Distributed Attack Detection and Secure Estimation of Networked Cyber-Physical Systems against False Data Injection Attacks and Jamming Attacks

Yanpeng Guan and Xiaohua Ge

***Abstract*—This paper is concerned with the problem of joint distributed attack detection and distributed secure estimation for a networked cyber-physical system under physical and cyber attacks. The system is monitored by a wireless sensor network in which a group of sensors is spatially distributed and the sensors' measurements are broadcast to remote estimators via a wireless network medium. A malicious adversary simultaneously launches a false data injection attack at the physical system layer to intentionally modify the system's state and jamming attacks at the cyber layer to block the wireless transmission channels between sensors and remote estimators. The sensors' measurements can be randomly dropped with mathematical probability if the corresponding transmission channels are deliberately jammed by the adversary. Resilient attack detection estimators are delicately constructed to provide locally reliable state estimations and detect the false data injection attack. Then, criteria for analyzing the estimation performance and designing the desired estimators are derived to guarantee the solvability of the problem. Finally, the effectiveness of the proposed approach is shown through an illustrative example.**

***Index Terms*—Distributed attack detection, distributed secure estimation, jamming attack, false data injection attack, wireless sensor network.**

I. Introduction

CYBER-physical systems (CPSs) represent a new generation of systems that integrate computation resources, communication medium and physical processes [1], [2]. CPSs have been intensively applied in a large number of practical areas, such as aerospace, civil infrastructures, power grids, water and gas distribution networks, and transportation networks. Different from classical control systems, the operation and communication of CPSs often occur through some shared wired or wireless network medium, such as the specialized real-time control networks CAN, BACnet and Fieldbus or the general-purpose wireless data communication networks Ethernet and Internet [3], [4]. This makes CPSs more open to the cyber-world [5]. In addition to being prone to failures or attacks on the physical processes as in classical control systems, CPSs are vulnerable to malicious cyber security threats on the data transmission or communication layer. There is no doubt that any severe attack on CPSs, launched in either the physical-process-domain or the cyber-domain, can have a significant impact on the economy, environment or even human life, such as attacks on the national power grids [6]. Therefore, it is of fundamental significance to consider security issues when designing safe and reliable CPSs.

The distributed secure estimation problem under consideration is motivated by security concerns of CPSs operated over wireless sensor networks (WSNs). Generally, a WSN consists of a large number of observation nodes that are spatially deployed in a monitoring region of the physical process or target plant. These nodes, which possess data sensing, processing and communication capabilities, collaborate among themselves to build a cooperative information processing paradigm. However, the broadcast nature of the nodes makes WSNs vulnerable to various malicious threats [7], because WSNs require the nodes to cooperatively perform an overall monitoring or estimation task by broadcasting their observations (e.g., measurements) among the neighboring nodes. In other words, the nodes' observations can be potentially manipulated by cyber attacks. A key concern of distributed secure estimation is how to assess the trustworthiness of nodes' measurements and compute locally reliable estimations of the physical system's state with the caveat that some of the nodes' measurements can be corrupted by a malicious adversary.

A. Relevant Work on Secure Estimation against Specific Attacks

Note that it is generally challenging to describe attacks by accurate mathematical models as malicious attacks usually occur in intelligent and erratic ways. The existing literature on the analysis of vulnerabilities of CPSs to malicious attacks has been confined to exploring some specific attacks against particular CPSs. For example, in [2], [8], *integrity attacks (or deception attacks)* on state estimation systems were defined, where integrity attacks intentionally compromised the integrity of sensor measurements or control packets. In [5], the problem of state estimation and control for linear systems was considered when some of the sensors or actuators were hijacked by deception attacks. Resilient estimators and output feedback controllers were designed such that the state of the system was accurately reconstructed and the resilience of the closed-loop system was improved. In [9], the effect of *sparse sensor attacks* was considered to achieve a state reconstruction of discrete-time linear CPSs where an adversary arbitrarily falsified measurements of a subset of sensors. In [10], [11],

This work was supported in part by the National Natural Science Foundation of China under Grant 61603232 and the Key project of Natural Science Foundation of Zhejiang Province, China under Grant LZ13F030001.

Y. Guan is with the Department of Automation, Shanxi University, Taiyuan 030006, China (e-mail: y.guan@sxu.edu.cn).

X. Ge is with the School of Software and Electrical Engineering, Swinburne University of Technology, John Street, Hawthorn, Melbourne, VIC 3122, Australia (e-mail: xge@swin.edu.au).

false data injection (FDI) attacks were considered in state estimation frameworks for electric power grids. Generally, FDI attacks are known as specific deception attacks or integrity attacks, where an adversary could access and modify the physical system's state, sensor data, or control commands by introducing arbitrary errors, fake information, or faults. In [12], the effect of FDI attacks on state estimation was studied over a sensor network. The attacker therein hijacked a subset of sensors and sent fake sensor measurements to compromise the integrity of the state estimator. Then, a steady-state Kalman filter and a failure detector were designed to provide a quantitative measure of the resilience of the system to such attacks. Very recently, [13] considered a specific FDI attack called a *fake-acknowledge attack* against remote state estimation for CPSs. The attacker was able to modify the acknowledgement-based online power schedule signal from the remote estimator and send fake information to the sensor. A game-theoretic framework was built to investigate the equilibrium for both the sensor and the attacker. In [6], *denial-of-service (DoS) attacks (or jamming attacks)* were studied for the remote state estimation of CPSs where the wireless channel from a sensor to a remote estimator was jammed by an external attacker. DoS attacks aim at deteriorating the communication channels to prevent information exchange, usually either sensor data or control commands, between components of CPSs. Note that a frequently used DoS technique is to launch jamming attacks on communication channels by interfering with their radio frequencies [14], [15]. By formulating a game-theoretic framework, the interactive decision-making process between a manipulated sensor and an energy-constrained attacker was investigated in [6]. To maximize the impact of DoS attacks on CPSs, [15], [16] presented optimal attack scheduling strategies for energy-constrained attackers. Hence, the attacker was able to decide when and where to jam the communication channel at each sampling time so as to degrade the remote estimation performance. Another particular form of attacks can be found in [17] where the effect of *replay attacks* (through which the sensor data or control commands were maliciously repeated) on control performance for CPSs was analyzed.

B. Relevant Work on Detection and/or Identification against False Data Injection Attacks

Despite the rich body of research about secure estimation in CPSs, there appear to be only a few studies on detecting and identifying FDI attacks in CPSs, see, e.g., [1], [11], [18]–[20]. Depending on different detection techniques, these results can arguably be classified into *three categories*. The first category is based on *statistical tests*. For example, in [18], the authors proposed a distributed average consensus algorithm in which each networked node locally computed the detection test statistic. The statistical distribution of the nodes' data was then exploited to devise techniques for mitigating the influence of data falsifying on the detection system. In [19], the problem of detecting and mitigating data injection attacks was studied in randomized gossip-based sensor networks. By analyzing the statistics of the sensors' states, decentralized consensus strategies were designed to detect and localize insider attackers. The second category is based on *data time-stamps*. For example, in [20], data time-stamps were used to detect the anomalies caused by the malicious node by evaluating the (average) temporal difference of the values held by normal nodes. The third category is based on *estimation residuals* and is inspired by the existing fault diagnosis/tolerance literature [21]–[24]. For example, in [11], a distributed estimation and false data detection algorithm was proposed to monitor the operation condition of a power network subject to FDI attacks. By analyzing the properties of an estimation residual between the measurement and its estimation, the presented algorithm detected the false data among the network measurements. Following similar analysis and design procedures, both centralized and distributed attack detection and identification monitors were proposed in [1] for a class of descriptor CPSs subject to attacks that affect the state and the measurements.

C. Motivations

A crucial feature of CPSs over WSNs is that system components such as observation nodes or sensor nodes are geographically distributed. This poses a significant difficulty in acquiring data from these spatially distributed nodes, especially in the presence of malicious attacks. Whereas, the majority of the existing results regarding secure estimation of CPSs are limited to the case of a single observation or sensor node (see, e.g., [5], [6], [8]–[10] and references therein), which renders the secure estimation algorithms therein inapplicable in WSN-based CPSs. On the other hand, the existing literature focuses mainly on one specific type of attack, and few results consider the simultaneous presence of various attacks on practical CPSs. In fact, it is quite common for a cunning attacker to launch different attacks on practical CPSs at the same time. Take automotive vehicles as an example, an attacker might be able to simultaneously compromise a car's external vehicle interfaces and internal network buses to pose threats to the vehicle control sub-systems [25]. To the best of the authors' knowledge, there are relatively few studies that have tackled the conjunct problem of attack detection and secure estimation for CPSs carried over WSNs when an adversary launches malicious attacks in both the physical-process-domain and cyber-domain, which motivates the present study.

D. Contributions

In this paper, we will address a distributed attack detection and secure estimation problem for a CPS over a WSN subject to both an FDI attack and jamming attacks. More specifically, the FDI attack will be launched by an attacker at the physical system layer so as to modify the system's state. Jamming attacks, however, will be considered during the wireless communication from sensors to remote estimators at the cyber layer. Unlike the information theoretic studies on secure communication, which primarily involve the protection of data and/or IT services, we will concentrate on investigating the distributed estimation performance under the attacks from a system theoretic perspective. We summarize the main contributions of this paper as follows.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2017.2749959, IEEE Transactions on Signal and Information Processing over Networks

- *A refined compensation-based measurement output model will be presented for each sensor*. A direct impact of jamming attacks on wireless transmission channels is that sensors' measurements will be randomly dropped with mathematical probability if the corresponding channels are deliberately jammed. Based on this measurement model, each remote estimator will pro-actively admit and utilize the corrupted sensor measurements from itself and its neighboring estimators to compute a local estimation.
- *Resilient attack detection estimators will be delicately constructed to deal with the simultaneous effects of the FDI attack, jamming attacks and process and measurement noises*. In particular, to deal with the FDI attack, each estimator will run a two-step attack detection mechanism to discern when the occurrence of the FDI attack can be detected and alarmed; and to handle the jamming attacks, each estimator will adopt the compensated measurements to increase the resilience of the estimation system. Based on the proposed estimators, the WSN-based secure estimation problem under attacks and noises will be mapped into an H_∞ estimation problem of an augmented estimation error system.
- *Criteria for analyzing secure estimation performance and designing desired estimators will be derived to guarantee the feasibility of the proposed distributed attack detection and secure estimation problem*. We will analytically and numerically investigate the impact of the FDI attack, jamming attacks and noises on the estimation performance, and show that under what conditions the resultant estimation error system will converge even in the presence of such attacks as well as process and measurement noises.

The reminder of this paper is organized as follows. In Section II, a compensation-based measurement model is presented and resilient attack detection estimators are constructed. The problem of distributed attack detection and secure estimation we propose to solve in this paper is also formulated at the end of this section. Section III presents the main results on secure estimation performance analysis and estimator design. Furthermore, an extension of the proposed results to the case of uncertain measurement-transmission probability is provided. In Section IV, an industrial continuous-stirred tank reactor model is employed to illustrate the effectiveness of the proposed method. Finally, Section V concludes the paper.

II. Problem Formulation

A. Notations

Throughout the paper, $\mathbb{R}^n$ stands for the n-dimensional Euclidean space and $\mathbb{R}^{n\times m}$ represents the set of all the real $n\times m$ matrices. For symmetric matrices X and Y, the notation $X<Y$ means that $X-Y$ is negative definite. $Prob\{\cdot\}$ represents the occurrence probability of an event. $\mathbb{E}\{\cdot\}$ represents the mathematical expectation of a stochastic variable. $\sup$ denotes the supremum of a set. $\|\cdot\|$ denotes the induced matrix 2-norm or the Euclidean vector norm as appropriate. $\otimes$ stands for the Kronecker product for matrices. $diag\{\cdot\}$ represents a diagonal matrix. $\mathbb{N}$ denotes the set of nonnegative integers. I is an identity matrix with an appropriate dimension. Let asterisk '*' denote a term that is induced by symmetry in symmetric block matrices. The superscript 'T' denotes the transpose of a matrix with vectors as a special case. If a matrix is invertible, the superscript '-1' represents the matrix inverse. The symbol $\sum$ denotes the summation of a sequence. The space of square-summable vector functions over $[0,\infty)$ is denoted as $l_2[0,\infty)$ and for any $w(k)\in l_2[0,\infty)$, its norm is given by $\|w(k)\|=\sqrt{\sum_{k=0}^{\infty}w^T(k)w(k)}$. Matrices, if not explicitly stated, are assumed to have appropriate dimensions.

B. System Dynamics under False Data Injection Attacks

Consider that the physical system is a discrete-time linear-invariant system of the following form

$$s(k+1)=As(k)+Bw(k)+Ep(k),\ s(0)=s_0 \tag{1}$$

for all $k\in\mathbb{N}$, where $s(k)\in\mathbb{R}^{n_s}$ is the state vector of the system at the k-th time step; $w(k)\in\mathbb{R}^{n_w}$ belonging to $l_2[0,\infty)$ is the process noise vector at the k-th time step; $p(k)\in\mathbb{R}^{n_p}$ is the false data injection (FDI) attack vector to be detected at the k-th time step. Here, the vector $p(k)$ is injected by the malicious attacker at the physical system layer to intentionally manipulate the system's state; s_0 is the initial state of the system; and A,B and E are known constant matrices with appropriate dimensions.

C. Communication Topology

In the following, a group of N spatially distributed sensor nodes will be deployed to monitor the system described in (1) and N cooperative estimator nodes which form an estimator network will be designed to compute local estimations of the system's state $s(k)$, as illustrated in Fig. 1. Moreover, sensors will be responsible for measuring the system's state and broadcasting their measurements to remote estimators. However, the estimators will coordinate their local estimations and received measurements with only their neighboring estimators in their communication ranges in order to achieve a satisfactory cooperative estimation task. We first recall some basic concepts of graph theory.

Denote a weighted directed graph by $\mathcal{G}=(\mathcal{V},\mathcal{E},\mathcal{A})$, where $\mathcal{V}=\{1,2,\cdots,N\}$ is the index set of N nodes, $\mathcal{E}\subseteq\mathcal{V}\times\mathcal{V}$ represents the edge set of paired nodes and $\mathcal{A}=[a_{ij}]\in\mathbb{R}^{N\times N}$ stands for the weighted adjacency matrix with positive adjacency elements a_{ij}. Then, the communication topology among the N estimator nodes can be modeled by the digraph $\mathcal{G}$. Moreover, $a_{ij}>0\Leftrightarrow(i,j)\in\mathcal{E}$ which means that node i can receive information from node j or node j can send its information to node i. It is assumed that self-loops exist in the graph, i.e., $a_{ii}>0$, $i\in\mathcal{V}$. The set of neighbors of node $i\in\mathcal{V}$ plus the node itself are denoted by $\mathcal{N}_i=\{j\in\mathcal{V}:(i,j)\in\mathcal{E}\}$.

D. Ideal Measurement Output Model

At time step k, the ideal measurement output model of system (1) on sensor i is given by

$$y_i(k)=C_is(k),\ \forall\ i\in\mathcal{V}, \tag{2}$$

where $y_i(k)\in\mathbb{R}^{n_y}$ is the measurement output on sensor i and will be broadcast through a wireless transmission channel to a

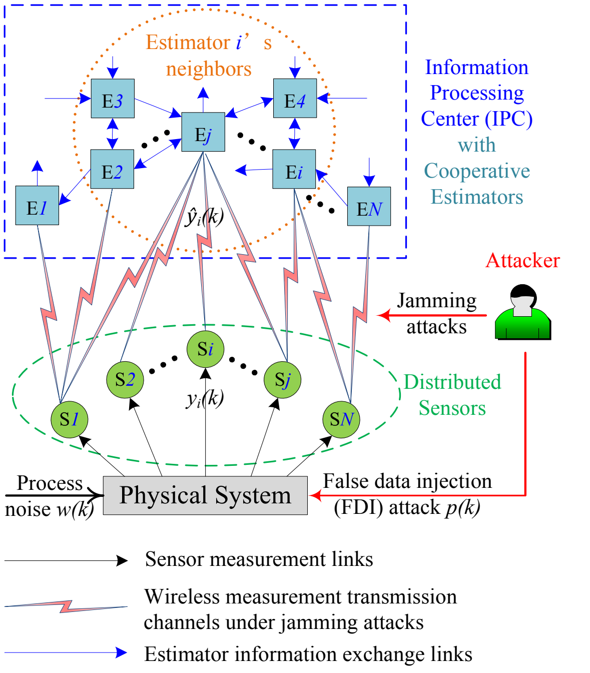

Estimator i's neighbors
E3
E4
Ej
E2
Ei
E1
EN
Information Processing Center (IPC) with Cooperative Estimators
ŷi(k)
Attacker
Jamming attacks
Si
S2
Sj
S1
SN
yi(k)
Distributed Sensors
Process noise w(k)
Physical System
False data injection (FDI) attack p(k)
Sensor measurement links
Wireless measurement transmission channels under jamming attacks
Estimator information exchange links

Fig. 1. A schematic diagram of distributed attack detection and secure estimation subject to a false data injection attack on physical system and jamming attacks on wireless measurement transmission channels

remote estimator i and estimator i's all underlying neighbors, thus being vulnerable to cyber attacks; and C_i, $\forall\ i \in \mathcal{V}$ is a known constant matrix with an appropriate dimension.

In the conventional distributed estimation framework [26]–[29], the measurement transmission channels from sensors j, $j \in \mathcal{N}_i$, to the remote estimator i are explicitly assumed to be ideal. In other words, the ideal measurements $y_j(k)$, $j \in \mathcal{N}_i$, are successfully and completely transmitted to estimator i at every instant of time, which leads to

$$F_i(k) = \sum_{j \in \mathcal{N}_i} a_{ij} y_j(k),\ \forall\ i \in \mathcal{V}, \tag{3}$$

where $F_i(k) \in \mathbb{R}^{n_y}$ is the combinational measurement which acts as an input of estimator i. However, this ideal assumption is not always true in practice when malicious cyber attacks occur in wireless transmission channels. For example, a DoS attacker aims at blocking the measurement transmission from sensor j to the remote estimator i, since typical DoS attacks can jam and interrupt the wireless channels. In this sense, the corresponding measurements $y_j(k)$ may be incomplete and lossy when they arrive at the side of estimator i.

E. Compensation-Based Measurement Output Model against Jamming Attacks

Consider a scenario where the attacker also launches multiple jamming attacks on the wireless measurement transmission channels between distributed sensors and remote estimators, as shown in Fig. 1, to deteriorate the overall estimation performance of the system (1). The attacker is an active adversary in the sense that sensor i's measurement will be dropped once the attacker successfully jams the corresponding wireless channel. Generally, there are three possible cases after the attacker launches a jamming attack on a wireless measurement transmission channel from sensor i to estimator j (i.e., $i \longrightarrow j$):

- Sensor i's measurement will successfully arrive at estimator j if the attacker fails to jam the transmission channel $i \longrightarrow j$. For example, in some circumstances, the attacker has to give up jamming certain channels due to a limited energy budget [16].
- Sensor i's measurement will be partially lost if the jamming of the transmission channel $i \longrightarrow j$ is not heavy.
- Sensor i's measurement will be completely lost if the transmission channel $i \longrightarrow j$ is severely jammed.

On the other hand, most malicious attackers have the energy constraint issue [16], which means that attackers may need to consider the energy budget when implementing various attack strategies. Understanding that a deterministic attack strategy not only leads to excess energy consumption but can also be readily handled by a robust estimator, a cunning attacker should randomly decide to jam the wireless transmission channels or to sleep in order to deceive designers or simply save energy. In this sense, in the presence of a smart attacker, random attack strategies may pose major difficulties for remote estimators.

Motivated by these facts, we adopt the following measurement output which is delivered through the wireless channel $i \longrightarrow j$ subject to random jamming attacks and adopted by the relevant estimators j, $\forall\ i \in \mathcal{N}_j; j \in \mathcal{V}$, as illustrated in Fig. 2,

$$\hat{y}_i(k) = y_i^{attk}(k) + y_i^{comp}(k) + y_i^{noise}(k),\ \forall\ i \in \mathcal{V}, \tag{4}$$

where the corrupted measurement $\hat{y}_i(k)$ on estimator j consists of three parts

$$\begin{cases} y_i^{attk}(k) = \theta_i(k) y_i(k) \\ y_i^{comp}(k) = (1 - \theta_i(k)) \hat{y}_i(k-1) \\ y_i^{noise}(k) = D_i v_i(k) \end{cases} \tag{5}$$

and specifically,

- $y_i^{attk}(k)$ stands for the attacked and manipulated measurement term. The stochastic variable $\theta_i(k) \in \mathbb{R}$ is a Bernoulli distributed white sequence taking values of 1 and 0 with the mathematical probability satisfying

$$\begin{cases} Prob\{\theta_i(k) = 1\} = \mathbb{E}\{\theta_i(k)\} = \beta_i \\ Prob\{\theta_i(k) = 0\} = 1 - \mathbb{E}\{\theta_i(k)\} = 1 - \beta_i, \end{cases} \tag{6}$$

where $\beta_i \in [0, 1]$ is a known constant. All stochastic variables $\theta_i(k)$, $\forall\ i \in \mathcal{V}$ and $k \in \mathbb{N}$ are assumed to be independent in i and k. Here, the stochastic variable $\theta_i(k)$ is employed to characterize the possibility of the measurement $y_i(k)$ being successfully transmitted to the remote estimator i (hereafter, β_i is known as the *measurement-transmission probability*). When the attacker launches a jamming attack and blocks the wireless channel $i \longrightarrow j$, sensor i's measurement $y_i(k)$ will be dropped with probability $1 - \beta_i$ (hereafter, $1 - \beta_i$ is known as the *measurement-loss probability*) [1]. Apparently, the larger

[1] Without causing confusion, we use the terms "measurement-transmission probability" and "measurement-loss probability" interchangeably throughout the paper, while the term "measurement-loss probability" emphasizes the lossy measurement caused by jamming attacks

the value of β_i, the higher the chance of successful transmission of the measurement $y_i(k)$.
Note that if $\beta_i \in (0, 1)$, only part of the measurement $y_i(k)$ is received by the estimators j, $\forall\ j \in \mathcal{N}_i$, and the channel $i \rightarrow j$ is partially jammed. In particular, $\beta_i \equiv 1$ corresponds to the ideal transmission case, which means that the measurement $y_i(k)$ is successfully and completely transmitted to estimators j and there is no jamming attack during the transmission, while $\beta_i \equiv 0$ reduces to the worse transmission case, where the measurement $y_i(k)$ is lost completely during the transmission.

- $y_i^{comp}(k)$ represents the compensated measurement term corresponding to the lossy measurement $y_i^{attk}(k)$ caused by the attacker. It is assumed that the sensor measurement $y_i(k)$ and its time-stamp k are encapsulated into a measurement packet $(k, y_i(k))$. As a result, whether or not this measurement packet is manipulated by the attacker can be checked by remote estimators according to the time-stamp of the arrived measurement packet. In this sense, it is reasonable to introduce the term $(1 - \theta_i(k))$ in the compensated measurement. On the other hand, it is shown in Fig. 2 that a buffer is equipped to store all the previous compensation-based measurements at instants $k = 0, 1, \cdots, k-1$, i.e., $(\hat{y}_i(k-1), \hat{y}_i(k-2), \cdots, \hat{y}_i(0))$. The buffer is accessed in a first-in-last-out mode, which means that remote estimator j, $\forall\ j \in \mathcal{V}$, can always use the compensated measurement $(1 - \theta_i(k))\hat{y}_i(k-1)$ on sensor i to reduce the effect of an attacked measurement $y_i^{attk}(k)$, $\forall\ i \in \mathcal{N}_j$. If at time step k, the measurement $y_i(k)$ is totally lost, the last transmitted measurement $\hat{y}_i(k-1)$ will be adopted to actuate estimators j. It is expected that such a compensation strategy will be helpful for estimators to generate accurate local estimations.
- $y_i^{noise}(k)$ denotes the perturbed measurement term with $v_i(k) \in \mathbb{R}^{n_v}$ being regarded as the measurement noise experienced through the wireless channel $i \rightarrow j$. Here, $v_i(k)$ is assumed to belong to $l_2[0, \infty)$ and D_i, $\forall\ i \in \mathcal{V}$ is a known constant matrix with an appropriate dimension.

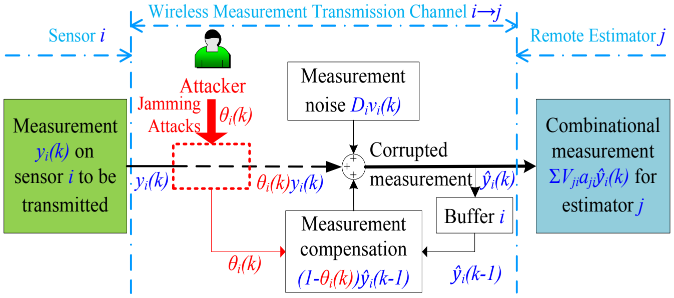

Wireless Measurement Transmission Channel $i \rightarrow j$
Sensor i
Remote Estimator j
Attacker
Jamming Attacks
$\theta_i(k)$
Measurement noise $D_i v_i(k)$
Measurement $y_i(k)$ on sensor i to be transmitted
$y_i(k)$
$\theta_i(k) y_i(k)$
Corrupted measurement
$\hat{y}_i(k)$
Combinational measurement $\Sigma V_{ji} a_{ji} \hat{y}_i(k)$ for estimator j
Buffer i
Measurement compensation $(1-\theta_i(k))\hat{y}_i(k-1)$
$\theta_i(k)$
$\hat{y}_i(k-1)$

Fig. 2. A compensation-based measurement output model against jamming attacks on wireless channels

F. Resilient Attack Detection Estimators

We are interested in constructing the following resilient attack detection estimators of the form

$$x_i(k+1) = \sum_{j \in \mathcal{N}_i} U_{ij} a_{ij} x_j(k) + \tilde{F}_i(k) \tag{7}$$

$$\tilde{F}_i(k) = \sum_{j \in \mathcal{N}_i} V_{ij} a_{ij} (\hat{y}_j(k) - \beta_j C_j x_j(k)) \tag{8}$$

$$r_i(k) = W_i x_i(k), \tag{9}$$

where $x_i(k) \in \mathbb{R}^{n_s}$ is the local state estimation computed by estimator i; $\sum_{j \in \mathcal{N}_i} U_{ij} a_{ij} x_j(k)$ is the estimation exchange term which represents how estimator i collects the estimations $x_j(k)$ from its neighboring estimators $j, \forall\ j \in \mathcal{N}_i$; $\tilde{F}_i(k)$ is the combinational measurement under the jamming attacks during the transmission with $\hat{y}_j(k)$ being defined in (4); $r_i(k) \in \mathbb{R}^{n_r}$ is the residual signal, which is assumed to be compatible with the FDI attack vector $p(k)$; and the initial condition of estimator i is $x_i(0) = x_i^0$. For all $i, j \in \mathcal{V}$, U_{ij}, V_{ij} and W_i are the estimator gain matrices to be determined.

As outlined in the preceding section, two types of malicious attacks are considered in the proposed distributed attack detection and secure estimation framework. By launching the FDI attack, the adversary injects false information to modify the system's state, whereas with jamming attacks, the attacker tries to block or interrupt the wireless measurement transmission channels between distributed sensors and remote estimators. Hence, we aim at designing resilient attack detection estimators that can work properly under both types of attacks. More specifically,

- To deal with the FDI attack, a two-step FDI attack detection mechanism will be established, as demonstrated in Fig. 3. The first step is to generate a residual signal $r_i(k)$ on each estimator. To achieve a desirable detection of the FDI attack, the second step is to evaluate the generated residual signal by analyzing the information about the FDI attack signal from the residual by means of post-processing of the residual. To this end, we define an evaluation function in terms of the norm of the residual signal on each estimator of the following form

$$f_i(T_0, T_e) = \left\{ \sum_{k=T_0}^{k=T_0+T_e} r_i^T(k) r_i(k) \right\}^{\frac{1}{2}}, \tag{10}$$

where $T_0 \geq 0$ denotes the initial evaluation time instant and $T_e > T_0$ denotes the evaluation time steps. Moreover, to discern when the occurrence of the FDI attack can be detected, a specific threshold should be pre-defined such that when the FDI attack occurs, a warning or an alarm message can be sent to remind the designer or operating engineer. Generally, the threshold should be the maximal value of the evaluated residual in the FDI attack-free case. For this purpose, we choose the residual evaluation threshold as

$$Th_i = \sup_{\substack{w(k), v_i(k) \in l_2[0,\infty) \\ p(k) \equiv 0}} \mathbb{E}\{\|r_i(k)\|\}. \tag{11}$$

Based on (10) and (11), the following evaluation logic is designed by comparing the evaluation function $f_i(T_0, T_e)$ with the threshold Th_i:

$$\begin{cases} f_i(T_0, T_e) > Th_i \Rightarrow \text{Alarm of FDI attack} \\ f_i(T_0, T_e) \leq Th_i \Rightarrow \text{No FDI attack.} \end{cases} \tag{12}$$

By virtue of the FDI attack alarm, further measures, such as attack isolation and false data correction, can be taken to guarantee the reliability and safety of the system, which serves as a possible direction of our future work. This paper mainly focuses on detecting the occurrence of the FDI attack.

- To handle the random jamming attacks, estimator i in (7) adopts the compensated measurements $\hat{y}_j(k)$, $\forall\ j \in \mathcal{N}_i$. This is critical because the jamming attacks are randomly launched to wireless channels. Some sensors' measurements can be lost completely during the transmission at some instants of time. Thus, by employing the compensated measurements, the resilience of the estimation system is expected to increase. It should be also noted that the proposed estimator i in (7) is based on its local estimation and received measurement as well as its all of its neighbors' estimations and measurements. This distributed cooperative estimation paradigm also aims to improve the resilience and reliability of the estimation system.

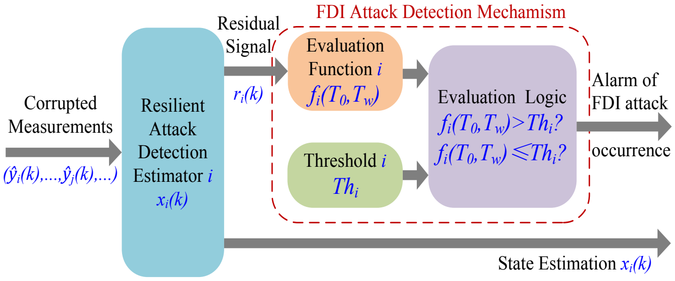


Fig. 3. Resilient attack detection estimators using a false data injection (FDI) attack detection mechanism

Remark 1: Note that the basic idea of the proposed FDI attack detection mechanism is to construct a residual signal and based on this, to determine a residual evaluation function to compare with a predefined threshold. If the residual evaluation function has a value larger than the threshold, then an FDI attack alarm is sent. However, it should be noted that the co-existence of some uncertain factors, such as process and measurement noises, and randomly dropped measurements caused by jamming attacks in this paper, may affect the residual and, furthermore, the evaluation function and threshold. Thus, it is possible that the jamming attack actions may falsely result in an FDI attack alarm at a specific instant of time. To accurately address the "false alarm" issue caused by jamming attacks, an intuition is to extract the information about the FDI attack from the residuals. However, this extraction/separation is generally difficult because 1) the jamming attacks are launched randomly; 2) the FDI attack and the jamming attacks are launched concurrently; and 3) the corrupted sensor measurements are disseminated in an epidemic manner over the WSN. In the proposed FDI attack detection mechanism, we consider the worst case scenario of all the possible effects of the process noise, the measurement noise and the randomly dropped measurements by using the "supremum" to select a proper threshold. To our knowledge, it remains open to establish a rigorous theoretical framework to investigate the "false alarm" issue of FDI attack in the simultaneous presence of jamming attacks on wireless measurement transmission channels.

G. The Distributed Attack Detection and Secure Estimation Problem

For node i, $\forall\ i \in \mathcal{V}$, define a state estimation error vector $e_i(k) = s(k) - x_i(k)$ and a residual error vector $h_i(k) = r_i(k) - p(k)$. Denote $\bar{e}(k) = [e_1^T(k),\ e_2^T(k),\ \cdots,\ e_N^T(k)]^T$, $\bar{s}(k) = [s^T(k), s^T(k), \cdots, s^T(k)]^T$, $\hat{y}(k) = [\hat{y}_1^T(k),\ \hat{y}_2^T(k),\ \cdots,\ \hat{y}_N^T(k)]^T$, $\bar{v}(k) = [v_1^T(k),\ v_2^T(k),\ \cdots,\ v_N^T(k)]^T$, $\bar{h}(k) = [h_1^T(k), h_2^T(k), \cdots, h_N^T(k)]^T$, $\bar{A} = I_N \otimes A$, $\bar{B} = [B^T,\ B^T,\ \cdots,\ B^T]^T$, $\bar{C} = diag\{\beta_1 C_1,\ \beta_2 C_2,\ \cdots,\ \beta_N C_N\}$, $\bar{D} = diag\{D_1,\ D_2,\ \cdots,\ D_N\}$, $\bar{E} = [E^T,\ E^T,\ \cdots,\ E^T]^T$, $\bar{\beta} = diag\{\beta_1,\ \beta_2,\ \cdots,\ \beta_N\}$, $\hat{I} = [I_{n_p}^T,\ I_{n_p}^T,\ \cdots,\ I_{n_p}^T]^T$, $\bar{W} = diag\{W_1,\ W_2,\ \cdots,\ W_N\}$, $\tilde{C}_i = diag\{\underbrace{0,\cdots,0}_{i-1},\ C_i,\ \underbrace{0,\cdots,0}_{N-i}\}$ and $\tilde{I}_i = diag\{\underbrace{0,\cdots,0}_{i-1},\ I_{n_y},\ \underbrace{0,\cdots,0}_{N-i}\}$ for all $i \in \mathcal{V}$.

To simplify subsequent development, we further set

$$\begin{cases} \bar{U} = [\bar{u}_{ij}]_{N\times N} & \text{with} \quad \bar{u}_{ij} = U_{ij} a_{ij} \\ \bar{V} = [\bar{v}_{ij}]_{N\times N} & \text{with} \quad \bar{v}_{ij} = V_{ij} a_{ij}. \end{cases} \tag{13}$$

It is easy to verify that $\bar{U}$ and $\bar{V}$ are two sparse matrices due to the fact that $a_{ij} = 0$ if $j \notin \mathcal{N}_i$. For sparse matrices, we recall the following lemma, which is helpful in deriving our subsequent results.

Lemma 1: [30] Let $\mathbb{S} = \{\bar{S} = [S_{ij}]_{Nn_s \times Nn_y} | S_{ij} \in \mathbb{R}^{n_s \times n_y},\ S_{ij} = 0 \text{ if } j \notin \mathcal{N}_i\}$ be the set of sparse matrices and $P = diag\{P_1,\ P_2,\ \cdots,\ P_N\}$ with $P_i \in \mathbb{R}^{n_s \times n_s}$, $\forall\ i \in \mathcal{V}$, being invertible matrices. For any matrix $\mathcal{F} \in \mathbb{R}^{Nn_s \times Nn_y}$, if $\mathit{F} = P\mathcal{F}$, then we have

$$\mathcal{F} \in \mathbb{S} \Longleftrightarrow \mathit{F} \in \mathbb{S}.$$

Substituting (4) into (8) and combining (1), (7)-(9), the estimation error system can be rewritten in a compact form as follows

$$\begin{aligned} \bar{e}(k+1) =& (\bar{A} - \bar{U})\bar{s}(k) + (\bar{U} - \bar{V}\bar{C})\bar{e}(k) + (\bar{V}\bar{\beta} - \bar{V})\hat{y}(k-1) \\ & + \sum_{i=1}^{N} (\theta_i(k) - \beta_i) \left(-\bar{V}\tilde{C}_i \bar{s}(k) + \bar{V}\tilde{I}_i \hat{y}(k-1)\right) \\ & + \bar{B}w(k) - \bar{V}\bar{D}\bar{v}(k) + \bar{E}p(k) \end{aligned} \tag{14}$$

$$\bar{h}(k) = \bar{W}\bar{s}(k) - \bar{W}\bar{e}(k) - \hat{I}p(k). \tag{15}$$

Setting $\xi(k) = [\bar{s}^T(k),\ \bar{e}^T(k),\ \hat{y}^T(k-1)]^T$ and $\eta(k) = [w^T(k),\ \bar{v}^T(k),\ p^T(k)]^T$, the combination of (1), (4), (14) and (15) yields the following augmented estimation error system

$$\xi(k+1) = \mathscr{A}\xi(k) + \sum_{i=1}^{N} (\theta_i(k) - \beta_i)\,\mathscr{B}_i \xi(k) + \mathscr{E}\eta(k) \tag{16}$$

$$\bar{h}(k) = \mathscr{C}\xi(k) + \mathscr{D}\eta(k), \tag{17}$$

where $\mathscr{A}$, $\mathscr{B}_i$, $\forall\ i \in \mathcal{V}$, $\mathscr{E}$, $\mathscr{C}$ and $\mathscr{D}$ are given in Box I.

As can be seen in (16) and (17), the estimation error system reveals the difference between the residual signal $r_i(k)$ and the FDI attack $p(k)$ to be detected. Moreover, the effects

$$\mathscr{A}=\left[\begin{array}{ccc}\bar{A} & 0 & 0\\ \bar{A}-\bar{U} & \bar{U}-\bar{V}\bar{C} & \bar{V}\bar{\beta}-\bar{V}\\ \bar{C} & 0 & I_N-\bar{\beta}\end{array}\right],\ \mathscr{B}_i=\left[\begin{array}{ccc}0 & 0 & 0\\ -\bar{V}\tilde{C}_i & 0 & \bar{V}\tilde{I}_i\\ \tilde{C}_i & 0 & -\tilde{I}_i\end{array}\right],\ \mathscr{E}=\left[\begin{array}{ccc}\bar{B} & 0 & \bar{E}\\ \bar{B} & -\bar{V}\bar{D} & \bar{E}\\ 0 & \bar{D} & 0\end{array}\right],\mathscr{C}=\left[\begin{array}{c}\bar{W}^T\\ -\bar{W}^T\\ 0\end{array}\right],\mathscr{D}=\left[\begin{array}{c}0\\ 0\\ -\hat{I}^T\end{array}\right].$$

Box I.

$$\Upsilon=\left[\begin{array}{ccc}\bar{A}^TP_1 & \bar{A}^TP_2-\tilde{U}^T & \bar{C}^TP_3\\ 0 & \tilde{U}-\bar{C}^T\tilde{V}^T & 0\\ 0 & (\bar{\beta}-I_N)\tilde{V}^T & (I_N-\bar{\beta})P_3\end{array}\right],\ \check{\Upsilon}=\left[\begin{array}{ccc}\bar{B}^TP_1 & \bar{B}^TP_2 & 0\\ 0 & -\bar{D}^T\tilde{V}^T & \bar{D}^TP_3\\ \bar{E}^TP_1 & \bar{E}^TP_2 & 0\end{array}\right],\ \Omega_i=\left[\begin{array}{ccc}0 & -\tilde{C}_i^T\tilde{V}^T & \tilde{C}_i^TP_3\\ 0 & 0 & 0\\ 0 & \tilde{I}_i\tilde{V}^T & -\tilde{I}_iP_3\end{array}\right].$$

Box II.

of the attacks, the process noise $w(k)$ and the measurement noise $v_i(k)$ through wireless channels can be minimized by making the H_∞ norm of the difference small. In this sense, the distributed attack detection and secure estimation problem to be tackled can be cast into an auxiliary H_∞ estimation problem, while the latter can be solved by employing the celebrated H_∞ optimization technique.

To proceed with, the following definition with regard to stochastic stability is recalled such that the main problem of this paper can be described more precisely.

Definition 1: System (16) with $w(k) \equiv 0$ and $v_i(k) \equiv 0$ is said to be stochastically stable if the following holds

$$\mathbb{E}\left\{\sum_{k=0}^{\infty}\|\xi(k)\|^2\right\}<\infty$$

for any initial condition s_0 and x_i^0.

Based on the above definition, the objective of this paper is to design desired resilient attack detection estimators of the form (7)-(9) such that

- The augmented estimation error system (16) and (17) with $w(k) \equiv 0$ and $v_i(k) \equiv 0$ is stochastically stable for any initial condition.
- For all nonzero $w(k), v_i(k) \in l_2[0,\infty)$, $\forall\ i \in \mathcal{V}$, the augmented estimation error system (16) and (17) satisfies the following performance constraint

$$\sup_{\substack{w(k)\neq 0,\, w(k)\in l_2[0,\infty)\\ v_i(k)\neq 0,\, v_i(k)\in l_2[0,\infty)}} \frac{\mathbb{E}\left\{\|\bar{h}(k)\|\right\}}{\|\eta(k)\|}<\gamma \tag{18}$$

for the zero initial condition, where the infimum of $\gamma > 0$ is made small in the feasibility of (18).

III. Main Results

In this section, criteria for analyzing estimation performance and designing resilient attack detection estimators (7)-(9) will be derived such that the augmented estimation error system (16) and (17) is stochastically stable under an optimized H_∞ performance level.

A. Performance Analysis on Distributed Attack Detection and Secure Estimation

We first present the following theorem which states under what conditions the augmented estimation error system (16) and (17) is stochastically stable with a prescribed H_∞ performance index.

Theorem 1: For prescribed scalars $\gamma > 0$, $\beta_i \in [0,1]$ for all $i \in \mathcal{V}$, and given estimator gain matrices U_{ij}, V_{ij} and W_i, $\forall\ i,j \in \mathcal{V}$, the augmented estimation error system (16) and (17) is stochastically stable and achieves a prescribed H_∞ performance level γ if there exists a real matrix $P > 0$ of an appropriate dimension such that

$$\Phi<0, \tag{19}$$

where $\Phi = [\Phi^{(mn)}]_{5\times 5}$ is a sparse block diagonal matrix with each nonzero entry given by $\Phi^{(11)} = -P$, $\Phi^{(22)} = -\gamma^2 I_{n_w+Nn_v+n_p}$, $\Phi^{(13)} = \mathscr{A}^T P$, $\Phi^{(23)} = \mathscr{E}^T P$, $\Phi^{(33)} = -P$, $\Phi^{(14)} = \mathscr{C}^T$, $\Phi^{(24)} = \mathscr{D}^T$, $\Phi^{(44)} = -I_{Nn_p}$, $\Phi^{(15)} = [\alpha_1\mathscr{B}_1^T P,\ \alpha_2\mathscr{B}_2^T P,\ \cdots,\ \alpha_N\mathscr{B}_N^T P]$ and $\Phi^{(55)} = -I_N \otimes P$ with $\alpha_i = \beta_i(1-\beta_i)$, $\forall\ i \in \mathcal{V}$.

Proof: Construct the following stochastic Lyapunov functional candidate $V(\xi(k)) = \xi^T(k)P\xi(k)$. Recall the facts of $\mathbb{E}\{(\theta_i(k) - \beta_i)\} = 0$, $\mathbb{E}\{(\theta_i(k) - \beta_i)^2\} = \alpha_i$ and $\mathbb{E}\{(\theta_i(k) - \beta_i)(\theta_j(k) - \beta_j)\} = 0$ for any $i \neq j$. For all nonzero $w(k), v_i(k) \in l_2[0,\infty)$, $\forall\ i \in \mathcal{V}$, calculating the forward difference of $V(\xi(k))$ along the system (16) yields

$$\begin{aligned}
&\mathbb{E}\{\mathbb{E}\{V(\xi(k+1))\}-V(\xi(k))\\
&\qquad+\bar{h}^T(k)\bar{h}(k)-\gamma^2\eta^T(k)\eta(k)\}\\
&\quad=\mathbb{E}\left\{\mathbb{E}\{\xi^T(k+1)P\xi(k+1)\}-\xi^T(k)P\xi(k)\right. && (20)\\
&\qquad\left.+\bar{h}^T(k)\bar{h}(k)-\gamma^2\eta^T(k)\eta(k)\right\}\\
&\quad=\phi^T(k)\tilde{\Phi}\phi(k), && (21)
\end{aligned}$$

where

$$\phi(k)=\left[\begin{array}{c}\xi(k)\\ \eta(k)\end{array}\right],\tilde{\Phi}=\left[\begin{array}{cc}\tilde{\Phi}^{(11)} & \tilde{\Phi}^{(12)}\\ * & \tilde{\Phi}^{(22)}\end{array}\right]$$

with $\tilde{\Phi}^{(11)} = \mathscr{A}^T P\mathscr{A} - P + \mathscr{C}^T\mathscr{C} + \sum_{i=1}^{N}\alpha_i\mathscr{B}_i^T P\mathscr{B}_i$, $\tilde{\Phi}^{(12)} = \mathscr{A}^T P\mathscr{E} + \mathscr{C}^T\mathscr{D}$ and $\tilde{\Phi}^{(22)} = \mathscr{E}^T P\mathscr{E} + \mathscr{D}^T\mathscr{D} - \gamma^2 I$.

Applying the Schur complement [31] to (19), it is straightforward to derive $\tilde{\Phi} < 0$. Thus, one has

$$\begin{aligned}
&\mathbb{E}\{\bar{h}^T(k)\bar{h}(k)-\gamma^2\eta^T(k)\eta(k)\}\\
&\qquad<\mathbb{E}\{V(\xi(k))-\mathbb{E}\{V(\xi(k+1))\}\}. && (22)
\end{aligned}$$

Summing up (22) from $k = 0$ to $k = k_T$, where $k_T \to \infty$, under the zero initial condition that $\mathbb{E}\{V(\xi(0))\} = 0$ and $\mathbb{E}\{V(\xi(k))\} \geq 0$, we finally obtain

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2017.2749959, IEEE Transactions on Signal and Information Processing over Networks

$\sum_{k=0}^{\infty}\mathbb{E}\left\{\|\bar{h}(k)\|^2-\gamma^2\|\eta(k)\|^2\right\}<0$, which means that the performance index (18) holds.

Next, we consider zero noise signals, i.e., $w(k)\equiv 0$ and $v_i(k)\equiv 0$, $\forall\ i\in\mathcal{V}$. Calculating the forward difference of $V(\xi(k))$ along the system (16) yields

$$\begin{aligned}&\mathbb{E}\{\mathbb{E}\{V(\xi(k+1))\}-V(\xi(k))\}\\&\quad=\xi^T(k)\left(\mathscr{A}^TP\mathscr{A}-P+\sum_{i=1}^{N}\alpha_i\mathscr{B}_i^TP\mathscr{B}_i\right)\xi(k),\quad(23)\end{aligned}$$

Similarly, applying the Schur complement to (19), it is straightforward to have $\mathscr{A}^TP\mathscr{A}-P+\sum_{i=1}^{N}\alpha_i\mathscr{B}_i^TP\mathscr{B}_i<0$. Hence, one has $\mathbb{E}\{\mathbb{E}\{V(\xi(k+1))\}-V(\xi(k))\}<0$. Then, following a similar pattern of the proof of Theorem 1 in [32], it can be shown that $\mathbb{E}\left\{\sum\limits_{k=0}^{\infty}\|\xi(k)\|^2\right\}<\infty$. By Definition 1, it can be concluded that the augmented estimation error system (16) is stochastically stable. This completes the proof. ■

B. Design of Resilient Attack Detection Estimators

Next, let us focus our attention on designing the desired resilient attack detection estimators (7)-(9) and solving out the estimator gain matrices U_{ij}, V_{ij} and W_i, $\forall\ i,j\in\mathcal{V}$.

The main result is stated in the following theorem which provides a design criterion for solving the proposed distributed attack detection and secure estimation problem.

Theorem 2: Given scalars $\gamma>0$ and $\beta_i\in[0,1]$ for all $i\in\mathcal{V}$, the proposed distributed attack detection and secure estimation problem for the augmented estimation error system (16) and (17) is solvable if there exist real matrices $P_1>0$, $P_2=diag\{P_{2,1},\ P_{2,2},\ \cdots,\ P_{2,N}\}>0$, $P_3>0$, $\check{U}$, $\check{V}$, $\bar{W}$ of an appropriate dimension such that

$$\check{\Phi}<0,\tag{24}$$

where $\check{\Phi}=[\check{\Phi}^{(mn)}]_{5\times 5}$ is a sparse block diagonal matrix with each nonzero entry given by $\check{\Phi}^{(11)}=-P=diag\{P_1,P_2,P_3\}$, $\check{\Phi}^{(22)}=-\gamma^2I_{n_w+Nn_v+n_p}$, $\check{\Phi}^{(13)}=\Upsilon$, $\check{\Phi}^{(23)}=\check{\Upsilon}$, $\check{\Phi}^{(33)}=-P$, $\check{\Phi}^{(14)}=\mathscr{C}^T$, $\check{\Phi}^{(24)}=\mathscr{D}^T$, $\check{\Phi}^{(44)}=-I_{Nn_p}$, $\check{\Phi}^{(15)}=[\alpha_1\Omega_1,\ \alpha_2\Omega_2,\ \cdots,\ \alpha_N\Omega_N]$ and $\check{\Phi}^{(55)}=-I_N\otimes P$ with Υ, $\check{\Upsilon}$ and $\Omega_i, \forall\ i\in\mathcal{V}$ being given in Box II. Moreover, the estimator gain matrices $\bar{U}$ and $\bar{V}$ can be computed by

$$\bar{U}=P_2^{-1}\tilde{U},\ \bar{V}=P_2^{-1}\tilde{V}\tag{25}$$

and $\bar{W}$ can be directly solved out from (24).

Proof: Choose the following diagonal structure of the matrix $P=diag\{P_1,P_2,P_3\}$, and define two new matrices $\tilde{U}=P_2\bar{U}$ and $\tilde{V}=P_2\bar{V}$. By Lemma 1, it can be shown that $\tilde{U}$ and $\tilde{V}$ are sparse matrices. Then, (19) implies that (24). This completes the proof. ■

With Theorem 2, the proposed distributed attack detection and secure estimation problem can be transformed into the following optimization problem

$$\underset{\digamma}{\text{minimize}}\ \ (\lambda)\quad\text{subject to (24)},$$

where $\lambda=\gamma^2$ and $\digamma$ is the set of all feasible solutions from the linear matrix inequality in Theorem 2. By using the available interior-point algorithms in many available commercial and noncommercial software products such as the Matlab LMI toolbox, one can solve the above minimization problem to obtain the desired resilient attack detection estimators (7)-(9) such that the stochastic stability and the optimal H_∞ performance level $\gamma=\sqrt{\lambda}$ of the augmented estimation error system (16) and (17) under the FDI attack and jamming attacks can be guaranteed.

C. Extension to the Case of Uncertain Measurement-Transmission Probability

Consider the case that the measurement-transmission probability β_i is uncertain. More specifically, it is assumed that the measurement-transmission probability β_i is subject to uncertainties of the polytopic type.

Assumption 1: The measurement-transmission probability β_i in (6) is uncertain and belongs to a given convex ployhedral domain described by S vertices:

$$\beta_i^\sigma\in\left\{\beta_i^\sigma|\beta_i^\sigma=\sum_{s=1}^{S}\sigma_s\beta_i^{(s)};\sum_{s=1}^{S}\sigma_s=1;\sigma_s\geq 0\right\},\tag{26}$$

where $\beta_i^{(s)}$ denotes the s-th vertex of the polotope for all $i\in\mathcal{V}$.

Remark 2: Note that the uncertain but bounded measurement-transmission probability β_i^σ satisfying

$$0\leq\beta_i^{(1)}\leq\beta_i^\sigma\leq\beta_i^{(2)}\leq 1,$$

where $\beta_i^{(1)}$ and $\beta_i^{(2)}$ are known real constants, is a special case of the polytopic-type uncertainty with only two vertices, i.e., $\beta_i^\sigma=\sum_{s=1}^{2}\sigma_s\beta_i^{(s)}$.

Remark 3: From the perspective of the attacker, it is promising to account for the uncertain measurement-transmission probability β_i^σ (or uncertain measurement-loss probability $1-\beta_i^\sigma$) because the probability of when and where the attacker decides to launch the jamming attacks is indeterminate to remote estimators or detectors. This will make the wireless channels highly vulnerable and will pose significant challenges for designing estimators or detectors.

First, we present a parameter-dependent design criterion that guarantees the feasibility of the proposed distributed attack detection and secure estimation problem.

Theorem 3: Given a positive scalar γ, the proposed distributed attack detection and secure estimation problem for the augmented estimation error system (16) and (17) is solvable if there exist real matrices $P_1^\sigma>0$, $P_2=diag\{P_{2,1},\ P_{2,2},\ \cdots,\ P_{2,N}\}>0$, $P_3^\sigma>0$, $\check{U}$, $\check{V}$, $\bar{W}$ of an appropriate dimension such that

$$\Psi_\sigma<0,\tag{27}$$

where $\Psi_\sigma=[\Psi_\sigma^{(mn)}]_{5\times 5}$ is a sparse block diagonal matrix with each nonzero entry given by $\Psi_\sigma^{(11)}=-P^\sigma=diag\{P_1^\sigma,P_2,P_3^\sigma\}$, $\Psi_\sigma^{(22)}=-\gamma^2I_{n_w+Nn_v+n_p}$, $\Psi_\sigma^{(13)}=\bar{\Upsilon}$, $\Psi_\sigma^{(23)}=\tilde{\Upsilon}$, $\Psi_\sigma^{(33)}=-P^\sigma$, $\Psi_\sigma^{(14)}=\mathscr{C}^T$, $\Psi_\sigma^{(24)}=\mathscr{D}^T$, $\Psi_\sigma^{(44)}=-I_{Nn_p}$, $\Psi_\sigma^{(15)}=[\bar{\Omega}_1,\ \bar{\Omega}_2,\ \cdots,\ \bar{\Omega}_N]$ and $\Psi_\sigma^{(55)}=-I_N\otimes 4P$ with $\bar{\Upsilon}$, $\tilde{\Upsilon}$ and $\bar{\Omega}_i$ being derived from Υ, $\check{\Upsilon}$ and Ω_i in Box II by replacing P_1,P_3,β_i with $P_1^\sigma,P_3^\sigma,\beta_i^\sigma$, respectively. The estimator gain matrices $\bar U$ and $\bar V$ can be computed by (25), and $\bar W$ can be directly solved out from (27).

Proof: Consider the following parameter-dependent stochastic Lyapunov functional candidate $\bar V(\xi(k)) = \xi^T(k)P^\sigma\xi(k)$. Calculating the forward difference of $V(\xi(k))$ along the system (16) yields

$$\begin{aligned}&\mathbb{E}\{\mathbb{E}\{\bar V(\xi(k+1))\} - \bar V(\xi(k))\\&\qquad + \bar h^T(k)\bar h(k) - \gamma^2\eta^T(k)\eta(k)\}\\&\le \phi^T(k)\begin{bmatrix}\tilde\Psi_\sigma^{(11)} & \tilde\Psi_\sigma^{(12)}\\ * & \tilde\Psi_\sigma^{(22)}\end{bmatrix}\phi(k),\end{aligned} \tag{28}$$

where $\tilde\Psi_\sigma^{(11)} = \mathscr{A}_\sigma^T P^\sigma \mathscr{A}_\sigma - P^\sigma + \mathscr{C}^T\mathscr{C} + \sum_{i=1}^N \frac{1}{4}\mathscr{B}_i^T P^\sigma \mathscr{B}_i$, $\tilde\Psi_\sigma^{(12)} = \mathscr{A}_\sigma^T P^\sigma \mathscr{E} + \mathscr{C}^T\mathscr{D}$ and $\tilde\Psi_\sigma^{(22)} = \mathscr{E}^T P^\sigma \mathscr{E} + \mathscr{D}^T\mathscr{D} - \gamma^2 I$. It should be noted that the inequality $\alpha_i = \beta_i(1-\beta_i) \le \frac{1}{4}$ is used to obtain (28). The rest of the proof is similar to the counterpart in the proof for Theorem 1. ■

A closer inspection of (27) reveals that the feasibility of Theorem 3 is dependent on the uncertain parameter σ, which means that Theorem 3 cannot be applied directly to solve out the estimator gain matrices. Therefore, one needs to convert the condition in Theorem 3 into a finite set of linear matrix inequality constraints. To achieve this goal, an alternative method is to set parameter-dependent matrices such as P_1^σ and P_3^σ to be linearly dependent on the uncertain parameter σ. For example, since β_i^σ takes on a polytopic form, one may set $P_1^\sigma = \sum_{s=1}^S \sigma_s P_1^{(s)}$ and $P_3^\sigma = \sum_{s=1}^S \sigma_s P_3^{(s)}$, where $P_1^{(s)}$ and $P_3^{(s)}$, $s = 1,2,\cdots,S$ are constant real matrices to be determined.

Next, we present the following result which provides a numerically tractable design criterion for solving the proposed distributed attack detection and secure estimation problem in the case of uncertain measurement-transmission probability. Recalling (28), the proof follows the similar pattern of Theorem 3, is thus omitted.

Theorem 4: Given scalars $\gamma > 0$ and $\beta_i^{(s)} \in [0,1]$ for all $i \in \mathcal{V}; s = 1,2,\cdots,S$, the proposed distributed attack detection and secure estimation problem for the augmented estimation error system (16) and (17) is solvable if there exist real matrices $P_1^{(s)} > 0$, $P_2 = diag\{P_{2,1}, P_{2,2}, \cdots, P_{2,N}\} > 0$, $P_3^{(s)} > 0$, $\check U$, $\check V$, $\bar W$ of an appropriate dimension such that

$$\check\Psi_s < 0, \quad s = 1,2,\cdots,S \tag{29}$$

where $\check\Psi_s = [\check\Psi_s^{(mn)}]_{5\times5}$ is a sparse block diagonal matrix derived from $\check\Phi$ in (24) by replacing $P_1, P_3, \beta_i, \bar C$ in $\check\Phi$ with $P_1^{(s)}, P_3^{(s)}, \beta_i^{(s)}, \bar C^{(s)}$, respectively. Moreover, the estimator gain matrices $\bar U$ and $\bar V$ can be computed by (25), and $\bar W$ can be directly solved out from (29).

IV. An Illustrative Example

To demonstrate the effectiveness and applicability of the proposed attack detection and secure estimation method, consider the system model as an industrial continuous-stirred tank reactor (CSTR), in which chemical species A reacts to form species B [33], as shown in Fig. 4. C_{A0} is the low concentration; C_A is the output concentration of the educt A; C_B is the output concentration of the desired product B within the reactor; T denotes the reactor temperature; and T_c is the cooling medium temperature. The discretized and linearized state-space model of the CSTR near the operating point is borrowed from [23], [24] and is given by

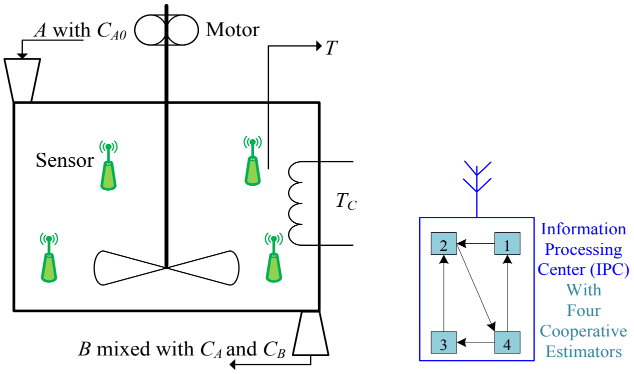

A with C_{A0}
Motor
T
Sensor
T_C
B mixed with C_A and C_B
2
1
3
4
Information Processing Center (IPC) With Four Cooperative Estimators

Fig. 4. A physical structure of a networked continuous stirred tank reactor (CSTR)

$$\begin{bmatrix} s^{(1)}(k+1)\\ s^{(2)}(k+1)\end{bmatrix} = \begin{bmatrix} 0.9719 & -0.0013\\ -0.0340 & 0.8628\end{bmatrix}\begin{bmatrix} s^{(1)}(k)\\ s^{(2)}(k)\end{bmatrix} + \begin{bmatrix} 0.3\\ 0.1\end{bmatrix} w(k) + \begin{bmatrix} -0.0839\\ 0.0761\end{bmatrix} p(k), \tag{30}$$

where $s^{(1)}(k)$ denotes the output concentration of the educt A; $s^{(2)}(k)$ represents the reactor temperature; $w(k)$ stands for the process noise, which may stem from poisoning of the reaction and/or from fouling of the cooling coils; and $p(k)$ is regarded as a potential false data injection attack launched by an adversary at the physical process side.

In this example, we apply the developed distributed attack detection and secure estimation method to estimate the state of the CSTR by using only the measurement of the reactor temperature. Four distributed sensors, i.e., $\mathcal{V} = \{1,2,3,4\}$, are deployed to monitor and measure the reactor temperature. Each sensor's measurement is then sent through a wireless channel to a remote information processing center for computing local estimations. The interaction topology of four cooperative estimators is depicted in Fig. 4, where the adjacency matrix of the topology is selected as a binary matrix, whose element is either 1 or 0. The measurement model is subject to measurement noise $v_i(k)$ and has the form of (3) with parameter matrices given by $C_i = [0 \;\; 0.1 + 1/i]$ and $D_i = 0.1/i$ for any $i \in \mathcal{V}$.

The objective of this case study is twofold: 1) each estimator detects when the FDI attack occurs and generates an alarm signal after its occurrence; and 2) each estimator computes local estimations of the output concentration of the educt A and the reactor temperature so as to attenuate the effects of random packet losses caused by the jamming attacks through measurement transmission channels and the process noise and measurement noise.

The uncertain measurement transmission probability of each sensor is assumed to be $0.3 \le \beta_1^{(\sigma)} \le 0.7$, $0.2 \le \beta_2^{(\sigma)} \le 0.6$, $0.4 \le \beta_3^{(\sigma)} \le 0.8$ and $0.5 \le \beta_4^{(\sigma)} \le 0.8$. The random variables $\theta_i(k)$, $i \in \mathcal{V}$ are shown in Fig. 5. By applying Theorem 3, it is

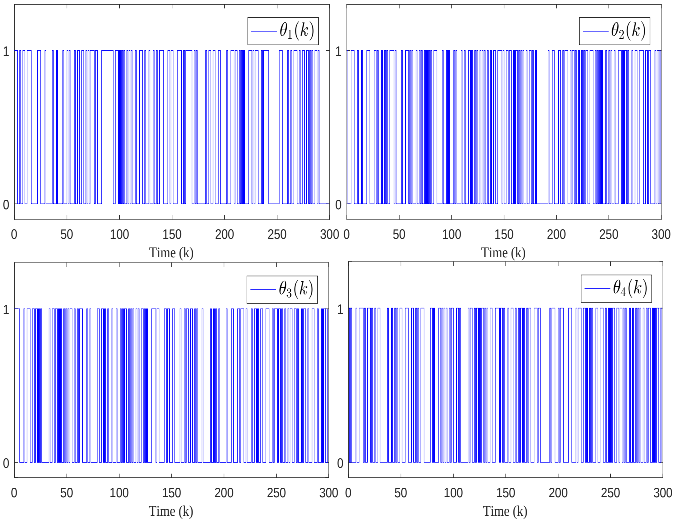


Fig. 5. The random variables $\theta_i(k)$, $i \in \mathcal{V}$

found that the proposed distributed secure estimation problem is solvable. Moreover, the optimal H_∞ noise attenuation level is obtained as $\gamma = 2$. To further illustrate the effectiveness of the designed resilient estimators, the process noise and measurement noise are taken as

$$w(k) = \begin{cases} rand - 0.6, & 30s \le k \le 80s \\ 0, & \text{otherwise} \end{cases}$$

$$v_i(k) = \begin{cases} 1.2rand - rand, & 30s \le k \le 80s \\ 0, & \text{otherwise} \end{cases}, \forall\ i \in \mathcal{V},$$

where $rand$ denotes a random scalar evenly distributively generated within $[0, 1]$. The false data injection attack signal $p(k)$ is simulated with unit amplitude at time steps $k = 50s, 51s, \cdots, 100s$, i.e.,

$$p(k) = \begin{cases} 1, & 50s \le k \le 100s \\ 0, & \text{otherwise.} \end{cases}$$

Connecting the designed estimators to the CSTR system and letting the simulation run for $300s$, Fig. 6 demonstrates the evolutions of the state estimation errors $e_i(k) = s(k) - x_i(k)$ for all $i \in \mathcal{V}$. It can be seen that the estimation errors eventually approach zero as time goes on. Thus, the designed estimators well estimate the CSTR's states. For each node, the residual response $r_i(k)$ with or without process noise $w(k)$ and measurement noise $v_i(k)$ to the above FDI attack signal $p(k)$ is depicted in Fig. 7. The residual evaluation functions $f_i(0, T_e)$ and thresholds Th_i for all $i \in \mathcal{V}$ are illustrated in Fig. 8. By a simple calculation, the threshold on each node can be obtained as $Th_1 = 5.9391 \times 10^{-11}$, $Th_2 = 5.9889 \times 10^{-11}$, $Th_3 = 1.4327 \times 10^{-10}$ and $Th_4 = 3.7877 \times 10^{-10}$, respectively. From Fig. 8, it is found that $f_1(0, 58) = 5.3951 \times 10^{-11}$ and $f_1(0, 59) = 6.0383 \times 10^{-11}$; $f_2(0, 59) = 5.4950 \times 10^{-11}$ and $f_2(0, 60) = 6.0751 \times 10^{-11}$; $f_3(0, 62) = 1.2906 \times 10^{-10}$ and $f_3(0, 63) = 1.4774 \times 10^{-10}$; $f_4(0, 73) = 3.5666 \times 10^{-10}$ and $f_4(0, 74) = 3.8976 \times 10^{-10}$, which means that $f_1(0, 58) < Th_1 < f_1(0, 59)$, $f_2(0, 59) < Th_2 < f_2(0, 60)$, $f_3(0, 62) < Th_3 < f_3(0, 63)$ and $f_4(0, 73) < Th_4 < f_4(0, 74)$. Thus, the FDI attack signal $p(k)$ can be detected in 9 time steps after its occurrence by the proposed resilient attack detector estimator 1, 10 time steps after its occurrence by estimator 2, 13 time steps after its occurrence by estimator 3 and 24 time steps after its occurrence by estimator 4, respectively. When the occurrence of the FDI attack is detected by the designed estimators, an alarm or a warning signal of FDI attack occurrence can be generated to remind the operating engineers or designer to take further measures such as attack signal isolation and false data correction. Furthermore, it can be seen that the residual signals can not only reflect the FDI attack signal in time, but can also detect the FDI attack signal without confusing it with the process noise $w(k)$ and the measurement noise $v_i(k)$, which verifies the effectiveness of the derived results.

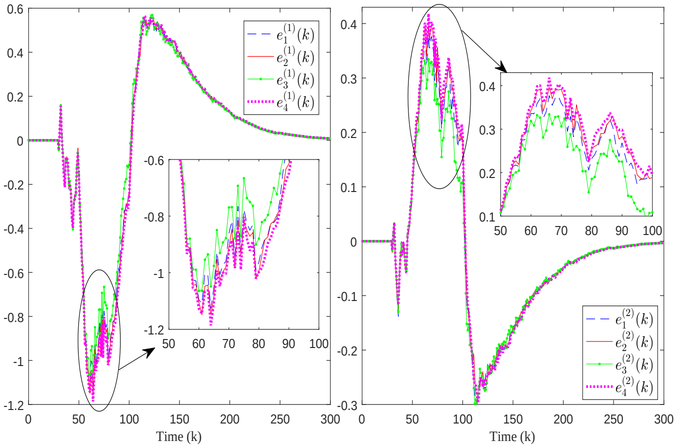


Fig. 6. The estimation errors $e_i(k) = s(k) - x_i(k)$, $i \in \mathcal{V}$

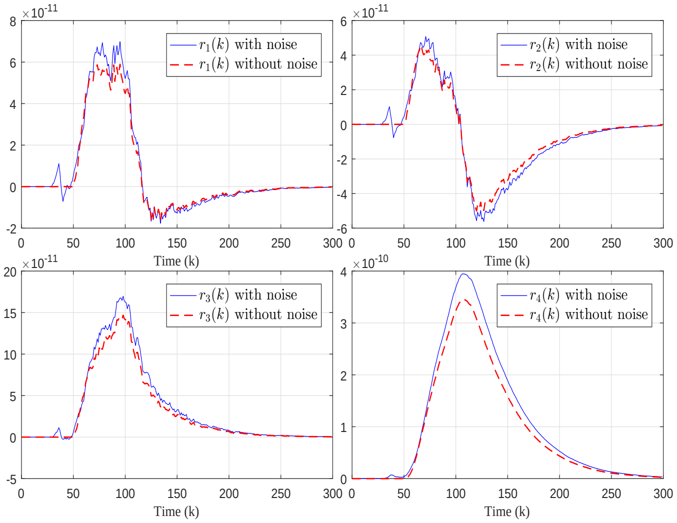


Fig. 7. The residual signals $r_i(k)$ with or without process noise $w(k)$ and measurement noise $v_i(k)$, $i \in \mathcal{V}$

V. Conclusion

The distributed attack detection and secure estimation problem for a CPS over a WSN in the presence of two types of malicious attacks have been studied. More specifically, an FDI attack has been considered at the physical system layer, where the adversary has injected false information to modify

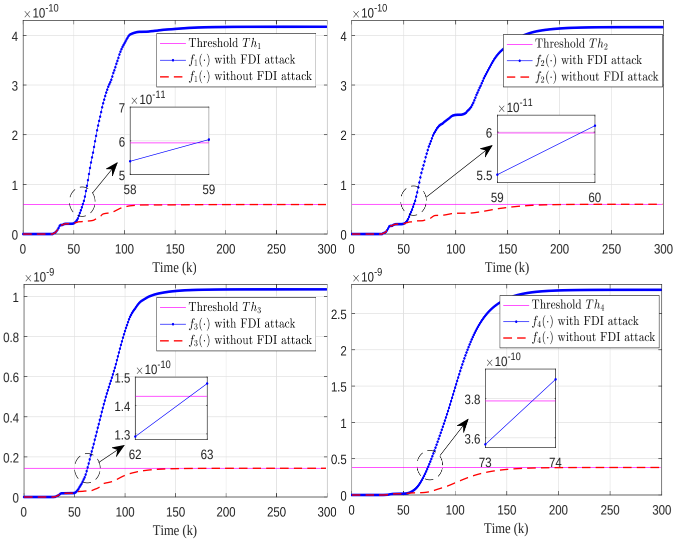


Fig. 8. The evaluation functions $f_i(\cdot)$ and thresholds Th_i, $i \in \mathcal{V}$

the system's state. Then, a class of random jamming attacks on wireless measurement transmission channels have been investigated. The effects of the physical and cyber attacks on the estimation performance of the resultant estimation error system have been analyzed. To handle the false data injection attack, a two-step attack detection mechanism has been established, through which the occurrence of the FDI attack can be detected and alarmed. To tackle the random jamming attacks, a refined measurement output model based on compensated measurements has been proposed and resilient estimators have been delicately constructed. Criteria for estimation performance analysis and estimator design have been derived to guarantee the feasibility of the problem. An extension of the proposed results to the case of uncertain measurement loss probability has also been studied. The effectiveness and applicability of the derived results have been verified via a networked continuous-stirred tank reactor system.

REFERENCES

[1] F. Pasqualetti, F. Dorfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Trans. Autom. Control*, vol. 58, no. 11, pp. 2715-2729, Nov. 2013.

[2] Y. Mo, R. Chabukswar, and B. Sinopoli, "Detecting integrity attacks on SCADA systems," *IEEE Trans. Control Syst. Technol.*, vol. 22, no. 4, pp. 1396-1407, Jul. 2014.

[3] L. Zhang, H. Gao, and O. Kaynak, "Network-induced constraints in networked control systems-A survey," *IEEE Trans. Ind. Informat.*, vol. 9, no. 1, pp. 403-416, Feb. 2013.

[4] X. Ge, F. Yang, and Q.-L. Han, "Distributed networked control systems: A brief overview," *Inf. Sci.*, vol. 380, pp. 117-131, Feb. 2017.

[5] 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.

[6] Y. Li, L. Shi, P. Cheng, *et al.*, "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.

[7] T. Bonaci, L. Bushnell, and R. Poovendram, "Node capture attacks in wireless sensor networks: A system theoretic approach," in *Proc. 49th IEEE Conf. Decision Control*, Atlanta, GA, USA, Dec. 2010, pp. 6765-6772.

[8] Y. Mo and B. Sinopoli, "Secure estimation in the presence of integrity attacks," *IEEE Trans. Autom. Control*, vol. 60, no. 4, pp. 1145-1151, Apr. 2015.

[9] Y. Shoukry and P. Tabuada, "Event-triggered state observers for sparse sensor noise/attacks," *IEEE Trans. Autom. Control*, vol. 61, no. 8, pp. 2079-2091, Aug. 2016.

[10] Y. Liu, P. Ning, and M. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM T Inform. Syst. Se.*, vol. 14, no. 1, pp. 1-33, May 2011.

[11] F. Pasqualetti, R. Carli, and F. Bullo, "A distributed method for state estimation and false data detection in power networks," in *Proc. IEEE Int. Conf. Smart Grid Commun.*, Brussels, Belgium, Oct. 2011, pp. 469-474.

[12] Y. Mo, E. Garone, A. Casavola, *et al.*, "False data injection attacks against state estimation in wireless sensor networks," in *Proc. 49th IEEE Conf. Decision Control*, Atlanta, GA, USA, Dec. 2010, pp. 5967-5972.

[13] Y. Li, D. Quevedo, S. Dey, *et al.*, "A game-theoretic approach to fake-acknowledgement attack on cyber-physical systems," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 1, pp. 1-11, Mar. 2017.

[14] R. Poisel, *Modern Communications Jamming Principles and Techniques*. Norwood, MA, USA: Artech House, 2011.

[15] H. Zhang, P. Cheng, L. Shi, *et al.*, "Optimal DoS attack scheduling in wireless networked control systems," *IEEE Trans. Control Syst. Tech.*, vol. 24, no. 3, pp. 843-852, May 2016.

[16] H. Zhang, P. Cheng, L. Shi, *et al.*, "Optimal denial-of-service attack scheduling with energy constraint," *IEEE Trans. Autom. Control*, vol. 60, no. 11, pp. 3023-3028, Nov. 2015.

[17] M. Zhu and S. Martínze, "On the performance analysis of resilient networked control systems under replay attacks," *IEEE Trans. Autom. Control*, vol. 59, no. 3, pp. 804-808, Mar. 2014.

[18] B. Kailkhura, S. Brahma, and P. Varshney, "Data falsification attacks on consensus-based detection systems," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 1, pp. 145-158, Mar. 2017.

[19] R. Gentz, S. Wu, H.-T. Wai, *et al.*, "Data injection attacks in randomized gossiping," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 2, no. 4, pp. 523-538, Dec. 2016.

[20] M. Zhu and S. Martínez, "On distributed constrained formation control in operator-vehicle adversarial networks," *Automatica*, vol. 49, no. 12, pp. 3571-3582, Dec. 2013.

[21] M. Blanke, M. Kinnaert, J. Lunze, *et al.*, Diagnosis and Fault-Tolerant Control. New York, NY, USA: Springer-Verlag, 2006.

[22] C. Chihaia, "Active fault-tolerance in wireless networked control systems," Ph.D. dissertation, Dept. Electr. Eng., University of Duisburg-Essen, Duisburg, Germany, 2010.

[23] X. Ge and Q.-L. Han, "Distributed fault detection over sensor networks with Markovian switching topologies," *Int. J. Gen. Syst.*, vol. 43, no. 3-4, pp. 305-318, Mar. 2014.

[24] H. Gao, T. Chen, and L. Wang, "Robust fault detection with missing measurements," *Int. J. Control*, vol. 81, no. 5, pp. 804-819, Apr. 2008.

[25] K. Koscher, A. Czeskis, F. Roesner, *et al.*, "Experimental security analysis of a modern automobile," in *Proc. 2010 IEEE Symp. Security Privacy*, Oakland, CA, USA, May 2010, pp. 447-462.

[26] C. Huang, D. Ho, and J. Lu, "Partial-information-based distributed filtering in two-targets tracking sensor networks," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 59, no. 4, pp. 820-832, Apr. 2012.

[27] X. Ge and Q.-L. Han, "Distributed event-triggered H_∞ filtering over sensor networks with communication delays," *Inf. Sci.*, vol. 291, pp. 128-142, Jan. 2015.

[28] S. Zhu, C. Chen, W. Li, *et al.*, "Distributed optimal consensus filter for target tracking in heterogeneous sensor networks," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1963-1976, Dec. 2013.

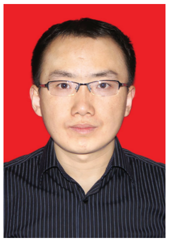
[29] X. Ge, Q.-L. Han, and X. Jiang, "Distributed H_∞ filtering over sensor networks with heterogeneous Markovian coupling intercommunication delays," *IET Control Theory Appl.*, vol. 9, no. 1, pp. 82-90, Jan. 2015.

[30] B. Shen, Z. Wang, and Y. Hung, "Distributed H_∞-consensus filtering in sensor networks with multiple missing measurements: The finite-horizon case," *Automatica*, vol. 46, no. 10, pp. 1682-1688, Oct. 2010.

[31] Q.-L. Han, "Absolute stability of time-delay systems with sector-bounded nonlinearity," *Automatica*, vol. 41, no. 12, pp. 2171-2176, Dec. 2005.

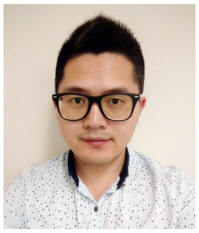
[32] E. Boukas and Z. Liu, "Robust H_∞ control of discrete-time Markovian jump linear systems with mode-dependent time-delays," *IEEE Trans. Autom. Control*, vol. 46, no. 12, pp. 1918-1924, Dec. 2001.

[33] K.-U. Klatt and S. Engell, "Gain-scheduling trajectory control of a continuous stirred tank reactor," *Computers Chem. Engng.*, vol. 22, no. 4/5, pp. 491-502, Aug. 1998.



Yanpeng Guan received the B.Sc. degree in mathematics from Changchun Normal University, Changchun, China, in 2005, the M.Eng. degree in control theory and control engineering from Hangzhou Dianzi University, Hangzhou, China, in 2010, and the Ph.D. degree in computer engineering from Central Queensland University, Rockhampton, QLD, Australia, in 2014, respectively. He is currently a lecturer with the Department of Automation, Shanxi University.

His research interests include networked control systems, cyber-physical systems, secure estimation, distributed control and estimation, and event-triggered control.



Xiaohua Ge received the B.Eng. degree in electronic and information engineering from Nanchang Hangkong University, Nanchang, China, in 2008, the M.Eng. degree in control theory and control engineering from Hangzhou Dianzi University, Hangzhou, China, in 2011, and the Ph.D. degree in computer engineering from Central Queensland University, Rockhampton, QLD, Australia, in 2014.

He was a Research Assistant with the Centre for Intelligent and Networked Systems, Central Queensland University, from 2011 to 2013. In 2014, he was a Research Fellow with the Centre for Intelligent and Networked Systems, Central Queensland University, Rockhampton, Australia. From 2015 to 2016, he was a Research Fellow with the Griffith School of Engineering, Griffith University, Gold Coast, Australia. He is currently a Lecturer with the School of Software and Electrical Engineering, Swinburne University of Technology, Melbourne, Australia.

His current research interests include networked control and filtering, distributed networked control systems, multi-agent systems and sensor networks.