) = H_{os}x(k-s) + H_{us}u_s(k) \tag{10}$$

where $y_s(k) = \big[y(k-s), y(k-s+1), \cdots, y(k)\big]^T$ and $u_s(k) = \big[u(k-s), u(k-s+1), \cdots, u(k)\big]^T$ are past output and input; H_{os} and H_{us} are known functions of system matrices determined by matrices A, B and C. The residuals are then defined as:

$$r_s(k) = v_s(y_s(k) - H_{us}u_s(k)) \tag{11}$$

where $v_s \neq 0$ is a vector which satisfies:

$$v_s H_{os} = 0 \tag{12}$$

Hence, a parity relation-based residual generator is constructed by:

$$r_s(k) = v_s(H_{fs}f_s(k) + H_{ds}d_s(k)) \tag{13}$$

where H_{fs} and H_{ds} are impact coefficients from attacks and disturbances to residuals.

Noticed that the design parameter of parity space-based residual generator is the vector v_s whose selection decisively affects the performance of the residuals. In addition, it can be seen from (13) that there is also a tradeoff between the robustness against the disturbances and sensitivity to the attacks. A natural way to improve the robustness of residual against disturbances is to select v_s wisely which aims to on one hand maximizing $\|v_s H_{fs}\|$, while on the other hand minimizing $\|v_s H_{ds}\|$.

The main advantage of parity space-based residual generations over the observer-based approaches is that the design can be carried out in a straightforward manner. However, it is mainly applied to discrete time systems, due to the need of past measurements and input data.

4) Binary hypothesis-based method: Binary hypothesis-based method [67]–[69] is mostly used to detect malicious attacks in multi-sensor networks, where the decision is made by two steps.

First, the output of system is collected by N sensors. Based on the observations, each sensor i makes a one-bit local decision regarding the absence or presence of attack using likelihood ratio test or energy detection scheme. After that, the local decisions are sent to a central decision-making center called Fusion Center (FC) [68].

Second, FC makes the final decisions about the states based on all the information received from the participating sensors. In this binary decision case, such a decision is often made by a *posteriori* probability comparison which is given by:

$$r = P(H_1|\mathbf{u}) - P(H_0|\mathbf{u}) \tag{14}$$

where H_1 and H_0 are two hypotheses which represent signal is present and absent, respectively; $\mathbf{u}$ are local decisions; P is probability function. No absent signal decision is made if $r > 0$, otherwise, the signal is absent.

In addition, a simple scheme was proposed by Ankit in [69] to identify the attacks, where a reputation metric of the i-th sensor is defined as:

$$n_i = \sum_{t=1}^{T} \mathcal{I}_{(u_i(t) \neq u_o(t))} \tag{15}$$

where $\mathcal{I}(\cdot)$ is the indicator function of inconsistency; T is sensing period; $\boldsymbol{u_i} = [u_i(1), u_i(2), \cdots, u_i(T)]$ represent the local decisions forwarded by sensors; $\boldsymbol{u_o} = [u_o(1), u_o(2), \cdots, u_o(T)]$ represent the final decisions of the FC. Then, sensors will be isolated or cut off from information fusion process whose reputation metrics are greater than a fixed threshold.

An interesting topic on this method is the optimal strategy design for attackers and fusion centers using minimax game theory approach. Here the objective of the fusion center is to design the strategy to help obtain accurate detection and, on the contrary, the attacker aims to deteriorate the detection performance.

5) Model free detection scheme: Model free-based detection approached have also been introduced for monitoring attacks in CPSs [31]–[33], [70]–[76]. These solutions generally rely on machine learning or statistical mechanisms to infer a model for the system under inspection directly from data. The motivation behind considering such an approach comes from the fact that some analytical models might not be accurate enough [72].

In [73], a graph-based Auto-Regressive model was built by time series measurement data under normal operating conditions. With this graph, if the system measurement data does not fit some (or all) of the relationships at some point, then an attack is detected.

Two data mining methodologies: Artificial Neural Networks [31], [70] and Support Vector Machine [71] were used to detect potential system intrusions. In [70], 5-day data that contains hundreds of megabytes with attacks were used to training the ANN and SVM. Then attack can be detected by the well trained network. In [74], He et al. exploited deep learning techniques to recognize the behavior features of attacks with the historical measurement data and employ the captured features to detect the attacks in real-time. Moreover, some attack detection approaches based on heuristic search have been proposed on several research works. Related reports can be found in [31]–[33].

Although the machine learning method is a useful tool to build a system model, it might introduce a heavy computational burden. Training a fully connected network in many cases is very difficult [75]. In order to reduce the difficulties of training neural networks, the reservoir computing method is widely adopted in the literature. For example, in [75] and [76], a reservoir computing-based method was proposed by Hamedani to detect single attacks and stealth attacks, respectively.

C. Attack Detection Against Actuator and Sensor Attack

Apart from attacks on the communication link, an adversarial attacker may also hijack and compromise an actuator and/or sensor network [77]. Therefore, reliable monitoring of CPSs by implementing proper measures against actuator and sensor attacks is of significant importance.

In order to provide security monitoring, secure state estimation, which aims to estimate the states from corrupted measurements, has attracted considerable attention [78]. The strategies of secure state estimation can be categorized into (i) attack space search method [79]–[83]; (ii) convex relaxation method and (iii) attack estimation method.

1) Attack space search method: This approach is a kind of method which searches over all attack space combinations [29]. Therefore, it is time consuming and may not be practically feasible in real control systems. To reduce the computation complexity, some search space reduction methods were proposed based on set theory [80]–[82], satisfiability modulo theory [83] and l_0 optimization approach [84].

2) Convex relaxation method: Convex relaxation method is another approach to estimate states by analyzing the sensor measurements collected within a time window of finite length [28]. This method can be categorized into l_1/l_r optimization formulation [12], [85], [86] and projected gradient descent algorithm [47]. A major drawback of this method is the correctness may only be guaranteed under restrictive assumptions on the system structure. There is a tradeoff between correctness estimation and computational complexity.

3) Attack estimation method: In addition, attack estimation is an approach that aims at identifying the attacks and providing resilient state reconstructions [78], [87], [88]. In [78], a secure Luenberger-like observer was proposed to estimate the states and attacks from the corrupted measurements. Furthermore, simultaneous unknown input and state estimation method was proposed by Yong at al. in [87], [88] to realize a resilient state estimation and attack mitigation.

It is worthy to known that the output signals are only guaranteed to be reconstructible if a certain upper bound on the number of attacked sensors is met [12], [29], [56], [89], [90] . Therefore, a protection-based approach was proposed in [91], [92] to prevent attacks by protecting measurements from certain sensors. However, the state estimation could still be in danger when the protection is penetrated by an attacker [93].

4) Watermarking method: Outside of the three categories above, watermarking is another effective approach to detect actuator attacks [94] and sensor attacks [56]. With this method, judiciously designed excitation signals are superimposed on the control commands to increase the detectability of the actuator attack, which can be presented as:

$$u_k = u_k^* + \Delta u_k \tag{16}$$

where u_k^* is the optimal control signal; Δu_k is an authentication signal from Gaussian distribution with zero mean, which is also called watermark signal; u_k is new control signal. The whole control system becomes more dynamic and converges slowly after the introduction of Δu_k, thus giving more chances for the detector to monitor the system.

Since the excitation signals act as additional disturbances to the system, it should be designed as on one hand having minimum effects on system performance and on the other hand, increasing the effectiveness in detecting attacks. In [95], a Hidden Markov Model (HMM) to select the statistical properties associated with the watermark was provided based on the tradeoff between desired detection performance and allowable control performance loss.

D. Attack Detection for Nonlinear Systems

In the real world, many industrial processes involve nonlinear properties due to their characteristics and external environment, which make the detection more challenging compared with linear systems. Recently, a variety of attack detection approaches have been developed for nonlinear systems.

1) Iterative state estimation: Iterative algorithms can be used in the state estimation for nonlinear systems [96], [97], where the system is linearized in each step. Similarly to (5), the state estimation can be obtained by solving the following optimization problem:

$$\min F(\mathbf{x}) = (\mathbf{z} - H(\mathbf{x}))^T \cdot W \cdot (\mathbf{z} - H(\mathbf{x})) \tag{17}$$

where H is the Jacobian matrix of $h(\mathbf{x})$. Details of solution algorithm can be found in [97], [98].

2) Kalman Filter-based method: A couple of modified Kalman Filter techniques have been constructed regarding attack detection in nonlinear systems, such as Extended Kalman Filter (EKF) [57], [99] and Unscented Kalman Filter (UKF) [100], [101]. EKF is an extension of KF where the system is linearized at each time step around the current operating mean and covariance. While UKF utilizes a series of sample data to approximate the probability density function to achieve a state estimation [102]. It was pointed out in [103] that UKF can achieve a high accuracy than EKF, as without the need to linearize the nonlinear equation.

3) Observer-based method: Observer-based method also plays a key role in attack detection of nonlinear systems in which the nonlinear part is modeled as system disturbance [104]–[107]. Therefore, the nonlinear function can be decoupled from the system. In [104], by means of proposed extended state observer, the states and total disturbances which involve both nonlinear dynamics and attacks were estimated. In addition, a nonlinear UIO was adopted in [105] to deal with the nonlinear term. In [106], a mode estimator was provided to estimate the states and attack vectors for a switched nonlinear system. In [107], a H_∞ filter was constructed for a T-S fuzzy-model-based nonlinear system.

Noticed that a common step of above mentioned method is to estimate or observe the states. Therefore, it is worthy to point out that χ^2 testing method can also be adopted if the system states are well estimated [93].

E. Attack Detection in the Presence of Noise

When taking into account the practical implementation, noise generated by sampling and actuation jitters and synchronization errors between system components usually presents in the systems [108]–[110]. The techniques considering systems without noise might not guarantee the effectiveness in systems with noise, which leads to the research on resilient state estimation methods [111].

Modified Kalman Filter is a well known approach for attack detection in systems with noise. In [101], an adaptive UKF was used for systems with Gaussian white noise. In terms of systems with stochastic noise, a multiple-model approach was adopted in [112] to estimate the hidden nodes of systems. Furthermore, a Mode Matched Filter was proposed in [109], [113] for systems in the presence of stochastic noise signals.

In addition, for systems with bounded noise, a novel robust filtering algorithm was proposed in [86] by Yong et al. considering the resilient state estimation problem that is robust to bounded multiplicative and additive modeling and noise errors. In [114], a H_∞ observer was provided for attack detection in the systems with bounded noise. Furthermore, a Luenberger-like observer was designed in [47] to make a state reconstruction under systems with sensor noise. In addition, system noise was filtered in [115] by means of a machine learning algorithm after analyzing the residual vectors both in time and frequency domains.

F. Discussion

State estimation is one approach to detect cyber-attacks. However, it may fail when detecting some intelligent attacks, where the false data might be consistent with the detection mechanism. Therefore, residuals should be carefully designed when considering the stealthy attacks. Similar to state estimation method, FDI approaches can also be adopted to monitor the systems with the help of proper designed observers and residuals. χ^2 detector is another approach to detect random attacks. However, it needs extra improvement in the case where the distribution of attack is unchanged. Binary hypothesis-based detection method is a kind of voting schemes, which leverages the potential presence of redundant sensors to detect attacks. Thus, it can only be used in multi-sensor systems. Model free method mainly detects system attacks based on data-driven modeling methods. Therefore, methodologies should be carefully adopted to build a reliable model. In addition, [116] is pointing out that information provided by current sensors may not be sufficient to estimate the states of the system irrespective of what method is used when there are inconsistencies between the process model and the states of the controlled process. Secure state estimation, which aims at estimating states from compromised sensors, is a useful approach to monitor actuator and sensor attacks. Furthermore, the attack can be identified and mitigated by the attack estimation method. In terms of nonlinear systems, the states can be monitored with modified state estimation approaches either by linearizing or nonlinear decoupling methods. In addition, various modified KF methods can be employed in systems with noise as well as discrete systems.

IV. Distributed Controllers

A. System Statement

System with distributed controllers can be described as:

$$\begin{cases} \dot{x}_i(t) = & A_i x_i(t) + B_i u_i(t) + E_i d_i(t) \\ & +R_i f_i(t) + \sum\limits_{j\in N_i} A_{ij} x_j(t) \\ y_i(t) = & C_i x_i(t) \end{cases} \tag{18}$$

where $i = 1, 2, \cdots, N$; $x_i(t)$, $u_i(t)$, $y_i(t)$, $d_i(t)$, and $f_i(t)$ are the state vectors, the input and output of system, external disturbances and attacks respectively; A_i, B_i, C_i, E_i, and R_i are system matrices with proper dimensions; $A_{ij}(t)$ is the coupling component among different systems; N_i is the neighbor of agent i. Comparing (1) with (18), the main difference between centralized and distributed controllers is the introduction of the coupling effect [117]–[122]. Therefore, the attack detection strategies should be robust not only against disturbances and measurement noises but also against incomplete measurements. The lack of knowledge of the distributed controller is the main challenge in the design of attack detection algorithms.

B. Attack Detection Design

Traditional attack detection schemes may not be applied to distributed systems, since not all measurements are available in distributed controllers. Generally, how to deal with the coupling component becomes the subject in the design of attack detection process.

1) Centralized method: The idea behind centralized method is to collect data from all nodes. Therefore, every distributed controller may have global measurements.

For instance, a design scheme for distributed control systems using a bank of observers was developed in [122], where each observer contains the model of the entire system. The system can be represented as follows:

$$\begin{cases} \dot{\mathbf{x}}(\mathbf{t}) = A\mathbf{x}(\mathbf{t}) + B\mathbf{u}(\mathbf{t}) + E\mathbf{d}(\mathbf{t}) + R\mathbf{f}(\mathbf{t}) \\ \mathbf{y}(\mathbf{t}) = C\mathbf{x}(\mathbf{t}) \end{cases} \tag{19}$$

where $\mathbf{x}(\mathbf{t}) = [x_1(t), x_2(t), \cdots, x_N(t)]^T$ are system states; $\mathbf{u}(\mathbf{t}) = [u_1(t), u_2(t), \cdots, u_N(t)]^T$ are system control input; $\mathbf{y}(\mathbf{t}) = [y_1(t), y_2(t), \cdots, y_N(t)]^T$ are system output; $\mathbf{d}(\mathbf{t}) = [d_1(t), d_2(t), \cdots, d_N(t)]^T$ are system disturbances and $\mathbf{f}(\mathbf{t}) = [f_1(t), f_2(t), \cdots, f_N(t)]^T$ are system attacks. The matrices can be written as:

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1N} \\ \vdots & \ddots & \vdots \\ A_{N1} & \cdots & A_{NN} \end{bmatrix}, C = \begin{bmatrix} C_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & C_N \end{bmatrix},$$

$$B = \begin{bmatrix} B_1 \\ \vdots \\ B_N \end{bmatrix}, E = \begin{bmatrix} E_1 \\ \vdots \\ E_N \end{bmatrix}, R = \begin{bmatrix} R_1 \\ \vdots \\ R_N \end{bmatrix}$$

By turning a distributed system into a centralized system, many centralized attack detection techniques mentioned in section III can be used here. However, this imposes a strict and heavy computational burden on each of the network nodes, especially when N is very large [118]. Thus, in order to detect the attack accurately, more effort should be put on the decoupling of states from its neighbors.

2) Singular value decomposition: Distributed attack detection can be achieved by Singular Value Decomposition (SVD) approach [118]–[120]. Taking advantage of SVD approach, (19) can be decoupled by following N isolated systems:

$$\hat{x}_{cli} = \widehat{A}_i x_{cli} + \widehat{B}_i \widehat{u}_i + \widehat{E}_i \widehat{d}_i + \widehat{R}_i \widehat{f}_i \tag{20}$$

where $\hat{x}_{cli}$, $\widehat{u}_i$, $\widehat{d}_i$, $\widehat{f}_i$, $\widehat{A}_i$, $\widehat{B}_i$, $\widehat{E}_i$, and $\widehat{R}_i$ are linear transformations of x, u_i, d_i, f_i, A_i, B_i, E_i, and R_i respectively.

It is shown that by applying the decomposition approach the attack detection problem of distributed systems can be solved by analyzing the problem of a set of decoupled systems (20), whose order and complexity are equal to that of a single agent. Then different attack detection methods could be designed based on the decoupled system.

3) Fault detection and identification method: Similarly, the FDI method can also be adopted to eliminate the influence of the coupling effect. The standard way is to consider the coupling effect as an unknown input, where the system model could be cast as:

$$\begin{cases} \dot{x}_i(t) = A_i x_i(t) + B_i u_i(t) + \bar{E}_i \bar{d}_i(t) + R_i f_i(t) \\ y_i(t) = C_i x_i(t) \end{cases} \tag{21}$$

where $\bar{E}_i \bar{d}_i(t) = E_i d_i(t) + \sum_{j \in N_i} A_{ij} x_j(t)$ is the combination of system disturbances and coupling effect of a distributed system.

Unknown Input Observer-based attack detection is the most adopted FDI approach in the literature, thanks to its explicit structure and design method [30], [58]–[62]. The typical structure of UIO is given by:

$$\begin{cases} \dot{z}_i(t) = F_i z_i(t) + T_i B_i u_i(t) + K_i y_i(t) \\ \hat{x}_i(t) = z_i(t) + H_i y_i(t) \end{cases} \tag{22}$$

where F_i, T_i, K_i, and H_i are matrices to be designed.

The residuals can be presented as:

$$r_i(t) = x_i(t) - \hat{x}_i(t) \tag{23}$$

Thus, the attack could be detected by comparing residuals (23) with a certain threshold. In order to design a UIO which converges to the actual value of the states, normally the system dynamics need to satisfy the following sufficient conditions [123]–[125]:

(i). Rank($C_i E_i$) = Rank(E_i);

(ii). The pair (C_i, $T_i A_i$) is detectable.

With these two conditions, feasibility of UIO should be considered in the designing processes.

In [61], Shames et al. gave a framework of distributed FDI scheme considering a double integrator dynamic system that fits some problems on power networks and distributed robotic systems. UIO can be always feasible in the proposed structure which is presented as follows:

$$\begin{cases} \dot{\xi}_i(t) = \zeta_i(t) \\ \dot{\zeta}_i(t) = u_i(t) + m_i(t) \end{cases} \tag{24}$$

where $m_i(t)$ are scalar known disturbances; $\xi_i(t)$ and $\zeta_i(t)$ are the scalar states; $u_i(t)$ are control input given by:

$$k_i u_i(t) = -l_i(t)\zeta_i(t) + \sum_{j \in N_i} \omega_{ij}(\xi_i(t) - \zeta_i(t)) \tag{25}$$

where k_i, l_i, $\omega_{ij} > 0$ for $i, j = 1, \cdots, N$.

4) Iteration method: Furthermore, Dörfler and Pasqualetti proposed a new decoupling method for the distributed systems in [29], [121] using waveform relaxation methods. The proposed iteration form of observer can be presented as:

$$\begin{cases} M_i \dot{\hat{x}}_i^{(k)}(t) = & (A_i + G_i C_i)\hat{x}_i^{(k)}(t) - G_i y_i(t) \\ & + \sum_{j \in N_i} A_{ij} \hat{x}_j^{(k-1)}(t) \\ r_i(t) = & y_i(t) - C_i \hat{x}_i^{(k)}(t) \end{cases} \tag{26}$$

where $k \in \mathbb{N}$ denotes the iteration index; $t \in [0, T]$ is the integration interval for some uniform time horizon $T > 0$; $\hat{x}_i(t)$ are the i-th estimates of $x_i(t)$ and output injection G_i is such that the pair $(M_i, A_i + G_i C_i)$ is regular and Hurwitz.

The system can be monitored by performing the following operations:

(i). Assuming $k = 0$ at start;

(ii). Set $k = k+1$, and compute signal $\hat{x}_i^{(k)}(t)$ by integrating the local filter (26);

(iii). Transmit $\hat{x}_i^{(k)}(t)$ to the j-th neighboring control center;

(iv). Update the input $\hat{x}_j^{(k)}(t)$ with the signal received from the j-th control center and go to (ii).

For k sufficiently large, the local residuals can be used for attack detection purposes.

C. Discussion

The main challenge in distributed attack detection comes to the fact that not all measurements are available for distributed controllers. More attention should by paid on the decouple of distributed systems. The centralized detection method has the feature of easy implementation. However, it is computational and may not be used in large-scale systems. Distributed detection can be realized by means of a singular value decomposition approach. Nevertheless, the computational complexity of SVD method also increases along the rising of distributed units, due to the needs to decompose the entire system. FDI based detection methods also play a role in distributed attack detection approaches, where neighboring states can be seen as an external disturbance. However, the feasibility of observers needs to be concerned in the designing process. The iteration method is an approach with less computational complexity. The limitation of this approach is the need for synchronous discrete time communication between neighboring controllers.

V. Potential Research Directions and Conclusions

A. Potential Research Directions

Attack detections on CPSs are an ongoing research topic, many new methodologies and results are required to meet various requirements in applications. Some potential research directions are suggested and listed as follows:

1) Distributed attack detection: Most of the existing results on attack detection methods in the literature are based on a centralized structure. However, the distributed control system becomes popular in recent years due to their lower computational complexity and the use of fewer network resources [61], [118]–[120], [122]. Due to the fact that a centralized detection method might not be used in a distributed system. Attack detection approaches for the distributed systems deserve further investigated.

2) Multi-attack detection: In real scenarios, multiple sensors or communication links might be attacked at the same time [12], [85], especially in the case where a large number of sensors are considered. However, most attack detection methods assume the single attack hypothesis in the system. Although methodologies based on multi-attack detection are of significant engineering importance, this research field still remains a number of challenges.

3) Scalable attack detection: Currently, attack detection methods have been studied in-depth for a single common system. However, those methods are undesirable for large scale cyber-physical systems due to high computational resources and communication bandwidth limitations [80]–[82], [118]. Therefore, scalability should be also considered in the design of attack detection algorithms. Order reducing approach may be a guide for such a system.

4) Detection of other attacks: Most existing literature only provide methods which are able to detect one specific kind of attack. However, they might not work against other types of attacks. For example, a well designed detection method for DoS attack may be ineffective for replay attack [55]. Therefore, the design of algorithms that can deal with various kinds of attack is of extremely importance.

B. Conclusions

A brief overview of recent developments on attack detection for cyber-physical systems has been presented. The controllers of different CPSs are divided into centralized and distributed controllers based on the knowledge of entire system. Specifically, existing centralized attack detection methodologies for LTI systems, sensor and actuator attack detection techniques, nonlinear systems and systems with noise are reviewed, respectively. Furthermore, attack detection approaches for distributed controllers have been surveyed following centralized method, singular value decomposition method, fault detection and identification method and iteration approach. Several potential research directions are discussed in terms of distributed attack detections, multi-attack detections, scalable attack detections and detections of other attacks.

References

[1] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber–physical system security for the electric power grid," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 210–224, 2011.

[2] C. Murguia, N. van de Wouw, and J. Ruths, "Reachable sets of hidden cps sensor attacks: Analysis and synthesis tools," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 2088–2094, 2017.

[3] H. Chen, "Applications of cyber-physical system: a literature review," *Journal of Industrial Integration and Management*, vol. 2, no. 03, p. 1750012, 2017.

[4] Y. Lu, "Cyber physical system (cps)-based industry 4.0: a survey," *Journal of Industrial Integration and Management*, vol. 2, no. 03, p. 1750014, 2017.

[5] S. K. Khaitan and J. D. McCalley, "Design techniques and applications of cyberphysical systems: A survey," *IEEE Systems Journal*, vol. 9, no. 2, pp. 350–365, 2014.

[6] R. Atat, L. Liu, H. Chen, J. Wu, H. Li, and Y. Yi, "Enabling cyber-physical communication in 5g cellular networks: challenges, spatial spectrum sensing, and cyber-security," *IET Cyber-Physical Systems: Theory & Applications*, vol. 2, no. 1, pp. 49–54, 2017.

[7] J. Wu, S. Guo, H. Huang, W. Liu, and Y. Xiang, "Information and communications technologies for sustainable development goals: state-of-the-art, needs and perspectives," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2389–2406, 2018.

[8] R. Atat, L. Liu, J. Wu, G. Li, C. Ye, and Y. Yang, "Big data meet cyber-physical systems: A panoramic survey," *IEEE Access*, vol. 6, pp. 73 603–73 636, 2018.

[9] E. A. Lee, "Cyber physical systems: Design challenges," in *2008 11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC)*. IEEE, 2008, pp. 363–369.

[10] C. Peng, H. Sun, M. Yang, and Y.-L. Wang, "A survey on security communication and control for smart grids under malicious cyber attacks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2019.

[11] M. S. Mahmoud, M. M. Hamdan, and U. A. Baroudi, "Modeling and control of cyber-physical systems subject to cyber attacks: A survey of recent advances and challenges," *Neurocomputing*, 2019.

[12] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic control*, vol. 59, no. 6, pp. 1454–1467, 2014.

[13] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, pp. 135–148, 2015.

[14] A. Teixeira, D. Pérez, H. Sandberg, and K. H. Johansson, "Attack models and scenarios for networked control systems," in *Proceedings of the 1st international conference on High Confidence Networked Systems*. ACM, 2012, pp. 55–64.

[15] H. Fawzi, P. Tabuada, and S. Diggavi, "Security for control systems under sensor and actuator attacks," in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*. IEEE, 2012, pp. 3412–3417.

[16] J. Slay and M. Miller, "Lessons learned from the maroochy water breach," in *International Conference on Critical Infrastructure Protection*. Springer, 2007, pp. 73–82.

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

[18] J. P. Farwell and R. Rohozinski, "Stuxnet and the future of cyber war," *Survival*, vol. 53, no. 1, pp. 23–40, 2011.

[19] C.-H. Lee, B.-K. Chen, N.-M. Chen, and C.-W. Liu, "Lessons learned from the blackout accident at a nuclear power plant in taiwan," *IEEE Transactions on Power Delivery*, vol. 25, no. 4, pp. 2726–2733, 2010.

[20] J. P. Conti, "The day the samba stopped [power blackouts]," *Engineering & Technology*, vol. 5, no. 4, pp. 46–47, 2010.

[21] Y. Liu and S. Hu, "Cyberthreat analysis and detection for energy theft in social networking of smart homes," *IEEE Transactions on Computational Social Systems*, vol. 2, no. 4, pp. 148–158, 2015.

[22] ——, "Smart home scheduling and cybersecurity: fundamentals," in *Smart Cities and Homes*. Elsevier, 2016, pp. 191–217.

[23] J. Liang, L. Sankar, and O. Kosut, "Vulnerability analysis and consequences of false data injection attack on power system state estimation," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3864–3872, 2015.

[24] S. Li, Y. Yılmaz, and X. Wang, "Quickest detection of false data injection attack in wide-area smart grids," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2725–2735, 2014.

[25] D. B. Rawat and C. Bajracharya, "Detection of false data injection attacks in smart grid communication systems," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1652–1656, 2015.

[26] 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 Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 411–423, 2016.

[27] K. Paridari, N. O'Mahony, A. E.-D. Mady, R. Chabukswar, M. Boubekeur, and H. Sandberg, "A framework for attack-resilient industrial control systems: Attack detection and controller reconfiguration," *Proceedings of the IEEE*, vol. 106, no. 1, pp. 113–128, 2017.

[28] A.-Y. Lu and G.-H. Yang, "Secure state estimation for cyber-physical systems under sparse sensor attacks via a switched luenberger observer," *Information sciences*, vol. 417, pp. 454–464, 2017.

[29] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE transactions on automatic control*, vol. 58, no. 11, pp. 2715–2729, 2013.

[30] A. J. Gallo, M. S. Turan, P. Nahata, F. Boem, T. Parisini, and G. Ferrari-Trecate, "Distributed cyber-attack detection in the secondary control of dc microgrids," in *2018 European Control Conference (ECC)*. IEEE, 2018, pp. 344–349.

[31] S. Altaf, A. Al-Anbuky, and H. GholamHosseini, "Fault diagnosis in a distributed motor network using artificial neural network," in *2014 International Symposium on Power Electronics, Electrical Drives, Automation and Motion*. IEEE, 2014, pp. 190–197.

[32] B. M. Sanandaji, E. Bitar, K. Poolla, and T. L. Vincent, "An abrupt change detection heuristic with applications to cyber data attacks on power systems," in *2014 American Control Conference*. IEEE, 2014, pp. 5056–5061.

[33] P. Arpaia, C. Manna, and G. Montenero, "Ant-search strategy based on likelihood trail intensity modification for multiple-fault diagnosis in sensor networks," *IEEE Sensors Journal*, vol. 13, no. 1, pp. 148–158, 2012.

[34] R. Han, L. Meng, G. Ferrari-Trecate, E. A. A. Coelho, J. C. Vasquez, and J. M. Guerrero, "Containment and consensus-based distributed coordination control to achieve bounded voltage and precise reactive power sharing in islanded ac microgrids," *IEEE Transactions on Industry Applications*, vol. 53, no. 6, pp. 5187–5199, 2017.

[35] R. Han, L. Meng, J. M. Guerrero, and J. C. Vasquez, "Distributed nonlinear control with event-triggered communication to achieve current-sharing and voltage regulation in dc microgrids," *IEEE Transactions on Power Electronics*, vol. 33, no. 7, pp. 6416–6433, 2017.

[36] J. M. Guerrero, J. C. Vasquez, J. Matas, L. G. De Vicuña, and M. Castilla, "Hierarchical control of droop-controlled ac and dc microgrids—a general approach toward standardization," *IEEE Transactions on industrial electronics*, vol. 58, no. 1, pp. 158–172, 2010.

[37] M. Tucci, L. Meng, J. M. Guerrero, and G. Ferrari-Trecate, "Stable current sharing and voltage balancing in dc microgrids: A consensus-based secondary control layer," *Automatica*, vol. 95, pp. 1–13, 2018.

[38] L. Meng, Q. Shafiee, G. F. Trecate, H. Karimi, D. Fulwani, X. Lu, and J. M. Guerrero, "Review on control of dc microgrids and multiple microgrid clusters," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 5, no. 3, pp. 928–948, 2017.

[39] X. Ge, F. Yang, and Q.-L. Han, "Distributed networked control systems: A brief overview," *Information Sciences*, vol. 380, pp. 117–131, 2017.

[40] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security (TISSEC)*, vol. 14, no. 1, p. 13, 2011.

[41] A. Teixeira, S. Amin, H. Sandberg, K. H. Johansson, and S. S. Sastry, "Cyber security analysis of state estimators in electric power systems," in *49th IEEE conference on decision and control (CDC)*. IEEE, 2010, pp. 5991–5998.

[42] R. Deng, G. Xiao, and R. Lu, "Defending against false data injection attacks on power system state estimation," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 1, pp. 198–207, 2015.

[43] L. Liu, M. Esmalifalak, Q. Ding, V. A. Emesih, and Z. Han, "Detecting false data injection attacks on power grid by sparse optimization," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 612–621, 2014.

[44] Y. Huang, J. Tang, Y. Cheng, H. Li, K. A. Campbell, and Z. Han, "Real-time detection of false data injection in smart grid networks: An adaptive cusum method and analysis," *IEEE Systems Journal*, vol. 10, no. 2, pp. 532–543, 2014.

[45] G. Dan and H. Sandberg, "Stealth attacks and protection schemes for state estimators in power systems," in *2010 First IEEE International Conference on Smart Grid Communications*. IEEE, 2010, pp. 214–219.

[46] G. Hug and J. A. Giampapa, "Vulnerability assessment of ac state estimation with respect to false data injection cyber-attacks," *IEEE Transactions on smart grid*, vol. 3, no. 3, pp. 1362–1370, 2012.

[47] Y. Shoukry and P. Tabuada, "Event-triggered state observers for sparse sensor noise/attacks," *IEEE Transactions on Automatic Control*, vol. 61, no. 8, pp. 2079–2091, 2015.

[48] W. Ao, Y. Song, and C. Wen, "Adaptive cyber-physical system attack detection and reconstruction with application to power systems," *IET Control Theory & Applications*, vol. 10, no. 12, pp. 1458–1468, 2016.

[49] A. A. Cardenas, S. Amin, and S. Sastry, "Secure control: Towards survivable cyber-physical systems," in *2008 The 28th International Conference on Distributed Computing Systems Workshops*. IEEE, 2008, pp. 495–500.

[50] Y. Mo, R. Chabukswar, and B. Sinopoli, "Detecting integrity attacks on scada systems," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 4, pp. 1396–1407, 2013.

[51] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, "Optimal linear cyber-attack on remote state estimation," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 4–13, 2016.

[52] Y. Mo and B. Sinopoli, "False data injection attacks in control systems," in *Preprints of the 1st workshop on Secure Control Systems*, 2010, pp. 1–6.

[53] ——, "On the performance degradation of cyber-physical systems under stealthy integrity attacks," *IEEE Transactions on Automatic Control*, vol. 61, no. 9, pp. 2618–2624, 2015.

[54] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli, "False data injection attacks against state estimation in wireless sensor networks," in *49th IEEE Conference on Decision and Control (CDC)*. IEEE, 2010, pp. 5967–5972.

[55] K. Manandhar, X. Cao, F. Hu, and Y. Liu, "Detection of faults and attacks including false data injection attack in smart grid using kalman filter," *IEEE transactions on control of network systems*, vol. 1, no. 4, pp. 370–379, 2014.

[56] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2009, pp. 911–918.

[57] A. Meng, H. Wang, S. Aziz, J. Peng, and H. Jiang, "Kalman filtering based interval state estimation for attack detection," *Energy Procedia*, vol. 158, pp. 6589–6594, 2019.

[58] S. Amin, X. Litrico, S. S. Sastry, and A. M. Bayen, "Cyber security of water scada systems—part ii: Attack detection using enhanced hydrodynamic models," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 5, pp. 1679–1693, 2012.

[59] J. Shi, X. He, Z. Wang, and D. Zhou, "Distributed fault detection for a class of second-order multi-agent systems: an optimal robust observer approach," *IET Control Theory & Applications*, vol. 8, no. 12, pp. 1032–1044, 2014.

[60] A. Teixeira, H. Sandberg, and K. H. Johansson, "Networked control systems under cyber attacks with applications to power networks," in *Proceedings of the 2010 American Control Conference*. IEEE, 2010, pp. 3690–3696.

[61] I. Shames, A. M. Teixeira, H. Sandberg, and K. H. Johansson, "Distributed fault detection for interconnected second-order systems," *Automatica*, vol. 47, no. 12, pp. 2757–2764, 2011.

[62] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "Distributed fault detection and isolation resilient to network model uncertainties," *IEEE transactions on cybernetics*, vol. 44, no. 11, pp. 2024–2037, 2014.

[63] S. X. Ding, *Model-based fault diagnosis techniques: design schemes, algorithms, and tools*. Springer Science & Business Media, 2008.

[64] H. M. Odendaal and T. Jones, "Actuator fault detection and isolation: An optimised parity space approach," *Control Engineering Practice*, vol. 26, pp. 222–232, 2014.

[65] M. Zhong, Y. Song, and S. X. Ding, "Parity space-based fault detection for linear discrete time-varying systems with unknown input," *Automatica*, vol. 59, pp. 120–126, 2015.

[66] M. Darouach, M. Zasadzinski, and S. J. Xu, "Full-order observers for linear systems with unknown inputs," *IEEE transactions on automatic control*, vol. 39, no. 3, pp. 606–609, 1994.

[67] B. Kailkhura, Y. S. Han, S. Brahma, and P. K. Varshney, "Asymptotic analysis of distributed bayesian detection with byzantine data," *IEEE Signal Processing Letters*, vol. 22, no. 5, pp. 608–612, 2014.

[68] ——, "Distributed bayesian detection in the presence of byzantine data," *IEEE transactions on signal processing*, vol. 63, no. 19, pp. 5250–5263, 2015.

[69] A. S. Rawat, P. Anand, H. Chen, and P. K. Varshney, "Collaborative spectrum sensing in the presence of byzantine attacks in cognitive radio

networks," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 774–786, 2010.

[70] W.-H. Chen, S.-H. Hsu, and H.-P. Shen, "Application of svm and ann for intrusion detection," *Computers & Operations Research*, vol. 32, no. 10, pp. 2617–2634, 2005.

[71] W. Hu, Y. Liao, and V. R. Vemuri, "Robust support vector machines for anomaly detection in computer security." in *ICMLA*, 2003, pp. 168–174.

[72] C. Alippi, S. Ntalampiras, and M. Roveri, "Model-free fault detection and isolation in large-scale cyber-physical systems," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 1, no. 1, pp. 61–71, 2016.

[73] A. B. Sharma, H. Chen, M. Ding, K. Yoshihira, and G. Jiang, "Fault detection and localization in distributed systems using invariant relationships," in *2013 43rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. IEEE, 2013, pp. 1–8.

[74] Y. He, G. J. Mendis, and J. Wei, "Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2505–2516, 2017.

[75] K. Hamedani, L. Liu, R. Atat, J. Wu, and Y. Yi, "Reservoir computing meets smart grids: Attack detection using delayed feedback networks," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 2, pp. 734–743, 2017.

[76] K. Hamedani, L. Liu, S. Hu, J. Ashdown, J. Wu, and Y. Yi, "Detecting dynamic attacks in smart grids using reservoir computing: A spiking delayed feedback reservoir based approach," *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2019.

[77] X. Huang and J. Dong, "Reliable control of cyber-physical systems under sensor and actuator attacks: An identifier-critic based integral sliding-mode control approach," *Neurocomputing*, vol. 361, pp. 229–242, 2019.

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

[79] L. An and G.-H. Yang, "Distributed secure state estimation for cyber–physical systems under sensor attacks," *Automatica*, vol. 107, pp. 526–538, 2019.

[80] ——, "Secure state estimation against sparse sensor attacks with adaptive switching mechanism," *IEEE Transactions on Automatic Control*, vol. 63, no. 8, pp. 2596–2603, 2017.

[81] A.-Y. Lu and G.-H. Yang, "Secure switched observers for cyber-physical systems under sparse sensor attacks: a set cover approach," *IEEE Transactions on Automatic Control*, 2019.

[82] L. An and G.-H. Yang, "State estimation under sparse sensor attacks: A constrained set partitioning approach," *IEEE Transactions on Automatic Control*, 2018.

[83] Y. Shoukry, P. Nuzzo, A. Puggelli, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada, "Secure state estimation for cyber-physical systems under sensor attacks: A satisfiability modulo theory approach," *IEEE Transactions on Automatic Control*, vol. 62, no. 10, pp. 4917–4932, 2017.

[84] C. Lee, H. Shim, and Y. Eun, "On redundant observability: from security index to attack detection and resilient state estimation," *IEEE Transactions on Automatic Control*, vol. 64, no. 2, pp. 775–782, 2018.

[85] M. Pajic, I. Lee, and G. J. Pappas, "Attack-resilient state estimation for noisy dynamical systems," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 82–92, 2016.

[86] S. Z. Yong, M. Q. Foo, and E. Frazzoli, "Robust and resilient estimation for cyber-physical systems under adversarial attacks," in *2016 American Control Conference (ACC)*. IEEE, 2016, pp. 308–315.

[87] S. Z. Yong, M. Zhu, and E. Frazzoli, "A unified filter for simultaneous input and state estimation of linear discrete-time stochastic systems," *Automatica*, vol. 63, pp. 321–329, 2016.

[88] ——, "Simultaneous input and state estimation for linear time-varying continuous-time stochastic systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 5, pp. 2531–2538, 2016.

[89] 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.

[90] 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*, p. 108687, 2019.

[91] Q. 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, no. 3, pp. 717–729, 2013.

[92] M. Talebi, C. Li, and Z. Qu, "Enhanced protection against false data injection by dynamically changing information structure of microgrids," in *2012 IEEE 7th Sensor Array and Multichannel Signal Processing Workshop (SAM)*. IEEE, 2012, pp. 393–396.

[93] G. Chaojun, P. Jirutitijaroen, and M. Motani, "Detecting false data injection attacks in ac state estimation," *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2476–2483, 2015.

[94] D. Muniraj and M. Farhood, "Detection and mitigation of actuator attacks on small unmanned aircraft systems," *Control Engineering Practice*, vol. 83, pp. 188–202, 2019.

[95] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 93–109, 2015.

[96] J. Wang, L. C. Hui, S.-M. Yiu, E. K. Wang, and J. Fang, "A survey on cyber attacks against nonlinear state estimation in power systems of ubiquitous cities," *Pervasive and Mobile Computing*, vol. 39, pp. 52–64, 2017.

[97] A. Abur and A. G. Exposito, *Power system state estimation: theory and implementation*. CRC press, 2004.

[98] L. Mili, M. Cheniae, N. Vichare, and P. J. Rousseeuw, "Robust state estimation based on projection statistics [of power systems]," *IEEE Transactions on Power Systems*, vol. 11, no. 2, pp. 1118–1127, 1996.

[99] S. Liu, G. Wei, Y. Song, and Y. Liu, "Extended kalman filtering for stochastic nonlinear systems with randomly occurring cyber attacks," *Neurocomputing*, vol. 207, pp. 708–716, 2016.

[100] H. Wang, A. Meng, Y. Liu, X. Fu, and G. Cao, "Unscented kalman filter based interval state estimation of cyber physical energy system for detection of dynamic attack," *Energy*, vol. 188, p. 116036, 2019.

[101] R. M. Asl, Y. S. Hagh, S. Simani, and H. Handroos, "Adaptive square-root unscented kalman filter: An experimental study of hydraulic actuator state estimation," *Mechanical Systems and Signal Processing*, vol. 132, pp. 670–691, 2019.

[102] H. Liu, D. Liu, C. Lu, and X. Wang, "Fault diagnosis of hydraulic servo system using the unscented k alman filter," *Asian Journal of Control*, vol. 16, no. 6, pp. 1713–1725, 2014.

[103] M. Laila, N. Naveen, and K. Vishakh, "Comparison of estimation capabilities of extended and unscented kalman filter in an rlv," *Int. J. Adv. Res. Elect. Electron. Instrum. Eng*, vol. 2, no. 7, pp. 3480–3488, 2013.

[104] Y. Yu and Y. Yuan, "Event-triggered active disturbance rejection control for nonlinear network control systems subject to dos and physical attacks," *ISA transactions*, 2019.

[105] X. Wang, X. Luo, M. Zhang, and X. Guan, "Distributed detection and isolation of false data injection attacks in smart grids via nonlinear unknown input observers," *International Journal of Electrical Power & Energy Systems*, vol. 110, pp. 208–222, 2019.

[106] H. Kim, P. Guo, M. Zhu, and P. Liu, "Attack-resilient estimation of switched nonlinear cyber-physical systems," in *2017 American Control Conference (ACC)*. IEEE, 2017, pp. 4328–4333.

[107] Z. Gu, X. Zhou, T. Zhang, F. Yang, and M. Shen, "Event-triggered filter design for nonlinear cyber–physical systems subject to deception attacks," *ISA transactions*, 2019.

[108] M. Pajic, J. Weimer, N. Bezzo, P. Tabuada, O. Sokolsky, I. Lee, and G. J. Pappas, "Robustness of attack-resilient state estimators," in *ICCPS'14: ACM/IEEE 5th International Conference on Cyber-Physical Systems (with CPS Week 2014)*. IEEE Computer Society, 2014, pp. 163–174.

[109] S. Z. Yong, "Simultaneous input and state set-valued observers with applications to attack-resilient estimation," in *2018 Annual American Control Conference (ACC)*. IEEE, 2018, pp. 5167–5174.

[110] Y. Nakahira and Y. Mo, "Dynamic state estimation in the presence of compromised sensory data," in *2015 54th IEEE Conference on Decision and Control (CDC)*. IEEE, 2015, pp. 5808–5813.

[111] M. Pajic, P. Tabuada, I. Lee, and G. J. Pappas, "Attack-resilient state estimation in the presence of noise," in *2015 54th IEEE Conference on Decision and Control (CDC)*. IEEE, 2015, pp. 5827–5832.

[112] S. Z. Yong, M. Zhu, and E. Frazzoli, "Simultaneous mode, input and state estimation for switched linear stochastic systems," *arXiv preprint arXiv:1606.08323*, 2016.

[113] ——, "Switching and data injection attacks on stochastic cyber-physical systems: Modeling, resilient estimation, and attack mitigation," *ACM Transactions on Cyber-Physical Systems*, vol. 2, no. 2, p. 9, 2018.

[114] M. Khajenejad and S. Z. Yong, "Simultaneous input and state set-valued h∞-observers for linear parameter-varying systems," in *American Control Conference*, 2019.

[115] C. M. Ahmed, M. Ochoa, J. Zhou, A. P. Mathur, R. Qadeer, C. Murguia, and J. Ruths, "Noiseprint: attack detection using sensor and process noise fingerprint in cyber physical systems," in *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*. ACM, 2018, pp. 483–497.

[116] S. K. Damodaran and P. D. Rowe, "Limitations on observability of effects in cyber-physical systems," in *Proceedings of the 6th Annual Symposium on Hot Topics in the Science of Security*. ACM, 2019, p. 2.

[117] F. Pasqualetti, F. Dörfler, and F. Bullo, "A divide-and-conquer approach to distributed attack identification," in *2015 54th IEEE Conference on Decision and Control (CDC)*. IEEE, 2015, pp. 5801–5807.

[118] M. Davoodi, K. Khorasani, H. Talebi, and H. Momeni, "A robust semi-decentralized fault detection strategy for multi-agent systems: An application to a network of micro-air vehicles," *International Journal of Intelligent Unmanned Systems*, vol. 1, no. 1, pp. 21–35, 2013.

[119] M. R. Davoodi, K. Khorasani, H. A. Talebi, and H. R. Momeni, "Distributed fault detection and isolation filter design for a network of heterogeneous multiagent systems," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 3, pp. 1061–1069, 2013.

[120] M. Davoodi, N. Meskin, and K. Khorasani, "Simultaneous fault detection and consensus control design for a network of multi-agent systems," *Automatica*, vol. 66, pp. 185–194, 2016.

[121] F. Dörfler, F. Pasqualetti, and F. Bullo, "Distributed detection of cyber-physical attacks in power networks: A waveform relaxation approach," in *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2011, pp. 1486–1491.

[122] S. X. Ding, P. Zhang, C. Chihaia, W. Li, Y. Wang, and E. Ding, "Advanced design scheme for fault tolerant distributed networked control systems," *IFAC Proceedings Volumes*, vol. 41, no. 2, pp. 13 569–13 574, 2008.

[123] S. Sundaram and C. N. Hadjicostis, "Delayed observers for linear systems with unknown inputs," *IEEE Transactions on Automatic Control*, vol. 52, no. 2, pp. 334–339, 2007.

[124] S. Z. Yong, M. Zhu, and E. Frazzoli, "Simultaneous input and state estimation with a delay," in *2015 54th IEEE Conference on Decision and Control (CDC)*. IEEE, 2015, pp. 468–475.

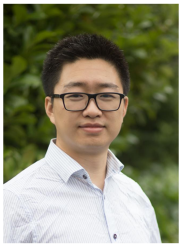
[125] A. Willsky, "On the invertibility of linear systems," *IEEE Transactions on Automatic Control*, vol. 19, no. 3, pp. 272–274, 1974.

Josep M. Guerrero (S'01-M'04-SM'08-FM'15) received the B.S. degree in telecommunications engineering, the M.S. degree in electronics engineering, and the Ph.D. degree in power electronics from the Technical University of Catalonia, Barcelona, in 1997, 2000 and 2003, respectively. Since 2011, he has been a Full Professor with the Department of Energy Technology, Aalborg University, Denmark, where he is responsible for the Microgrid Research Program. From 2014 he is chair Professor in Shandong University; from 2015 he is a distinguished guest Professor in Hunan University; and from 2016 he is a visiting professor fellow at Aston University, UK, and a guest Professor at the Nanjing University of Posts and Telecommunications. From 2019, he became a Villum Investigator by The Villum Fonden, which supports the Center for Research on Microgrids (CROM) at Aalborg University, being Prof. Guerrero the founder and Director of the same centre (www.crom.et.aau.dk).

His research interests is oriented to different microgrid aspects, including power electronics, distributed energy-storage systems, hierarchical and cooperative control, energy management systems, smart metering and the internet of things for AC/DC microgrid clusters and islanded minigrids. Specially focused on microgrid technologies applied to offshore wind, maritime microgrids for electrical ships, vessels, ferries and seaports, and space microgrids applied to nanosatellites and spacecrafts. Prof. Guerrero is an Associate Editor for a number of IEEE TRANSACTIONS. He has published more than 500 journal papers in the fields of microgrids and renewable energy systems, which are cited more than 50,000 times. He received the best paper award of the IEEE Transactions on Energy Conversion for the period 2014-2015, and the best paper prize of IEEE-PES in 2015. As well, he received the best paper award of the Journal of Power Electronics in 2016. During six consecutive years, from 2014 to 2019, he was awarded by Clarivate Analytics (former Thomson Reuters) as Highly Cited Researcher. In 2015 he was elevated as IEEE Fellow for his contributions on "distributed power systems and microgrids."

Peilin Xie (S'19) received the B.S. degree in Electrical Engineering from Beijing Jiaotong University, Beijing, China in 2015, and the M.S. degree in Electrical Engineering and Automation from North China Electric Power University, Beijing, China, in 2018. She is currently working toward her Ph.D. degree with the Department of Energy Technology, Aalborg University, Aalborg, Denmark.

Her research mainly focus on the power management control for shipboard microgrids.



Sen Tan (S'20) received the B.S. degree in Automation, the M.S. degree in Control Engineering both from Northeastern University, Liaoning, China, in 2014 and 2017 respectively. He is currently pursuing the Ph.D. degree with the Department of Energy Technology, Aalborg University, Denmark.

His research interests include distributed control in Microgrid, fault detection and motor drive technologies.

Renke Han (S'16-M'18) received the B.S. degree in Automation, the M.S. degree in Control Theory and Control Engineering both from Northeastern University, Liaoning, China, in 2013 and 2015 respectively. He received the Ph.D. degree in Power Electronics Systems from Aalborg University, Aalborg, Denmark, in 2018.

From February 2017 to September 2017, he was as a Visiting Scholar with Laboratoire d'Automatique, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. Since November 2018, he has been with Power Electronics Group, University of Oxford, UK, as a postdoctoral researcher. His research interests include multiport converter design, modeling, control and stability of Microgrid, embedded system development, and cyber-physical system.

He received an outstanding presentation award in Annual Conference of the IEEE Industrial Electronics Society, Italy in 2016 and the Outstanding Master Degree Thesis Award from Liaoning Province, China, in 2014.

Juan C. Vasquez (M'12-SM'14) received the B.S. degree in electronics engineering from the Autonomous University of Manizales, Manizales, Colombia, and the Ph.D. degree in automatic control, robotics, and computer vision from BarcelonaTech-UPC, Spain, in 2004 and 2009, respectively. In 2011, He was Assistant Professor and in 2014, Associate Professor at the Department of Energy Technology, Aalborg University, Denmark. In 2019, He became Professor in Energy Internet and Microgrids and currently He is the Co-Director of the Villum Center for Research on Microgrids (see crom.et.aau.dk). He was a Visiting Scholar at the Center of Power Electronics Systems (CPES) at Virginia Tech, USA and a visiting professor at Ritsumeikan University, Japan. His current research interests include operation, advanced hierarchical and cooperative control, optimization and energy management applied to distributed generation in AC/DC Microgrids, maritime microgrids, advanced metering infrastructures and the integration of Internet of Things and Energy Internet into the SmartGrid. Prof.Vasquez is an Associate Editor of IET POWER ELECTRONICS and a Guest Editor of the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS Special Issue on Energy Internet. Prof. Vasquez was awarded as Highly Cited Researcher by Thomson Reuters from 2017 to 2019 and He was the recipient of the Young Investigator Award 2019. He has published more than 450 journal papers in the field of Microgrids, which in total are cited more than 19000 times. Dr. Vasquez is currently a member of the IEC System Evaluation Group SEG4 on LVDC Distribution and Safety for use in Developed and Developing Economies, the Renewable Energy Systems Technical Committee TC-RES in IEEE Industrial Electronics, PELS, IAS, and PES Societies.