Doostmohammadian, Mohammadreza; Zarrabi, Houman; Rabiee, Hamid R.; Khan, Usman A.; Charalambous, Themistoklis

Distributed Detection and Mitigation of Biasing Attacks over Multi-Agent Networks

Published in:
IEEE Transactions on Network Science and Engineering

DOI:
10.1109/TNSE.2021.3115032

Published: 01/12/2021

Document Version
Publisher's PDF, also known as Version of record

Distributed Detection and Mitigation of Biasing Attacks Over Multi-Agent Networks

Mohammadreza Doostmohammadian, *Member, IEEE*, Houman Zarrabi, *Member, IEEE*, Hamid R. Rabiee, *Senior Member, IEEE*, Usman A. Khan, *Senior Member, IEEE*, and Themistoklis Charalambous, *Senior Member, IEEE*

***Abstract*—This paper proposes a distributed attack detection and mitigation technique based on distributed estimation over a multi-agent network, where the agents take partial system measurements susceptible to (possible) biasing attacks. In particular, we assume that the system is not locally observable via the measurements in the direct neighborhood of any agent. First, for performance analysis in the attack-free case, we show that the proposed distributed estimation is unbiased with bounded mean-square deviation in steady-state. Then, we propose a residual-based strategy to locally detect possible attacks at agents. In contrast to the deterministic thresholds in the literature assuming an upper bound on the noise support, we define the thresholds on the residuals in a probabilistic sense. After detecting and isolating the attacked agent, a system-digraph-based mitigation strategy is proposed to replace the attacked measurement with a new observationally-equivalent one to recover potential observability loss. We adopt a graph-theoretic method to classify the agents based on their measurements, to distinguish between the agents recovering the system rank-deficiency and the ones recovering output-connectivity of the system digraph. The attack detection/mitigation strategy is specifically described for each type, which is of polynomial-order complexity for large-scale applications. Illustrative simulations support our theoretical results.**

***Index Terms*—Biasing Attacks, False-Data Injection, Distributed Observability, Distributed Estimation, Structural Analysis.**

I. Introduction

DATA (or measurements) regarding many real-world systems, such as wireless sensor networks, multi-agent robotic systems, block-chain and cloud-computing, smart energy networks, are naturally distributed over large geographical regions [1], [2]. Collecting all these to a central coordinator (or a fusion center) for the purposes of processing and learning is tedious and impractical in many applications. Distributed learning or inference is thus typically preferred, due to the fact that it does not require long-range communication to a central unit. The corresponding distributed strategies are practically feasible as they rely on local data processing and local communication only among the neighboring agents. However, such decentralized strategies are vulnerable to malicious attacks. In this paper, we consider distributed detection and mitigation of biasing attacks at sensors/agents performing distributed estimation over a large-scale dynamical system. Potential applications include secure distributed estimation over Cyber-Physical-Systems (CPS) [3]–[9], Internet-of-Things (IoT) [10]–[12], smart cities [13], social networks [14]–[16], and power-grid monitoring systems [2], [17]–[25] among others.

In distributed estimation (or filtering) applications [26]–[28] a *multi-agent network* is referred to a group of agents with sensing, data-processing, and communication capabilities, which take (noisy) output or measurements of the dynamical system, share their information over a network, and process the received data locally to track the system state. In case of erroneous or biased data [29], [30], the distributed estimation performance is significantly degraded if the biased measurements are necessary for *observability*. Recall that observability refers to the possibility of inferring the (entire) states of the dynamical system via tracking outputs/measurements of a *subset* of states over a finite time. This is more challenging in *single time-scale* estimation with only one step of data-fusion between every two consecutive time-steps of system dynamics, and with no local observability (i.e., the system is not observable in the neighborhood of any agent) [26], [27], [31]–[35]. This differs from double time-scale estimation where all necessary information for observability is *directly* communicated to every agent from its neighbors. This requires considerably more communication traffic and information exchange over the network. This implies that the biased (attacked) measurement affects the *residual* (defined as the deviation of the estimated/expected output from the original system output [36]) at more agents, making it harder to locally *isolate* the faulty sensor. Such additive bias could be, for example, due to false-data injection attacks [37]. The general idea in

Manuscript received January 28, 2021; revised September 8, 2021; accepted September 19, 2021. Date of publication September 24, 2021; date of current version December 9, 2021. The work of UAK was supported in part by NSF under Awards CMMI-1903972 and CBET-1935555. The work of Themistoklis Charalambous was supported by the Academy of Finland under Grant 317726. Recommended for acceptance by Dr. Xiaoming Fu. *(Corresponding author: Mohammadreza Doostmohammadian.)*

Mohammadreza Doostmohammadian is with the Faculty of Mechanical Engineering, Semnan University, Semnan 35131, Iran, and also with the School of Electrical Engineering, Aalto University, 02150 Espoo, Finland (e-mail: mohammadreza.doostmohammadian@aalto.fi).

Houman Zarrabi is with the Iran Telecommunication Research Center (ITRC), Tehran 19841, Iran (e-mail: h.zarrabi@itrc.ac.ir).

Hamid R. Rabiee is with the School of Computer Engineering, Sharif University of Technology, Tehran 11155, Iran (e-mail: rabiee@sharif.edu).

Usman A. Khan is with the Electrical and Computer Engineering Department, Tufts University, Medford, MA 02155 USA (e-mail: khan@ece.tufts.edu).

Themistoklis Charalambous is with the School of Electrical Engineering, Aalto University, 02150 Espoo, Finland (e-mail: themistoklis.charalambous@aalto.fi).

Digital Object Identifier 10.1109/TNSE.2021.3115032

this work is to *locally* detect and isolate such attacks and, further, reconfigure the multi-agent network using substitute measurements to recover (potential) loss of observability.

The distributed estimator in this paper performs consensus (on the received data) at the same time-scale of the underlying system (single time-scale), see e.g., [38], [39] for details. We use structured systems theory [39]–[41]) to guarantee *generic* or *structural* observability. This helps to partition the system outputs (fed to the agents) into certain observationally-equivalent classes [42]. This gives the set of necessary agents for estimation (whose removal makes the system unobservable) and the set of redundant agents (whose removal results in no observability loss). Subsequently, different strategies are used to substitute the faulty sensor and design inter-agent communications. We propose our attack detection and mitigation strategy based on this specific *agent classification*. In particular, we show that isolation of the attacks related to system rank-deficiency is more challenging and requires certain constrained gain design. Recall that system rank refers to the rank of the associated matrix to the linear system of differential equations (in the state-space representation), see Section II-A for more details.

Comparison with related literature: This work develops a *joint* distributed estimation and attack detection/isolation technique, and extends the prior works on resilient distributed estimation subject to *unreliable* sensor measurements [43], [44] and adversarial attacks [32], [45]–[52]. These literature do not detect/isolate the attack, but estimate the system in the presence of (specific) attacks with bounded (steady-state) error, while making simplifying assumptions, e.g., a *noise-free* model. Our work extends [32], [43]–[52] by further considering *distributed/localized* techniques to locate the attacked sensor. Further, this work differs from many works on distributed estimation in the literature by relaxing the observability assumption; for example, [26], [27], [31]–[35] assume local observability at some (or all) agents. In contrast, and similar to [28], [53], [54], our work makes no such restrictive assumption. However, [28], [53], [54] perform many iterations of data-fusion (consensus) between two consecutive system steps (*double time-scale* estimation), requiring much faster data-processing/communication rate.

In the context of adversarial attacks, most observer-based detection scenarios assume system and/or measurement noise with bounded support, i.e., they consider an upper bound on the noise variable [55]–[59]. In this paper, we make no such assumption; instead, the noise is assumed to be of infinite support (i.e., it can take any arbitrarily large value with bounded second-order moment). Therefore, we propose probabilistic attack-detection thresholds, in contrast to the deterministic threshold design (or *flag value*) in observer-based detection methods [55]–[59]. In another line of research [60]–[67], distributed attack detection without observer/estimator design is considered. These works consider a multi-agent network aiming to detect (typically Byzantine) attack in a sensed signal in a distributed way, with no estimation purpose (due to unknown system model). For example, [7] uses innovation variance to detect attacks (component malfunctions) in linear-quadratic-Gaussian (LQG) CPS models. However, our main goal is to detect the attacks (in form of biasing anomalies changing the true output values [30]) deteriorating distributed estimation performance, and, further, to provide a mitigation strategy to restore observability (more precisely, *distributed observability* [68]). In this regard, this paper performs *simultaneous distributed* estimation and attack-detection, which makes it different from [7], [60]–[67] performing *only* detection.

Of relevance are also *watermarking* strategies [69], [70], that inject a known input signal (watermark) into the system and track this watermark in the outputs using Chi-square testing (χ^2-detector). Such input injection is not possible for tracking *autonomous* systems, and thus, the physical watermarking is impractical in such cases. The distributed strategy in this work is not limited to full-rank LTI systems, in contrast to distributed estimators in [29], [33]–[35], [71] over strongly-connected (SC) sensor-networks. Further, unlike the static parameter estimation in [72] and noiseless centralized attack-detection/estimation in [31], this work is based on *distributed* estimation of *noise-corrupted* linear systems. Another relevant topic is compressive sensing [73]–[78] to translate the data into a compressed dimension, share and combine the data, reconstruct it to the full dimension, and perform diffusion-based [78] or least mean square (LMS) update [75]–[77] to estimate the original signal. Although the compressed *transmit* of data is applicable in our work (to reduce the communication burden), *distributed dynamic* observability makes our work different from [25], [75]–[78] based on *static* observability irrespective of the dynamic system model. Recall that this is referred to as the *Static Linear State-Space* (SLS) model in detection literature [36] and differs from our solution considering *Linear Dynamical State-Space* (LDS) model[1]. Similarly this work differs from *centralized* estimation in [73], [74] with certain assumptions on the sparsity of the initial states [73], [74] or system rank [73]. Autoencoder-based learning is used in some works [80]–[83] to distinguish (classify) faulty/attacked data from non-attacked measurement data. In smart-grid applications, the PMU measurements are used to train the detector via either supervised learning [80], [81] or unsupervised learning [82]. No dynamics is considered in these works (SLS model), contrasting our (distributed) observability-based LDS model. Further, [80]–[82] only perform detection with no aim of estimation in the absence of attacks, while some works (see references in [83]) only perform learning-based estimation with no possibility of detection. Recall that noise (in system dynamics and/or output) plays a key role in the LDS detection. As mentioned before, the assumption on the noise support (finite or infinite) and its value in the finite-case affects the performance of the detection mechanism

[1] Using the dynamic model of the system (LDS case), fewer outputs are needed to reconstruct the full state of the system (dynamic observability), while in the static or SLS case (with no information of system dynamics) in general more outputs (as many as system states) are needed. Having fewer outputs (than the number of states) in the SLS case results in under-determined system of linear equations (unobservability), which mandates substitute recovering solutions such as compressive-sensing or auto-encoder neural networks [79]. A compressive-sensing-based example for the smart-grid application is given in [25], which requires no rank condition on the SLS model.

TABLE I
NOTATIONS IN THE PAPER (SUBSCRIPT k IMPLIES PARAMETER'S TIME INDEX)

n, N	number of system states, measurements (agents)
k	time index
$\mathbf{x}_k, \mathbf{y}_k$	column-vector of states, measurements
$\boldsymbol{\nu}_k, \boldsymbol{\zeta}_k$	zero-mean system and measurement noise
$\boldsymbol{\tau}_k$	attack vector at time k
A, C	system and measurement matrix
E, R	system and measurement noise covariance
$\mathbf{c}_j$	measurement matrix (column-vector) at agent j
α, β, γ	types of agents
$\mathcal{G}_A$	system digraph associated with system matrix A
$\mathcal{G}_N, \mathcal{G}_\alpha, \mathcal{G}_\beta$	communication network of agents
$\mathcal{N}_\alpha(i), \mathcal{N}_\beta(i)$	set of neighbors of agent i over networks $\mathcal{G}_\alpha, \mathcal{G}_\beta$
$\mathcal{N}(\cdot,\cdot)$	Gaussian distribution
$\mathcal{C}, \mathcal{S}^p$	contraction and parent SCC in the system graph
W	stochastic fusion-matrix associated to $\mathcal{G}_\beta$
U	adjacency matrix of $\mathcal{G}_\alpha$
K	gain matrix (with K_i as i-th diagonal-block)
$\widehat{\mathbf{x}}^i_{k\|k-1}, \widehat{\mathbf{x}}^i_{k\|k}$	*a priori* and *a posteriori* estimate of agent i
$\mathbf{e}^i_k, r^i_k$	estimation error and residual of agent i
θ_κ	detection threshold associated with probability κ
κ,$\varkappa$	threshold probability, probability of false-alarm
$\mathbf{1}_{NN}, \mathbf{1}_N$	all 1's matrix/column-vector of size N
I_N	Identity matrix of size N
$\mathbb{E}$	Expected value operator

[55]–[59]. Similarly, noise in the output data affects the SLS detection performance, e.g., in power-system applications [25], [75]–[78]. See more details along with a review of centralized physics-based detection mechanisms in [36].

Main contributions:

i) Our observer-based detection strategy is *localized* and *distributed* over the multi-agent network with *no local observability* assumption at any agent, but *global observability* at the group of agents. This is key in large-scale, as it enables each agent to detect a (possible) attack on its received output with no central coordination, in contrast to centralized detection scenarios.

ii) Using certain agent classification based on system-rank, we develop detection and attack isolation strategies which are specific to the measurement types based on the system dynamics (LDS model) (see Section II-B for detailed explanation).

iii) The noise is considered over an infinite range with no constraint/bound on its support, which is more realistic for real-world applications (see Remark 1). In this sense, our attack detection and mitigation is categorized as probabilistic (vs. deterministic) thresholding.

iv) In order to prevent repetitive attacks at the same agent by the adversary, we consider an attack mitigation strategy to replace the biased measurement with an observationally-equivalent one (borrowing results from [42], [84]).

We emphasize that the proposed algorithms for threshold design, agent classification, and mitigation via observational equivalency are of polynomial-order complexity.

Notation: Throughout this paper, scalar and (column) vector variables are respectively represented by lower-case and bold lower-case letters. Further, capital letters represent matrices. The induced 2-norm of the matrix A is defined as $\|A\|_2 = \sqrt{\lambda_n}$ where $\lambda_n = \rho(A^\top A)$ and $\rho(\cdot)$ denotes the spectral radius of matrix. Further, $|\cdot|$ denotes the Euclidean norm. Table I summarizes the notation in this paper.

II. PROBLEM SETUP

A. Linear Dynamical System

Following the discussions in Section I, we consider *noise-corrupted* linear discrete-time systems (LDS model [36]) as,

$$\mathbf{x}_{k+1} = A\mathbf{x}_k + \nu_k, \tag{1}$$

with $\mathbf{x}_k \in \mathbb{R}^n$ as the column-vector of states at time k, A as the system matrix, and $\nu_k \sim \mathcal{N}(0, E)$ as the system noise vector. Throughout the paper, the system-rank refers to the rank of the system matrix A. Consider a group of N agents with scalar outputs given by $y^i_k = \mathbf{c}_i^\top \mathbf{x}_k + \zeta^i_k + \tau^i_k$ and the vector form as,

$$\mathbf{y}_k = C\mathbf{x}_k + \zeta_k + \tau_k, \tag{2}$$

with $\mathbf{y}_k \in \mathbb{R}^N$ as the column-vector of state measurements (or system outputs) $\mathbf{y}_k = (y^1_k, \ldots, y^N_k)^\top$, $\zeta_k = (\zeta^1_k, \ldots, \zeta^N_k)^\top \sim \mathcal{N}(0, R)$ as the measurement noise vector, and $\tau_k = (\tau^1_k, \ldots, \tau^N_k)^\top$ as the column-vector of biasing attack at the agents. We assume arbitrary attack τ_k by the adversary, e.g., both fixed stationary attack and non-stationary attacks are considered for simulation (Section V). Further, the measurement matrix $C = [\mathbf{c}_1^\top; \ldots; \mathbf{c}_N^\top]$ is the column concatenation of row-vectors $\mathbf{c}_i^\top$ associated with agent i (with ";" as column concatenation). Standard assumptions on Gaussianity and independence of noise terms are considered. For example, it is typical to assume that the sensor measurements are independent, making the measurement noise covariance matrix R diagonal.

Remark 1: Several papers in the literature (e.g., [55]–[59]) assume constrained noise $|\nu_k| < \delta$ and/or $|\zeta_k| < \delta$, where the upper bound δ on the noise support sets the deterministic thresholds for attack detection. For example, in [56] the deterministic threshold at sensor i is defined as $\mathcal{D}_i = \|\mathcal{O}_i\|_2 \|\mathbf{e}^i_k\|_2 + 2\delta$ with $\|\mathcal{O}_i\|_2$ and $\|\mathbf{e}^i_k\|_2$ as the 2-norm of the observability Grammian and the state-estimation error, respectively. In contrast, we make no such finite support assumption (loosely speaking, $\delta \to \infty$), while it is standard to assume that the second moments of the noise terms are finite, i.e., $\mathbb{E}(\nu_k^\top \nu_k) < \infty$ and $\mathbb{E}(\zeta_k^\top \zeta_k) < \infty$. Assuming unbounded δ, the deterministic threshold, for example $\mathcal{D}_i$ in [56], also goes unbounded ($\to \infty$), and thus, no attack can be detected. Similar arguments hold for [55], [57]–[59].

Remark 2: Note that noise Gaussianity is a standard assumption in most distributed estimation/filtering and attack detection literature, e.g., see [2]–[6], [8], [17]–[20], [28], [30], [31], [43], [44], [47], [48], [60], [63], [67], [69]–[71], [85]–[87].

B. Agent Classification Based on Structural Analysis

The notion of observability used throughout this paper is structural [40], [86], [88] and the theory is build on this notion.

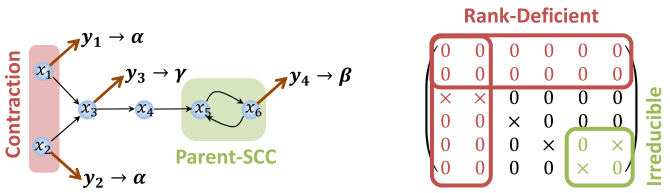


Fig. 1. This figure illustrates the proposed agent/measurement classification over a simple system digraph and its associated system matrix: α-agents with outputs y_1 and y_2 from a contraction (two state nodes contracting into one state node), β-agent with output y_4 from a parent SCC (two linked state nodes with no outgoing link to other components), and a redundant γ-agent (with output y_3) which is neither type α nor type β. As illustrated in the zero-nonzero pattern of the system matrix (right), the contraction represents system (structural) rank-deficiency and the parent-SCC is associated with the irreducible block (with zero entries in the upper/lower non-diagonal blocks). See more details in [42].

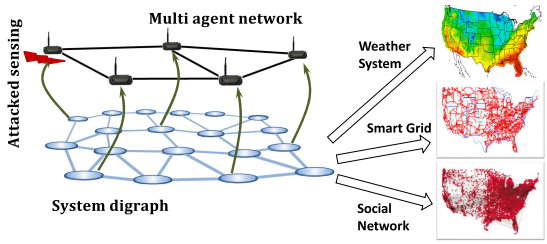


Fig. 2. In this work, there exist two graph representations: (i) system digraph $\mathcal{G}_A$ (see Section II-B), representing the interactions of system states (or system nodes), and (ii) multi-agent network (denoted by $\mathcal{G}_N = \mathcal{G}_\alpha \cup \mathcal{G}_\beta$). The system digraph models a large-scale state-space system, e.g., social network, power grid, or weather system. The green arrows show the state measurements/outputs which vulnerable to (possible) adversarial attacks. The agents/sensors are classified based on their state measurements from specific components in $\mathcal{G}_A$ (see examples in Section V) and track the global system state locally (i.e., performing distributed estimation) via sharing information over the network $\mathcal{G}_N$. Attacked (biased) measurements may affect the estimation performance at all agents. The proposed algorithm in this work enables each agent to *locally* detect if its measurement/output is attacked or not, and further provides mitigation techniques for resilient estimation.

It is known that the rank deficiency of matrix A and strong-connectivity of *system digraph* $\mathcal{G}_A$ affect its structural observability properties, and further, its estimation performance. In this direction, using structured systems theory and generic analysis [40], [88], we propose specific sensor/agent classification based on the structure (zero-nonzero pattern) of the system matrix A and system digraph $\mathcal{G}_A$. Using the theory developed in [5], [42], the agents are partitioned into different classes based on their state-measurements. We specifically show in Section IV-B that the detection and mitigation logic differs for each class. First, we describe some relevant graph-theoretic notions. In $\mathcal{G}_A$, every node represents a state and every link represents a fixed non-zero entry of A (A_{ij} implies $j \to i$ as a link from node j to node i). In $\mathcal{G}_A$ a strongly-connected-component (SCC) is a component in which every node is connected to every other node via a path. Define a parent SCC $\mathcal{S}_l^p$ as an SCC with no out-going links to other SCCs. Further, a contraction $\mathcal{C}_l \in \mathcal{C}$ is a component for which $|\mathcal{N}_\mathcal{G}(\mathcal{C}_l)| < |\mathcal{C}_l|$, with $\mathcal{N}_\mathcal{G}(\mathcal{C}_l) = \{b | a \to b, a \in \mathcal{C}_l\}$ and $|\cdot|$ as the set cardinality. Based on these graph components, three types of agents are defined as follows,

- **α-agent** is an agent with measurement of a state node in a contraction $\mathcal{C}_l$.
- **β-agent** is an agent with measurement of a state node in a parent SCC $\mathcal{S}_l^p$.
- **γ-agent** is any agent which is neither type α nor β.

An example of such classification is given in Fig. 1 (and Section V). This partitioning has two advantages: (i) it allows using a different communication topology for different types of agents and simpler topology design when one or the other type of agents is not present; and (ii) it allows for the attack detection and mitigation strategy to be specifically defined for each type (see details in Section IV-B). In particular, following [85], it can be shown that any α-agent recovers the (structural) rank condition for observability, while the β-agent recovers the output-connectivity of the system digraph [86]. Therefore, both α and β-agents are necessary for observability, while removing (redundant) γ-agents has no effect on system observability. Recall that the structural properties are irrespective of the numerical values of system parameters [40]; therefore, for a structure-invariant matrix A the proposed classification is fixed and time-invariant.

C. Problem Statement

This paper considers a group of sensors/agents taking *noise-corrupted* measurements in the form (2) of a dynamical system (e.g., social network or power grid) in the form (1) represented by a *system digraph* $\mathcal{G}_A$, see Fig. 2. The agents perform distributed estimation over a network, denoted by $\mathcal{G}_N = \mathcal{G}_\alpha \cup \mathcal{G}_\beta$ to track the state of the noisy dynamical system (1). Note that the networks $\mathcal{G}_\alpha$, $\mathcal{G}_\beta$, and their union $\mathcal{G}_N$ include all the agents of type α, β, and γ. It is assumed that an adversarial attacker aims to add an arbitrary value τ_k^i (at any time k) to make the measurement at (one or more) agent i biased from its original value. Since the dynamical system is not necessarily observable at any agent, the biased measurements (at α/β-agents) affect the estimation error at all agents and result in the degradation of the distributed estimation performance. The problem here is to find a strategy to detect (and isolate) such instantaneous attacks *locally at each agent*. In particular, we propose a probabilistic detection strategy that returns the probability of attack (at each agent), instead of deterministic strategies returning 0-1 (NoAttack-Attack). The next question addressed in this paper is how to recover the potential loss of observability due to removing the attacked measurement depending on its type (α, β, or γ). Such countermeasures prevent the same adversarial attack by removing the attacked agent/measurement. As explained in Section IV-B, the attacked measurement can be replaced with a new *observationally-equivalent* one to avoid possible repetitive attacks at the same agent.

D. Assumptions

i) The pair (A, C) is observable. The pairs $(A, \mathbf{c}_j^\top)$ and $(A, \mathbf{c}_{\mathcal{N}_\alpha(j)}^\top)$ are not necessarily observable at any sensor j or in its neighborhood denoted by $\mathcal{N}_\alpha(j) \cup \mathcal{N}_\beta(j)$ (see details in Section III). This implies that the

underlying system A is not necessarily observable in the neighborhood of any agent.

ii) The noise terms ν_k, ζ_k are iid Gaussian, see Remark 1.

iii) The known system matrix A is not necessarily stable, i.e., its spectral radius $\rho(A)$ can be potentially greater than 1. In other words, this paper applies to both stable and unstable systems.

iv) The adversary can manipulate the state measurements at a subset of sensors by adding erroneous additive term τ_k at any time k. For example, τ_k^i can be from a uniform distribution over $[-l_\tau, l_\tau]$ with $l_\tau \gg \|R\|_2$, $l_\tau \gg \|E\|_2$ ($l_\tau \to \infty$ in general) or τ_k^i can be a fixed value. In general, the term τ_k^i may be non-zero at some time-instants k (instantaneous attack) and zero at some other times.

III. Distributed Estimation Under Possible Measurement Attacks

In this section, we propose a consensus-based distributed estimation (filtering) protocol over the multi-agent network. The proposed protocol performs *one* iteration of information sharing and consensus between every two consecutive steps of system dynamics as follows:

$$\widehat{\mathbf{x}}^i_{k|k-1} = \sum_{j\in\mathcal{N}_\beta(i)} W_{ij} A \widehat{\mathbf{x}}^j_{k-1|k-1}, \tag{3}$$

$$\widehat{\mathbf{x}}^i_{k|k} = \widehat{\mathbf{x}}^i_{k|k-1} + K_i \sum_{j\in\mathcal{N}_\alpha(i)} \mathbf{c}_j \Big(y^j_k - \mathbf{c}_j^\top \widehat{\mathbf{x}}^i_{k|k-1} \Big), \tag{4}$$

where y_k^j is the measurement of agent j at time k that could be attack-corrupted (or biased), $\mathcal{N}_\beta(i)$ and $\mathcal{N}_\alpha(i)$ are the neighborhood of agent i, respectively, over network $\mathcal{G}_\beta$ and $\mathcal{G}_\alpha$, K_i is the local feedback gain (or the observer gain) matrix at agent i, and $\widehat{\mathbf{x}}^i_{k|k-1}$ and $\widehat{\mathbf{x}}^i_{k|k}$ are the (column-vector of) estimates of system state $\mathbf{x}_k$ at agent i given the measurements, respectively, up to time $k-1$ and k. In fact, $\widehat{\mathbf{x}}^i_{k|k-1}$ is the *a-priori* estimate (or *prediction*) and $\widehat{\mathbf{x}}^i_{k|k}$ is the *posteriori* estimate after *measurement-update* at time-step k.

Remark 3: In this work, the combination of the following two graphs forms the multi-agent network: (i) $\mathcal{G}_\beta$ over which agents share the estimates $\widehat{\mathbf{x}}^j_{k-1|k-1}$, and (ii) $\mathcal{G}_\alpha$ over which agents share their measurements y^j_k. Define matrices W and U as the associated matrices to the graphs $\mathcal{G}_\beta$ and $\mathcal{G}_\alpha$, respectively. The matrix $U = \{U_{ij}\}$ is the 0–1 adjacency matrix of $\mathcal{G}_\alpha$, with $U_{ij} = 1$ associated to the link $j \to i$ in $\mathcal{G}_\alpha$ from α-agent j to every agent i. The *non-zero* entries of $W = \{W_{ij}\}$ take values in the range $0 < W_{ij} \le 1$ associated to the link $j \to i$ in $\mathcal{G}_\beta$.

Matrix W is row-stochastic to ensure *consensus* on a-priori estimates, i.e., $\sum_{j=1}^n W_{ij} = \sum_{j\in\mathcal{N}_\beta(i)} W_{ij} = 1$ for all i, j. Such a matrix W (and the graph $\mathcal{G}_\beta$) can be formed via distributed algorithms in [89]. The structure of $\mathcal{G}_\beta$ and $\mathcal{G}_\alpha$ (and the associated matrices) need to be designed properly for bounded steady-state estimation error, see Section III-A.

Remark 4: The proposed protocol (3)-(4) is a single time-scale distributed estimator, where the estimation is performed at the same time-scale of the system dynamics. This is in contrast to the double time-scale protocols [28], [53], [54], which require much faster estimation and communication rate than the sampling rate of the system dynamics, and, therefore, demand more costly communication and processing equipment. However, the observability assumption in [28], [53], [54] is similar to Assumption (ii), which makes such scenarios suitable for large-scale applications as the proposed protocol (3)-(4); see examples in Section V.

Denote the estimation error at agent i at time k by $\mathbf{e}^i_k \triangleq \mathbf{x}_{k|k} - \widehat{\mathbf{x}}^i_{k|k}$ and let $\mathbf{e}_k = (\mathbf{e}^1_k; \ldots; \mathbf{e}^N_k)$ be the global or collective error. Then, the following proposition defines the error dynamics of the protocol (3)-(4).

Proposition 1: The global error dynamics for protocol (3)-(4) is,

$$\mathbf{e}_k = (W\otimes A - KD_C(W\otimes A))\mathbf{e}_{k-1} + \eta_k = \widehat{A}\mathbf{e}_{k-1} + \eta_k, \tag{5}$$

$$\eta_k = \mathbf{1}_N \otimes \nu_{k-1} - KD_C(\mathbf{1}_N \otimes \nu_{k-1}) - K\overline{D}_C\zeta_k - K\overline{D}_C\tau_k, \tag{6}$$

where η_k collects the noise terms, $\widehat{A} := W\otimes A - KD_C(W\otimes A)$, $K \triangleq \text{blockdiag}[K_i]$, $D_C \triangleq \text{blockdiag}[\sum_{j\in\mathcal{N}_\alpha(i)} \mathbf{c}_j\mathbf{c}_j^\top]$, and $\overline{D}_C \triangleq (U\otimes \mathbf{1}_n)\circ(\mathbf{1}_N \otimes C^\top)$ with "$\circ$" and "$\otimes$," respectively, as the entrywise (Hadamard) and Kronecker product.

Proof: The error at each agent i is as follows,

$$\mathbf{e}^i_k = \mathbf{x}_k - \Bigg(\sum_{j\in\mathcal{N}_\beta(i)} W_{ij} A \widehat{\mathbf{x}}^j_{k-1|k-1} + K_i \sum_{j\in\mathcal{N}_\alpha(i)} \mathbf{c}_j \Bigg(y^j_k - \mathbf{c}_j^\top \sum_{j\in\mathcal{N}_\beta(i)} W_{ij} A \widehat{\mathbf{x}}^j_{k-1|k-1} \Bigg) \Bigg).$$

Recalling stochasticity of W matrix, we have $A\mathbf{x}_{k-1} = \sum_{j\in\mathcal{N}_\beta(i)} W_{ij} A\mathbf{x}_{k-1}$. Substituting this along with equations (1)-(2),

$$\begin{aligned}
\mathbf{e}^i_k &= \sum_{j\in\mathcal{N}_\beta(i)} W_{ij} A\mathbf{x}_{k-1} - \sum_{j\in\mathcal{N}_\beta(i)} W_{ij} A \widehat{\mathbf{x}}^j_{k-1|k-1} \\
&\quad - K_i \sum_{j\in\mathcal{N}_\alpha(i)} \mathbf{c}_j\mathbf{c}_j^\top \Bigg(\sum_{j\in\mathcal{N}_\beta(i)} W_{ij} A\mathbf{x}_{k-1} - \sum_{j\in\mathcal{N}_\beta(i)} W_{ij} A \widehat{\mathbf{x}}^j_{k-1|k-1} \Bigg) \\
&\quad + \nu_{k-1} - K_i \sum_{j\in\mathcal{N}_\alpha(i)} \mathbf{c}_j\zeta^j_k + \mathbf{c}_j\tau^j_k + \mathbf{c}_j\mathbf{c}_j^\top \nu_{k-1} \\
&= \sum_{j\in\mathcal{N}_\beta(i)} W_{ij} A\mathbf{e}^j_{k-1} - K_i \sum_{j\in\mathcal{N}_\alpha(i)} \mathbf{c}_j\mathbf{c}_j^\top \sum_{j\in\mathcal{N}_\beta(i)} W_{ij} A\mathbf{e}^j_{k-1} + \eta^i_k,
\end{aligned} \tag{7}$$

with $\eta^i_k \triangleq \nu_{k-1} - K_i \sum_{j\in\mathcal{N}_\alpha(i)} (\mathbf{c}_j\zeta^j_k + \mathbf{c}_j\tau^j_k + \mathbf{c}_j\mathbf{c}_j^\top \nu_{k-1})$. Using the definition of Kronecker and entrywise products, the collective error and noise term follow (5)-(6). ■

A. Error Stability

The following lemma establishes the stability condition of the error dynamics (5)-(6).

Lemma 1: The necessary condition for error dynamics (5)-(6) to be stable is that the pair $(W\otimes A, D_C)$ is observable.

Proof: The proof follows the Kalman stability theorem on the error dynamics (6). More information can be found in [86], [90], [91] on error stability of linear observer design. ■

Note that $(W \otimes A, D_C)$-observability is also referred to as the distributed observability [68]. Using structured system theory (generic analysis), distributed observability can be formulated as the observability of the Kronecker product of the graphs $\mathcal{G}_A$ and $\mathcal{G}_\beta$. Following the observability analysis of Kronecker composite networks in [92], the following lemma determines the sufficient connectivity of $\mathcal{G}_\beta$ and $\mathcal{G}_\alpha$.

Lemma 2: The pair $(W \otimes A, D_C)$ is observable if and only if the following conditions hold:

1) $\mathcal{G}_\beta$ is strongly-connected (SC) with self-link at each agent, which further implies that W is irreducible.
2) $\mathcal{G}_\alpha$ is a hub-network in which every α-agent is a hub, i.e., there is a directed link from every α-agent to every other agent in $\mathcal{G}_\alpha$. Further, $i \in \mathcal{N}_\alpha(i)$ for every agent i.

Proof: We provide the sketch of the proof here and refer the interested reader to [92] for more details. For (structural) observability two conditions on the associated composite graph need to be satisfied [86], [88]: (i) the output connectivity condition, implying the existence of a directed path from every state node in the system graph $\mathcal{G}_A$ to an agent (output), and (ii) the rank condition, implying a direct output of (at least) one state node in every contraction in $\mathcal{G}_A$ for system-output rank recovery. In this work, the global system graph associated with $W \otimes A$ is the Kronecker-product of $\mathcal{G}_A$ and $\mathcal{G}_\beta$. Recall that for $(W \otimes A, D_C)$-observability (or distributed observability) the global system state must be observable to every agent. Therefore, to satisfy condition (i), every state node needs to be connected via a directed path to every agent, which justifies strong-connectivity of $\mathcal{G}_\beta$. On the other hand, to satisfy condition (ii), the outputs from state nodes measured by all α-agents (including one node in every contraction) need to be directly shared among all agents to recover their system-output rank. This implies that for any α-agent j, we have $j \in \mathcal{N}_\alpha(i), \forall i \in \{1, \ldots, N\}$. This justifies the connectivity of $\mathcal{G}_\alpha$, and completes the proof. ■

With $\mathcal{G}_\beta$ and $\mathcal{G}_\alpha$ satisfying the conditions in Lemma 2, the block-diagonal gain matrix K can be designed such that $\rho(\widehat{A}) < 1$, i.e., $\widehat{A}$ is a Schur matrix. In fact, the gain matrix K is known to be the solution to the Linear-Matrix-Inequality (LMI) $X - \widehat{A}^\top X \widehat{A} \succ 0$ or equivalently,

$$\begin{pmatrix} X & \widehat{A}^\top X \\ X\widehat{A} & X \end{pmatrix} \succ 0, \tag{8}$$

for some $X \succ 0$ (where "$\succ$" denotes positive-definiteness). However, to satisfy the distributed condition, K needs to be further block-diagonal in order to satisfy information locality. Following [91], [93], iterative cone-complementarity optimization method is adopted to design the proper K matrix with polynomial-order complexity. Applying such K matrix, we have $\rho(\widehat{A}) < 1$, which implies stability and steady-state boundedness of the error in the attack-free case.

B. Performance Analysis in the Attack-Free Case

Next, we provide the performance analysis of the proposed distributed estimator (filter) (3)-(4) in the attack-free case. Following the same analogy as in [26]–[28], [87], we analyze the mean performance and mean-square performance of the protocol (3)-(4) for $\tau_k = \mathbf{0}$.

Lemma 3: Let $\mathbf{e}_\infty \triangleq \lim_{k\to\infty} \mathbf{e}_k$ denote the steady-state error of the proposed estimator (3)-(4). Then, $\mathbb{E}(\mathbf{e}_\infty) = \mathbf{0}$.

Proof: Taking expectation of the error dynamics (5),

$$\mathbb{E}(\mathbf{e}_k) = \widehat{A}\mathbb{E}(\mathbf{e}_{k-1}) + \mathbb{E}(\eta_k). \tag{9}$$

Recall from Section III-A that $\rho(\widehat{A}) < 1$ and following from [26], [87], it is clear that the first term in (9) vanishes asymptotically. Then, from (6) in the attack-free case ($\tau_k = \mathbf{0}$),

$$\begin{aligned}\mathbb{E}(\mathbf{e}_\infty) &= \mathbb{E}(\eta_\infty) \\ &= \mathbf{1}_N \otimes \mathbb{E}(\nu_\infty) - KD_C(\mathbf{1}_N \otimes \mathbb{E}(\nu_\infty)) - K\overline{D}_C\mathbb{E}(\zeta_\infty).\end{aligned}$$

Recall from Section II-A that $\mathbb{E}(\nu_k) = \mathbf{0}$ and $\mathbb{E}(\zeta_k) = \mathbf{0}$. This implies that $\mathbb{E}(\mathbf{e}_\infty) = \mathbf{0}$ and the lemma follows. ■

Lemma 4: Define $Q_k := \mathbb{E}(\mathbf{e}_k\mathbf{e}_k^\top)$ and $\Phi := \mathbb{E}(\eta_k\eta_k^\top)$. Let $Q_\infty = \lim_{k\to\infty} Q_k$ denote the collective error covariance at the steady-state. For error dynamics (5) in the attack-free case,

$$\|Q_\infty\|_2 \le \frac{a_1 N\|E\|_2 + a_2\|\overline{R}\|_2}{1 - b^2}, \tag{10}$$

with $a_1 \triangleq \|I_{Nn} - KD_C\|_2^2$, and $a_2 \triangleq \|K\|_2^2$, $\overline{R} \triangleq \text{blockdiag}\,[\sum_{j\in\mathcal{N}_\alpha(i)} \mathbf{c}_j R_{jj} \mathbf{c}_j^\top]$.

Proof: Following [87] with $\|\widehat{A}\|_2 \triangleq b$,

$$\|Q_\infty\|_2 \le \frac{\|\Phi\|_2}{1 - b^2}. \tag{11}$$

From (6) we have,

$$\begin{aligned}\eta_k\eta_k^\top = (I_{Nn} - KD_C)(\mathbf{1}_{NN} \otimes \nu_{k-1}\nu_{k-1}^\top)(I_{Nn} - KD_C)^\top \\ + (K\overline{D}_C)\zeta_k\zeta_k^\top(K\overline{D}_C)^\top.\end{aligned} \tag{12}$$

Then, from (6),

$$\begin{aligned}\|\Phi\|_2 \le \|(I_{Nn} - KD_C)(\mathbf{1}_{NN} \otimes E)(I_{Nn} - KD_C)^\top\|_2 \\ + \|(K\overline{D}_C)R(K\overline{D}_C)^\top\|_2.\end{aligned}$$

Using the fact that $\|\mathbf{1}_{NN} \otimes E\|_2 = N\|E\|_2$,

$$\|\Phi\|_2 \le \|I_{Nn} - KD_C\|_2^2\ N\|E\|_2 + \|K\|_2^2\|\overline{R}\|_2, \tag{13}$$

and applying (11) results in (10). ■

In fact, Lemma 3 implies that the estimator (3)-(4) is unbiased in the absence of attacks, while Lemma 4 states that its mean-square estimation error (also known as mean-square deviation [27]) is bounded in steady-state.

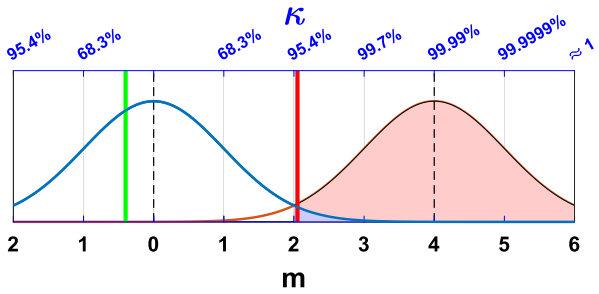


Fig. 3. This figure illustrates the attack detection logic in Lemma 5. The confidence intervals for the normalized residual in the absence of attack (blue curve) are shown. Each value of m in (17) associated with a confidence interval represents a probability threshold κ associated with the Gaussian PDF of the residual. As an example, the red and green lines represent two normalized residual values $\frac{r_k^i}{\Theta_2^i}$ via (14) and (17). Following the binary hypothesis testing (maximum-likelihood case), the threshold on the residual is the intersection (midpoint) of the two PDFs, where the residual belongs to the PDF in the presence of attack (red curve). Since the residual is over the threshold θ_κ with $m = 2$ ($r_k > 2\Theta_2$), probability of attack is more than $\kappa = 95.4\%$. This probability is equal to the red shaded area (since both PDFs follow the same normal distribution), while the probability of false alarm ($1 - \kappa = 4.6\%$) is shaded by blue. Clearly, this gives the highest probability of detection, while for higher threshold values (larger κ) the residual is not detected as biased/attacked. For the green residual with $|r_k| < \Theta_2$, the residual is most likely due to (system/measurement) noise, which is also evident from the (blue) PDF. Recall that, in general, m may take (positive) real values over infinite range ($m \to \infty$).

IV. Main Algorithm

We now describe the attack detection logic. Define the residual at every agent i as the absolute difference value between the original output y_k^i and the estimated output,

$$r_k^i \triangleq |y_k^i - \widehat{y}_k^i| = |\mathbf{c}_i^\top \widehat{A}_i \mathbf{e}_{k-1} + \mathbf{c}_i^\top \eta_k^i + \zeta_k^i + \tau_k^i|. \tag{14}$$

Note that the residual defined above based on the absolute-value is a standard definition, which is irrespective of the attack being positive ($\tau_k^i > 0$) or negative ($\tau_k^i < 0$) and works for both sign-preserving and sign-changing attacks. As shown in Lemmas 3 and 4, in the attack-free case with $\tau_k^i = 0$, the estimation error $\mathbf{e}_k^i$, and therefore, the residual r_k^i is bounded steady-state stable and unbiased at all agents. Note that in general $\widehat{A}_i \mathbf{e}_{k-1} \to 0$ due to Schur stability of $\widehat{A}$, while the second term in (14) is,

$$\mathbf{c}_i^\top \eta_k^i = \mathbf{c}_i^\top \nu_{k-1} - \mathbf{c}_i^\top K_i \sum_{j \in \mathcal{N}_\alpha(i)} \Big(\mathbf{c}_j \zeta_k^j + \mathbf{c}_j \tau_k^j + \mathbf{c}_j \mathbf{c}_j^\top \nu_{k-1} \Big). \tag{15}$$

In case of an attack on agent i, i.e., $\tau_k^i \neq 0$, the term $\mathbf{c}_i^\top \eta_k^i$ is biased at agent i. This biased residual can be used to find (isolate) the attacked agent. In this sense, first, we need to define a threshold on the residuals to distinguish the effect of noise terms (in absence of attacks) and the biasing attacks.

A. Probabilistic Threshold Design

Here, the probabilistic detection thresholds are defined based on Q_∞ in (10). For each agent define,

$$\frac{\|Q_\infty\|_2}{N} \leq \frac{a_1 N \|E\|_2 + a_2 \|\overline{R}\|_2}{N(1-b^2)} =: \Theta_1. \tag{16}$$

TABLE II
Different Threshold Probabilities κ for Integer m in (17)

$\frac{m}{2}$	1	2	3	4
Threshold probability κ	68.3%	95.4%	99.7%	99.99%

Then, for specific false alarm rates and attack detection probabilities κ, one can consider different detection-levels $m \in \mathbb{R}_{>0}$ as described in Fig. 3. A detection-level m represents a specific probability threshold κ associated with the Gaussian PDF of the estimation error in the attack-free case. Then, the thresholds θ_κ are designed as follows.

Lemma 5: Following the assumptions in Section II-D, given the noise covariance R and E and the residuals r_k^i from (14), the attack detection threshold for a detection-level $m \in \mathbb{R}_{>0}$ is,

$$\theta_\kappa := m\Theta_2^i, \;\; \Theta_2^i := |\mathbf{c}_i^\top| \Theta_1 + R_{ii} \tag{17}$$

where $\kappa = \mathrm{erf}(\frac{m}{\sqrt{2}})$ is detection probability (with $\mathrm{erf}(\cdot)$ as the *Gauss error function*), $\mathbf{c}_i$ is the measurement column-vector at agent i, and Θ_1 follows (16).

Proof: The proof directly follows from Lemma 3 and 4 and the results in [87]. From Lemma 3 and 4, $\mathbb{E}(\mathbf{e}_k^i) = \mathbf{0}$ for attack-free case, and following the zero-mean Gaussian distribution of the noise terms in η_k (including ν_k and ζ_k) and linearity of the error dynamics (5)-(6) and the protocol (3)-(4), it is straightforward to see that $\mathbf{e}_k^i$ and r_k^i are Gaussian; see details in [87]. Then, from standard textbooks on Gaussian distribution (e.g., [94]) and (14) in attack-free case, the probability of $|r_k^i| \leq m\Theta_2^i$ with $\Theta_2^i = |\mathbf{c}_i^\top| \Theta_1 + R_{ii}$ is determined via the value of the normal deviate less than $m\Theta_2^i$, i.e., $\kappa = \mathrm{erf}(\frac{m}{\sqrt{2}})$. Recall that Θ_2^i is the residual variance and R_{ii} is the measurement noise variance at agent i. Then, in presence of attack, both error $\mathbf{e}_k^i$ and residual r_k^i are biased by some products of $\tau_k^i \neq 0$ (due to linearity). In this case, the residual follows a biased Gaussian distribution with non-zero mean. Following statistical hypothesis testing for the two Gaussian distributions with equal variance (assuming equally likely a-priori hypothesis), if the residual $|r_k^i|$ is greater than $m\Theta_2^i$ then the probability of attack is κ and probability of false alarm is $1 - \kappa$. This justifies the probability thresholds $\theta_\kappa \triangleq m\Theta_2^i$ (as illustrated in Fig. 3) and completes the proof. ■

The parameter m in (17) and Lemma 5 can take any real (or integer) value in $\mathbb{R}_{>0}$. Some typical threshold probability values κ for integer values of m are given in Table II. Clearly, higher values of m (and κ) implies lower false alarm rates.

Remark 5: A straightforward sequel to Lemma 5 is that one can design the threshold θ_κ for a given false-alarm rate $\varkappa = 1 - \kappa$ as $\theta_\kappa = \sqrt{2}\mathrm{erf}^{-1}(\kappa)\Theta_2^i$.

Remark 6: The magnitude of the residual r_k^i is tightly related to the magnitude of the biasing attack τ_k^i. In other words, greater measurement bias τ_k^i results in greater residual r_k^i exceeding the threshold θ_κ with higher attack probability κ and lower probability of false alarm $\varkappa = 1 - \kappa$.

Recall from Remark 1 that, unlike [55]–[59] considering a fixed (deterministic) threshold based on the upper bound on

ζ_k, (17) assigns probability κ to the threshold θ_κ with no such upper bound assumption on the noise terms, implying the probabilistic threshold design.

B. Attack Detection and Mitigation Logic

Recall that, following Lemma 2, the connectivity of the α, β, and γ-agents over $\mathcal{G}_\alpha$ and $\mathcal{G}_\beta$ results in the next lemma.

Lemma 6: Following the connectivity condition in Lemma 2 and residual formulations in (14)-(15),

i) In case of having no α-agent[2], attack at any β or γ-agent is isolated.

ii) For isolation of attack in presence of an α-agent j, the gain matrix K needs to satisfy,

$$\left|\frac{\mathbf{c}_i^\top K_i \mathbf{c}_j}{\mathbf{c}_j^\top K_j \mathbf{c}_j - 1}\right| \leq \epsilon, \text{ for } i \neq j, \tag{18}$$

where $0 \leq \epsilon < 1$ is a pre-specified constant determining the residual ratio.

Proof: From Lemma 2, in absence of any α-agent, $\mathcal{N}_\alpha(i) = \{i\}$ for any agent i of type β and γ. Thus, from (14)-(15), biasing attack $\tau_k^i \neq 0$ at a β or γ-agent i only affects the residual r_k^i. This implies that r_k^i is biased while r_k^j ($j \neq i$) is unbiased, implying that attack τ_k^i is isolated at any β/γ-agent. On the other hand, in the presence of an α-agent j subject to attack $\tau_k^j \neq 0$, (14)-(15) implies that the residual r_k^i at every agent i is affected by the attack at agent $j \in \mathcal{N}_\alpha(i)$ via the term $\mathbf{c}_i^\top K_i \mathbf{c}_j$, while the residual r_k^j at α-agent j is affected by the factor $\mathbf{c}_j^\top K_j \mathbf{c}_j - 1$. Therefore, (18) ensures that $|\frac{r_k^i}{r_k^j}| > \frac{1}{\epsilon} > 1$ (for $i \neq j$), implying greater residual at α-agent j by factor $\frac{1}{\epsilon}$. This constraint ensures that the attack can be isolated at every α-agent j. ■

Following Lemma 5 and 6, for the attacked agent i (of any type) the residual r_k^i is (more) biased over θ_κ in (17), while the residuals at other agents are less biased (or unbiased). Largest κ such that $r_k^i \geq \theta_\kappa$ declares the probability of attack (or probability of false alarm $1-\kappa$). Likewise, from Remark 5 and 6, the attack detection logic can be designed for a *given* false alarm rate $\varkappa_i$ (and probabilistic threshold θ_{κ_i}) at sensor i. Then, similar to the deterministic case, the following hypothesis testing locally declares "Attack" or "No-Attack" at sensor i (under certain false alarm rate $\varkappa_i$),

$$\text{If } \begin{cases} r_k^i \geq \theta_{\kappa_i} \\ r_k^i < \theta_{\kappa_i} \end{cases} \text{ Then } \begin{cases} \mathcal{H}_1^i : \text{Attack Detected} \\ \mathcal{H}_0^i : \text{No Attack} \end{cases} \tag{19}$$

Remark 7: A relevant concept is *nodal/local consistency* of measurement/prediction information (data) set at agent i and $j \in \mathcal{N}_\alpha(i) \cup \mathcal{N}_\beta(i)$ at every time k, denoted by $\mathcal{I}_k^i, \mathcal{I}_k^j$ [95]. Recall that nodal consistency checks the *statistical consistency* of $\mathcal{I}_k^i$ with the information $\mathcal{I}_{[k-T,k]}^i$ over a sliding time-window T, declaring that $\mathcal{I}_k^i$ is trustable or not. In this direction, one can track the information over such time-window T and apply, for example, a *chi-square detector* on the residuals over T [16] instead of instantaneous residuals (14). Local consistency, on the other hand, checks the *statistical consistency* of the common information (e.g., on the shared observable subspace) between $\mathcal{I}_{[k-T,k]}^i$ and received information $\mathcal{I}_{[k-T,k]}^j$, $j \in \mathcal{N}_\alpha(i) \cup \mathcal{N}_\beta(i)$, and declares if $\mathcal{I}_k^j$ is trustable or not. Note that for (necessary) α/β-agents, weak local consistencies imply certain loss of observability information and degradation of estimation performance.

[2] Number of α-agents is equal to the rank-deficiency of the system matrix A [85]. Therefore, for a full-rank system the associated distributed estimator has no α-agent [71].

Algorithm 1. Attack Detection and Mitigation

Input: System digraph $\mathcal{G}_A$ and its contractions $\mathcal{C}$ and parent SCCs $\mathcal{S}^p$, $\alpha/\beta/\gamma$ classification, local estimate $\widehat{\mathbf{x}}_{k|k}^i$ at every agent $i \in \{1, \ldots, N\}$.
Initialization: $\widehat{\mathbf{x}}_{0|0}^i$ is set randomly at all agents i
Every agent i does the following:
Finds the thresholds θ_κ based on (17);
Finds $\widehat{\mathbf{x}}_{k|k-1}^i$ and $\widehat{\mathbf{x}}_{k|k}^i$ for $k \geq 1$ via (3)-(4);
Finds the residual r_k^i via (14);
if $r_{k>k_a}^i > \theta_\kappa$ **then**
  Declares: attack detected with probability κ;
  if agent i is type α **then**
    Substitutes new agent i' with output from another state in the same contraction $\mathcal{C}_l$;
  else
    if agent i is type β **then**
      Substitutes new agent i' with output from another state in the same parent SCC $\mathcal{S}_l^p$;
    else
      Remove γ-agent i with no substitution;
Output Attack probability κ and substitute agent i';

Remark 8: (**Attack mitigation**) From Section II-B, α/β-agents are necessary for observability; therefore, in case of attacks, their erroneous information of their observable subsystems makes those subsystems unobservable to all agents, causing unstable estimation error. To recover the loss of observability, recall that the states in the same parent SCC $\mathcal{S}_l^p$ and in the same contraction $\mathcal{C}_l$ are *observationally-equivalent*, in the sense that measurement of two states in $\mathcal{S}_l^p$ or in $\mathcal{C}_l$ provide information on the same observable subsystem. In other words, the information $\mathcal{I}_k^i, \mathcal{I}_k^j$ offered by two state measurements (agents i, j) are said to be observationally-equivalent if they equally contribute to the rank recovery of the *observability Gramian* (see detailed definition in [42], [84]). In this regard, for attack mitigation, the biased measurement can be replaced with a new measurement of an observationally-equivalent state in $\mathcal{S}_l^p$ or $\mathcal{C}_l$. Note that, after mitigating the attacks, the performance analysis follows as in Section III-B.

Remark 9: (**Cost-optimal mitigation**) Given an observationally-equivalent set of state nodes $\mathcal{S}_l^p$ or $\mathcal{C}_l$, the substitute/replacement state measurement can be chosen based on its sensing cost. Combinatorial optimization strategies [96], e.g., the well-known *Hungarian* algorithm, can be adopted to find the minimal-cost equivalent measurement to reduce the overall sensing cost. Similar arguments hold for cost-optimal design

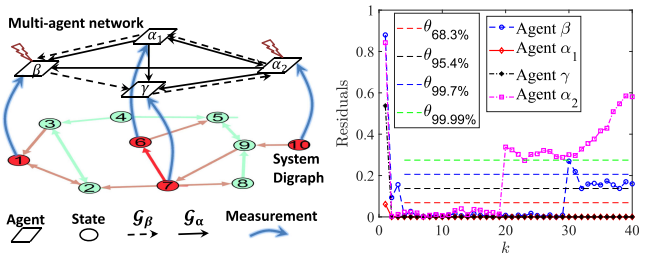


Fig. 4. (Left) The multi-agent network (top network including $\mathcal{G}_\alpha$ and $\mathcal{G}_\beta$) estimates the states of the dynamical system (bottom network) by taking output measurements of the states in red color. The agents α_2 and β are under attack. (Right) The residuals at 4 agents over time along with the thresholds θ_κ are shown. As expected, the residuals at the attacked agents are biased over the thresholds.

of the multi-agent network $\mathcal{G}_N = \mathcal{G}_\alpha \cup \mathcal{G}_\beta$, e.g., using the so-called *minimum spanning strong sub-graph* algorithm [97].

Remark 8 along with Lemma 5 and 6 result in Algorithm IV-B.

Note that the terms D_C in (5) and $\overline{R}$ in (16) are defined locally, i.e., the i-th diagonal block of D_C and $\overline{R}$ related to agent i are defined based on received measurement information $\mathbf{c}_j$ and R_j from its direct neighbors (summation is over $j \in \mathcal{N}_\alpha(i)$). Therefore, the calculations of these terms are distributed and localized over the network. The thresholds θ_κ in (17), agent types, and the sets of observationally-equivalent states in the system digraph $\mathcal{G}_A$ are determined by a central entity once off-line, then, broadcasted and transmitted to every agent. This procedure is done once and the information is stored at all agents; then, the agents can perform estimation and detect the attack locally with no further role of the centralized entity. See similar assumptions in [87], [91] for distributed estimation/filtering.

Remark 10: The DM (Dulmage-Mendelsohn) decomposition and DFS (depth-first-search) or Kosaraju-Sharir algorithms can be used, respectively, to find contractions and SCCs (along with their topological order) with computational complexity $\mathcal{O}(n^{2.5})$ and $\mathcal{O}(n^2)$ [98]. The residual calculation at agents is of $\mathcal{O}(n)$ complexity, while the complexity of the threshold design based on 2-norm calculation is $\mathcal{O}(N^3n^3)$. Overall, the complexity of Algorithm IV-B is $\mathcal{O}(N^3n^3)$. This polynomial order complexity suits large-scale applications.

V. Simulation

For simulation we consider a dynamical system with 10 states associated with the system digraph $\mathcal{G}_A$ in Fig. 4-(Left). The link weights in $\mathcal{G}_A$ are considered randomly (such that $\rho(A) > 1$). Following Remark 10, the contractions and parent SCCs in $\mathcal{G}_A$ are: $\mathcal{S}_1^p = \{1,2,3\}$, $\mathcal{C}_1 = \{5,6,8\}$, and $\mathcal{C}_2 = \{8,10\}$. From Section II-B, one output from each of these node sets ensure observability of $\mathcal{G}_A$. As shown in Fig. 4-(Left), agents β, α_1, and α_2 take output of state 1, 6, and 10, respectively, along with a redundant agent γ with output of state 7 (which is not necessary for observability). Following Section III, the network $\mathcal{G}_\beta$ is considered as a cycle, while in $\mathcal{G}_\alpha$ agents α_2 and α_1 are two hubs of the network. Each agent adopts the proposed protocol (3)-(4) to estimate all 10 system states (with partial observability via its measurement and neighboring information). The link weights in $\mathcal{G}_\beta$ (the nonzero W_{ij}s) are chosen randomly such that W is row-stochastic. The noise terms follow $\nu = \mathcal{N}(0, 0.01\mathbf{1}_{nn})$ and $\zeta = \mathcal{N}(0, 0.01 I_N)$. The block-diagonal gain K is determined via heuristic LMIs such that, for example: $|\mathbf{c}_\beta^\top K_\beta \mathbf{c}_{\alpha_1}| = 0.008$, $|\mathbf{c}_\gamma^\top K_\gamma \mathbf{c}_{\alpha_1}| = 0.00001$, $|\mathbf{c}_{\alpha_2}^\top K_{\alpha_2} \mathbf{c}_{\alpha_1}| = 0.005$, $|\mathbf{c}_{\alpha_1}^\top K_{\alpha_1} \mathbf{c}_{\alpha_1}| = 0.24$, satisfying Lemma 6 for any $0.011 \leq \epsilon < 1$ with j as agent α_1 in (18). Likewise, $0.01 \leq \epsilon < 1$ for agent α_2, implying that, for this given K, the attack-related portion of the residual at attacked agent α_2 is almost 100 times greater than the residuals at other (non-attacked) agents. Therefore, any attack at agents α_2, α_1 can be isolated. The parameters in (16) are $a_1 = 2.937$, $a_2 = 0.183$, $b = 0.682$, which result in $\Theta_1 = 0.068$ and $\Theta_2 = 0.078$. We consider fixed attack $\tau_{k\geq 30} = 1$ at agent β (following Assumption (iv)) along with an auto-regressive non-stationary attack for $k \geq 20$ at agent α_2 in the form $\tau_{k+2} = 2\tau_{k+1} - \tau_k + \vartheta$ with $\tau_{20} = \tau_{21} = 0.3$ and $\vartheta \in [0, 0.02]$ as a uniform random variable. The residuals (14) (shown in Fig. 4-(Right)) at the attacked agents β and α_2 are biased, respectively, over $\theta_{95.4\%}$ and $\theta_{99.99\%}$, implying false alarm probabilities[3] approximately less than 4.6% and 0.01%.

TABLE III
Parameter Values for the Detection and Estimation Protocol in [53], [54]

L	40	α	2	β	0.4
$\|\|A\|\|$	1.35	γ	0.57	N	4
s	2	b_w	0.06	b_v	0.06
λ_0	1.95	η_0	0.1	ρ_{t_0}	0.1

Comparison with recent literature: next, we use the estimation and detection strategy in [53], [54] for comparison. Recall that from Remark 4, the distributed observer in [53], [54] is a double time-scale protocol, which requires many iterations of consensus between every two time-steps of system dynamics. Therefore, it needs much faster information sharing/processing rate as compared to the proposed protocol (3)-(4). *The reason for choosing [53], [54] for comparison study is that double time-scale protocols make similar relaxed observability assumption as Assumption (ii) in Section II-D (irrespective of system rank-deficiency). This is in contrast to many exisitng single time-scale protocols, e.g., [26], [27], [31]–[35], which assume that the underlying system is observable in the neighborhood of each agent and/or is full-rank. In other words, the mentioned references generally require more network connectivity, and therefore, do not result in steady-state stable error over the given $\mathcal{G}_\alpha$ and $\mathcal{G}_\beta$ networks in Fig. 4-(Left).* We set the parameters in [53], [54] as in Table III (which seem to provide the best outcome).

In this simulation, agents need to perform $L = 40$ consensus iterations for estimation/detection, which requires 40-times faster communication and computation rate as compared to the proposed protocol (3)-(4). The results are shown in

[3] The auto-regressive attack is given as an example of possible extension of the results to the case of non-stationary attacks, where the attack probabilities can be approximated by Lemma 5.

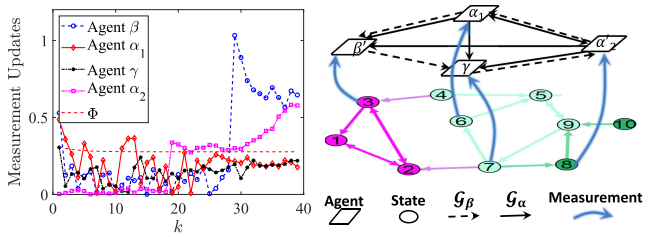


Fig. 5. (Left) This figure shows the measurement-updates at all agents based on the methodology in [53], [54]. The attack is detected via the threshold Φ. Clearly, the protocol in [53], [54] with parameters given in Table III detects both attacks at agents α_2 and β, while also raising false alarm on agent α_1 at some times. In contrast, our proposed detection strategy only raise alarm on the attacked agents as shown in Fig. 4-(Right). (Right) Using Algorithm IV-B, the detected attacks are mitigated by adding equivalent agents α_2' and β' to recover the loss of observability. The new agents α_2' and β' measure observationally-equivalent states, respectively, in the same contraction (green nodes) and in the same parent SCC (purple nodes).

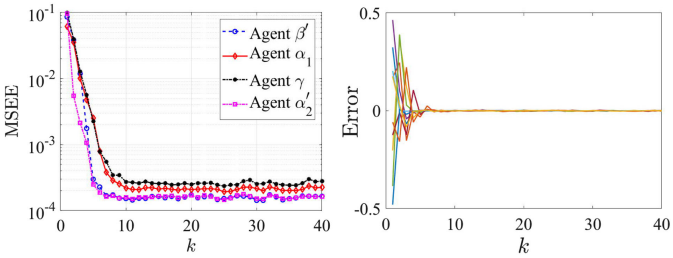


Fig. 6. (Left) This figure presents Monte-Carlo time-evolution of the MSEEs (in log-scale) at 4 agents after attack mitigation. The bounded steady-state MSEEs imply observable estimation/filtering. (Right) This figure shows the (Monte-Carlo) time-evolution of the estimation errors of all 10 states at agent β', which are unbiased in steady-state.

Fig. 5-(Left).Following the attack detection logic in [53], [54], the agents can detect possible attacks if their *measurement-updates* are over a certain threshold Φ. From Fig. 5-(Left), both attacks are detected, while also falsely alarming attack at agent α_1 at some times.

Attack mitigation and performance analysis: next, using the mitigation strategy in Algorithm IV-B, we replace the detected attacked agents β and α_2 with substitute agents β' and α_2', respectively measuring observationally-equivalent state 3 in $\mathcal{S}_1^p$ and state 8 in $\mathcal{C}_2$. The connectivity of the new agents follows the same connectivity of $\mathcal{G}_\alpha$ and $\mathcal{G}_\beta$ as shown in Fig. 5-(Right).We perform Monte-Carlo simulation (averaged over 100 repetitions) of the proposed protocol (3)-(4) for the attack-mitigated case of Fig. 5-(Right). The mean-square performance and mean performance are shown in Fig. 6. As it is clear, the mean-square estimation errors (MSEEs) are bounded steady-state stable at all agents as expected from Lemma 4. Further, from Lemma 3, the steady-state errors at all agents are unbiased; Fig. 6-(Right) shows unbiased state errors at agent β' as an example.

VI. Conclusion

This paper considers a decentralized attack detection over distributed estimation networks. The detection, isolation, and mitigation strategy is designed specifically for α, β, and γ-agents in polynomial-order complexity. As future research direction, network reconfiguration [11], [99] to reduce attack vulnerability and design of attack-tolerant/ resilient engineered networks is promising. Further, one can track the history of residuals (for general rank-deficient systems) over a sliding time-window (known as *stateful detection* [36]), similar to χ^2-detection [16] or trust-index evolution [95].

References

[1] S. Asefi, Y. Madhwal, Y. Yanovich, and E. Gryazina, "Application of blockchain for secure data transmission in distributed state estimation," 2021, *arXiv:2104.04232*.

[2] U. A. Khan and M. Doostmohammadian, "A sensor placement and network design paradigm for future smart grids," in *Proc. 4th Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process.*, San Juan, Puerto Rico, 2011, pp. 137–140.

[3] W. Yang, W. Luo, and X. Zhang, "Distributed secure state estimation under stochastic linear attacks," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 3, pp. 2036–2047, 1 Jul.–Sep. 2021.

[4] M. Doostmohammadian and U. A. Khan, "Topology design in networked estimation: A generic approach," in *Proc. Amer. Control Conf.*, Washington, DC, 2013, pp. 4140–4145.

[5] M. Doostmohammadian, H. R. Rabiee, and U. A. Khan, "Cyber-social systems: Modeling, inference, and optimal design," *IEEE Syst. J.*, vol. 14, no. 1, pp. 73–83, Mar. 2020.

[6] S. Xu, R. C. De Lamare, and H. V. Poor, "Distributed estimation over sensor networks based on distributed conjugate gradient strategies," *IET Signal Process.*, vol. 10, no. 3, pp. 291–301, 2016.

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

[8] H. Shiri, M. A. Tinati, M. Codreanu, and G. Azarnia, "Distributed sparse diffusion estimation with reduced communication cost," *IET Signal Process.*, vol. 12, no. 8, pp. 1043–1052, 2018.

[9] M. Doostmohammadian and U. A. Khan, "Vulnerability of CPS inference to DoS attacks," in *Proc. 48th IEEE Asilomar Conf. Signals, Syst., Comput.*, 2014, pp. 2015–2018.

[10] Y. Chen, S. Kar, and J. M. F. Moura, "The Internet of Things: Secure distributed inference," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 64–75, Sep. 2018.

[11] M. Doostmohammadian and H. R. Rabiee, "On the observability and controllability of large-scale IoT networks: Reducing number of unmatched nodes via link addition," *IEEE Contr. Syst. Lett.*, vol. 5, no. 5, pp. 1747–1752, Nov. 2021.

[12] O. J. Pandey, V. Gautam, H. H. Nguyen, M. K. Shukla, and R. M. Hegde, "Fault-resilient distributed detection and estimation over a SW-WSN using LCMV beamforming," *IEEE Trans. Netw. Service Manag.*, vol. 17, no. 3, pp. 1758–1773, Sep. 2020.

[13] Y. Guo, T. Ji, Q. Wang, L. Yu, G. Min, and P. Li, "Unsupervised anomaly detection in IoT systems for smart cities," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 4, pp. 2231–2242, Oct.–Dec. 2020.

[14] S. Pequito, S. Kar, and A. P. Aguiar, "Minimum number of information gatherers to ensure full observability of a dynamic social network: A structural systems approach," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, 2014, pp. 750–753.

[15] M. Doostmohammadian, H. R. Rabiee, and U. A. Khan, "Centrality-based epidemic control in complex social networks," *Social Netw. Anal. Mining*, vol. 10, pp. 1–11, 2020.

[16] M. Doostmohammadian, T. Charalambous, M. Shafie-khah, N. Meskin, and U. A. Khan, "Simultaneous distributed estimation and attack detection/isolation in social networks: Structural observability, Kronecker-product network, and chi-square detector," in *Proc. 1st IEEE Int. Conf. Auton. Syst*, 2021, *arXiv:2105.10639*.

[17] M. Dehghani, A. Kavousi-Fard, M. Dabbaghjamanesh, and O. Avatefipour, "Deep learning based method for false data injection attack detection in AC smart Islands," *IET Gener., Transmiss. Distrib.*, vol. 14, no. 24, pp. 5756–5765, 2020.

[18] S. Cui, Z. Han, S. Kar, T. T. Kim, H. V. Poor, and A. Tajer, "Coordinated data-injection attack and detection in the smart grid: A detailed look at enriching detection solutions," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 106–115, Sep. 2012.

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

[20] E. Drayer and T. Routtenberg, "Detection of false data injection attacks in smart grids based on graph signal processing," *IEEE Syst. J.*, vol. 14, no. 2, pp. 1886–1896, Jun. 2020.

[21] T. Chakravorti, R. K. Patnaik, and P. K. Dash, "Detection and classification of islanding and power quality disturbances in microgrid using hybrid signal processing and data mining techniques," *IET Signal Process.*, vol. 12, no. 1, pp. 82–94, 2017.

[22] X. Luo, X. Wang, X.and Pan, and X. Guan, "Detection and isolation of false data injection attack for smart grids via unknown input observers," *IET Gener., Transmiss. Distrib.*, vol. 13, no. 8, pp. 1277–1286, 2019.

[23] R. Babu and B. Bhattacharyya, "Optimal allocation of phasor measurement unit for full observability of the connected power network," *Int. J. Elect. Power Energy Syst.*, vol. 79, pp. 89–97, 2016.

[24] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Trans. Inf. System Secur.*, vol. 14, no. 1, pp. 1–33, 2011.

[25] J. Chen, W. Li, C. Wen, J. Teng, and P. Ting, "Efficient identification method for power line outages in the smart power grid," *IEEE Trans. Power Syst.*, vol. 29, no. 4, pp. 1788–1800, Jul. 2014.

[26] U. A. Khan and J. M. F. Moura, "Distributing the Kalman filter for large-scale systems," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4919–4935, Oct. 2008.

[27] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, "Diffusion strategies for distributed kalman filtering: Formulation and performance analysis," *Proc. Cogn. Inf. Process.*, 2008, pp. 36–41.

[28] R. Olfati-Saber and P. Jalalkamali, "Collaborative target tracking using distributed Kalman filtering on mobile sensor networks," in *Proc. Amer. Control Conf.*, San Francisco, CA, 2011, pp. 1100–1105.

[29] M. Deghat, V. Ugrinovskii, I. Shames, and C. Langbort, "Detection and mitigation of biasing attacks on distributed estimation networks," *Automatica*, vol. 99, pp. 369–381, 2019.

[30] J. Miloševič, T. Tanaka, H. Sandberg, and K. H. Johansson, "Analysis and mitigation of bias injection attacks against a Kalman filter," IFAC-*Papers OnLine*, vol. 50, no. 1, pp. 8393–8398, 2017.

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

[32] Y. Chen, S. Kar, and J. M. F. Moura, "Resilient distributed estimation: Sensor attacks," *IEEE Trans. Autom. Control*, vol. 64, no. 9, pp. 3772–3779, Sep. 2019.

[33] G. Battistelli, L. Chisci, G. Mugnai, A. Farina, and A. Graziano, "Consensus-based algorithms for distributed filtering," in *Proc. 51st IEEE Conf. Decis. Control*, 2012, pp. 794–799.

[34] S. Tu and A. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6217–6234, Dec. 2012.

[35] S. Park and N. Martins, "Necessary and sufficient conditions for the stabilizability of a class of LTI distributed observers," in *Proc. 51st IEEE Conf. Decis. Control*, 2012, pp. 7431–7436.

[36] J. Giraldo *et al.*, "A survey of physics-based attack detection in cyber-physical systems," *ACM Comput. Surv.*, vol. 51, no. 4, pp. 1–36, 2018.

[37] Y. Guan and X. Ge, "Distributed attack detection and secure estimation of networked cyber-physical systems against false data injection attacks and jamming attacks," *IEEE Trans. Signal Inf. Process. Over Netw.*, vol. 4, no. 1, pp. 48–59, Mar. 2018.

[38] U. A. Khan, S. Kar, A. Jadbabaie, and J. M. F. Moura, "On connectivity, observability, and stability in distributed estimation," in *Proc. 49th IEEE Conf. Decis. Control*, 2010, pp. 6639–6644.

[39] M. Doostmohammadian and U. Khan, "On the genericity properties in distributed estimation: Topology design and sensor placement," *IEEE J. Sel. Top. Signal Process.*, vol. 7, no. 2, pp. 195–204, Apr. 2013.

[40] J. M. Dion, C. Commault, and J. van der Woude, "Generic properties and control of linear structured systems: A survey," *Automatica*, vol. 39, pp. 1125–1144, Mar. 2003.

[41] M. Doostmohammadian and U. Khan, "Graph-theoretic distributed inference in social networks," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 4, pp. 613–623, Aug. 2014.

[42] M. Doostmohammadian and U. A. Khan, "Measurement partitioning and observational equivalence in state estimation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2016, pp. 4855–4859.

[43] S. S. Pereira, R. Lòpez-Valcarce, and A. Pagès-Zamora, "A diffusion-based EM algorithm for distributed estimation in unreliable sensor networks," *IEEE Signal Process. Lett.*, vol. 20, no. 6, pp. 595–598, Jun. 2013.

[44] M. Doostmohammadian, H. R. Rabiee, H. Zarrabi, and U. A. Khan, "Distributed estimation recovery under sensor failure," *IEEE Signal Process. Lett.*, vol. 24, no. 10, pp. 1532–1536, Oct. 2017.

[45] R. G. Dutta, T. Zhang, and Y. Jin, "Resilient distributed filter for state estimation of cyber-physical systems under attack," in *Proc. Amer. Control Conf.*, 2019, pp. 5141–5147.

[46] A. Mitra, J. Richards, S. Bagchi, and S. Sundaram, "Resilient distributed state estimation with mobile agents: Overcoming byzantine adversaries, communication losses, and intermittent measurements," *Auton. Robots*, vol. 43, no. 3, pp. 743–768, 2019.

[47] A. Mustafa and H. Modares, "Secure event-triggered distributed kalman filters for state estimation," 2019, *arXiv:1901.06746*.

[48] F. Wen and Z. Wang, "Distributed Kalman filtering for robust state estimation over wireless sensor networks under malicious cyber attacks," *Digit. Signal Process.*, vol. 78, pp. 92–97, 2018.

[49] Z. Yang, A. Gang, and W. Bajwa, "Adversary-resilient distributed and decentralized statistical inference and machine learning: An overview of recent advances under the byzantine threat model," *IEEE Signal Proc. Mag.*, vol. 37, no. 3, pp. 146–159, May 2020.

[50] L. Su and S. Shahrampour, "Finite-time guarantees for byzantine-resilient distributed state estimation with noisy measurements," *IEEE Trans. Autom. Control*, vol. 65, no. 9, pp. 3758-3771, Sep. 2020.

[51] Q. Li, B. Shen, Z. Wang, and F. E. Alsaadi, "A sampled-data approach to distributed h∞ resilient state estimation for a class of nonlinear time-delay systems over sensor networks," *J. Franklin Inst.*, vol. 354, no. 15, pp. 7139–7157, 2017.

[52] X. Wang and E. Yaz, "Stochastically resilient extended Kalman filtering for discrete-time nonlinear systems with sensor failures," *Int. J. Syst. Sci.*, vol. 45, no. 7, pp. 1393–1401, 2014.

[53] X. He, X. Ren, H. Sandberg, and K. H. Johansson, "Secure distributed filtering for unstable dynamics under compromised observations," in *Proc. IEEE 58th Conf. Decis. Control*, 2019, pp. 5344–5349.

[54] X. He, X. Ren, H. Sandberg, and K. H. Johansson, "How to secure distributed filters under sensor attacks?," *IEEE Trans. Autom. Control*, 2021, to be published, doi: 10.1109/TAC.2021.3092603

[55] J. Kim, C. Lee, H. Shim, Y. Eun, and J. H. Seo, "Detection of sensor attack and resilient state estimation for uniformly observable nonlinear systems having redundant sensors," *IEEE Trans. Autom. Control*, vol. 64, no. 3, pp. 1162–1169, Mar. 2019.

[56] M. Pajic, P. Tabuada, I. Lee, and G. J. Pappas, "Attack-resilient state estimation in the presence of noise," in *Proc. 54th IEEE Conf. Decis. Control*, 2015, pp. 5827–5832.

[57] M. S. Chong, M. Wakaiki, and J. P. Hespanha, "Observability of linear systems under adversarial attacks," in *Proc. Amer. Control Conf.*, 2015, pp. 2439–2444.

[58] C. Lee, H. Shim, and Y. Eun, "Secure and robust state estimation under sensor attacks, measurement noises, and process disturbances: Observer-based combinatorial approach," in *Proc. Eur. Control Conf.*, 2015, pp. 1872–1877.

[59] 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 Trans. Autom. Control*, vol. 62, no. 10, pp. 4917–4932, Oct. 2017.

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

[61] P. Wang, M. Govindarasu, A. Ashok, S. Sridhar, and D. McKinnon, "Data-driven anomaly detection for power system generation control," in *Proc. IEEE Int. Conf. Data Mining Workshops*, 2017, pp. 1082–1089.

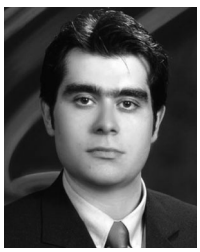
[62] W. Hashlamoun, S. Brahma, and P. K. Varshney, "Mitigation of byzantine attacks on distributed detection systems using audit bits," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 1, pp. 18–32, Mar. 2018.

[63] P. Chen, Y. S. Han, H. Lin, and P. K. Varshney, "Optimal byzantine attack for distributed inference with m-ary quantized data," in *Proc. IEEE Int. Symp. Inf. Theory*, 2016, pp. 2474–2478.

[64] F. Rosas, J. Hsiao, and K. Chen, "A technological perspective on information cascades via social learning," *IEEE Access*, vol. 5, pp. 22605–22633, 2017.

[65] E. Soltanmohammadi, M. Orooji, and M. Naraghi-Pour, "Decentralized hypothesis testing in wireless sensor networks in the presence of misbehaving nodes," *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 1, pp. 205–215, Jan. 2013.
[66] B. Kailkhura, Y. S. Han. Brahma, and P. K. Varshney, "Asymptotic analysis of distributed Bayesian detection with byzantine data," *IEEE Signal Process. Lett.*, vol. 22, no. 5, pp. 608–612, May 2015.
[67] X. Zheng, L. Xie, and H. Chen, "Steady-state performance analysis of consensus-based distributed detection under sensing data falsification attack," in *Proc. 9th Int. Conf. Wireless Commun. Signal Process.*, 2017, pp. 1–6.
[68] M. Doostmohammadian and U. A. Khan, "On the characterization of distributed observability from first principles," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, 2014, pp. 914–917.
[69] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Syst. Mag.*, vol. 35, no. 1, pp. 93–109, Feb. 2015.
[70] B. Satchidanandan and P. R. Kumar, "Dynamic watermarking: Active defense of networked cyber-physical systems," *Proc. IEEE*, vol. 105, no. 2, pp. 219–240, Feb. 2017.
[71] M. Doostmohammadian and N. Meskin, "Sensor fault detection and isolation via networked estimation: Full-rank dynamical systems," *IEEE Control Netw. Syst.*, vol. 8, no. 2, pp. 987–996, Jun. 2021.
[72] Y. Chen, S. Kar, and J. M. F. Moura, "Resilient distributed estimation through adversary detection," *IEEE Trans. Signal Process.*, vol. 66, no. 9, pp. 2455–2469, May 2018.
[73] G. Joseph and C. R. Murthy, "On the observability of a linear system with a sparse initial state," *IEEE Signal Process. Lett.*, vol. 25, no. 7, pp. 994–998, Jul. 2018.
[74] M. B. Wakin, B. M. Sanandaji, and T. L. Vincent, "On the observability of linear systems from random, compressive measurements," in *Proc. 49th IEEE Conf. Decis. Control*, 2010, pp. 4447–4454.
[75] L. Li and D. Li, "A distributed estimation method over network based on compressed sensing," *Int. J. Distrib. Sensor Netw.*, vol. 15, no. 4, 2019, Art. no. 1550147719841496.
[76] M. Majidi, M. Etezadi-Amoli, and H. Livani, "Distribution system state estimation using compressive sensing," *Int. J. Elect. Power Energy Syst.*, vol. 88, pp. 175–186, 2017.
[77] R. J. Hamidi, H. Khodabandelou, H. Livani, and M. Sami-Fadali, "Hybrid state estimation using distributed compressive sensing," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2016, pp. 1–5.
[78] S. Xu, R. C. De Lamare, and H. V. Poor, "Distributed compressed estimation based on compressive sensing," *IEEE Signal Process. Lett.*, vol. 22, no. 9, pp. 1311–1315, Sep. 2015.
[79] P. Agarwal, M. Tamer, and H. Budman, "Assessing observability using supervised autoencoders with application to tennessee eastman process," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 206–211, 2020.
[80] C. Wang, S. Tindemans, K. Pan, and P. Palensky, "Detection of false data injection attacks using the autoencoder approach," in *Proc. Int. Conf. Probabilistic Methods Appl. Power Syst.*, 2020, pp. 1–6.
[81] J. Wang, D. Shi, Y. Li, J. Chen, H. Ding, and X. Duan, "Distributed framework for detecting PMU data manipulation attacks with deep autoencoders," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 4401–4410, Jul. 2019.
[82] D. Wilson, Y. Tang, J. Yan, and Z. Lu, "Deep learning-aided cyber-attack detection in power transmission systems," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2018, pp. 1–5.
[83] M. Khodayar, G. Liu, J. Wang, and M. E. Khodayar, "Deep learning in power systems research: A review," *CSEE J. Power Energy Syst.*, vol. 7, no. 2, pp. 209–220, 2021.
[84] M. Doostmohammadian, H. R. Rabiee, H. Zarrabi, and U. Khan, "Observational equivalence in system estimation: Contractions in complex networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 5, no. 3, pp. 212–224, Jul.–Sep. 2018.
[85] M. Doostmohammadian and U. A. Khan, "On the distributed estimation of rank-deficient dynamical systems: A generic approach," in *Proc. 38th Int. Conf. Acoust., Speech, Signal Process.*, Vancouver, CA, 2013, pp. 4618–4622.
[86] M. Doostmohammadian and U. A. Khan, "Communication strategies to ensure generic networked observability in multi-agent systems," in *Proc. 45th Annu. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, 2011, pp. 1865–1868.
[87] U. A. Khan and A. Jadbabaie, "Collaborative scalar-gain estimators for potentially unstable social dynamics with limited communication," *Automatica*, vol. 50, no. 7, pp. 1909–1914, 2014.
[88] Y. Y. Liu, J. J. Slotine, and A. L. Barabási, "Observability of complex systems," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 7, pp. 2460–2465, 2013.
[89] T. Charalambous and C. N. Hadjicostis, "Distributed formation of balanced and bistochastic weighted digraphs in multi-agent systems," in *Proc. Eur. Control Conf.*, 2013, pp. 1752–1757.
[90] J. Bay, *Fundamentals of Linear State Space Systems*. New York, NY, USA: McGraw-Hill, 1999.
[91] U. A. Khan and A. Jadbabaie, "Coordinated networked estimation strategies using structured systems theory," in *Proc. 49th IEEE Conf. Decis. Control*, 2011, pp. 2112–2117.
[92] M. Doostmohammadian and U. A. Khan, "Minimal sufficient conditions for structural observability/controllability of composite networks via Kronecker product," *IEEE Trans. Signal Inf. Process. over Netw.*, vol. 6, pp. 78–87, 2020.
[93] L. El Ghaoui, F. Oustry, and M. Ait Rami, "A cone complementarity linearization algorithm for static output-feedback and related problems," *IEEE Trans. Autom. Control*, vol. 42, no. 8, pp. 1171–1176, Aug. 1997.
[94] K. Krishnamoorthy, *Handbook of Statistical Distributions With Applications*. Boca Raton, FL, USA: CRC Press, 2016.
[95] U. Khan and A. Stankovic, "Secure distributed estimation in cyber-physical systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2013, pp. 5209–5213.
[96] M. Doostmohammadian and U. A. Khan, "On the complexity of minimum-cost networked estimation of self-damped dynamical systems," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 3, pp. 1891–1900, Jul.–Sep. 2020.
[97] M. Doostmohammadian, H. R. Rabiee, and U. A. Khan, "Structural cost-optimal design of sensor networks for distributed estimation," *IEEE Signal Proc. Lett.*, vol. 25, no. 6, pp. 793–797, Jun. 2018.
[98] K. Murota, *Matrices and Matroids for Systems Analysis*. Berlin, Germany: Springer, 2000.
[99] M. Doostmohammadian, T. Charalambous, M. Shafie-khah, H. R. Rabiee, and U. A. Khan, "Analysis of contractions in system graphs: Application to state estimation," in *Proc. 1st IEEE Int. Conf. Auton. Syst.*, 2021.

Mohammadreza Doostmohammadian (Member, IEEE) received the B.Sc. and M.Sc. degrees in mechanical engineering from the Sharif University of Technology (SUT), Tehran, Iran, and the Ph.D. degree in electrical engineering from Tufts University, Medford, MA, USA. He was a Postdoc with the AICT, School of Computer Engineering, SUT and a Researcher with ITRC. He is currently an Assistant Professor of mechatronics with Semnan University, Semnan, Iran, and a Researcher with Aalto University, Espoo, Finland. His general research interests include distributed optimization, control, and estimation over networks. Recognition of his work includes IEEE JSTSP journal cover and IEEE MSC09 and ICNSC14 conference awards.



Houman Zarrabi (Member, IEEE) received the Ph.D. degree from Concordia University, Montreal, QC, Canada, in 2011. Since then, he has been involved in various industrial and research projects. His main expertise includes IoT, M2M, CPS, Big Data, embedded systems, and VLSI. He is currently the national IoT program Director and an Assistant Professor with Iran Telecommunication Research Center (ITRC).

Hamid R. Rabiee (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from CSULB, and the EEE and Ph.D. degrees in electrical and computer engineering from USC, in 1993, and from Purdue University, West Lafayette, IN, USA, in 1996. He was with AT&T Bell Laboratories, Intel Corporation, as a Senior Software Engineer, and with PSU, OGI, and OSU as an Adjunct Professor. He was also a Visiting Professor with Imperial College London, London, U.K., for the 2017–2018 academic year. He is the Founder of AICT, SATI, DML, VASL, BCB, and Cognitive Neuroengineering Research Center. He is currently a Professor of computer engineering with SUT.

Usman A. Khan (Senior Member, IEEE) received the B.S. degree in electrical and computer engineering from the University of Engineering & Technology Lahore, Lahore, Pakistan, the M.S. degree in electrical and computer engineering from the University of Wisconsin-Madison, Madison, WI, USA, and the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA. He held a Postdoc position with GRASP Laboratory, UPenn. He was a Visiting Professor with KTH, and currently is an Associate Professor of electrical and computer engineering (ECE) with Tufts University, Medford, MA, USA, where he is also an Adjunct Professor of computer science. Recognition of his work includes the prestigious NSF Career Award, several NSF REU awards, an IEEE journal cover, three IEEE best student paper awards.

Themistoklis Charalambous (Senior Member, IEEE) received the B.A. and M.Eng. in electrical and information sciences from Trinity College, Cambridge University, Cambridge, U.K., and the Ph.D. degree from Control Laboratory, Engineering Department, Cambridge University. He joined Human Robotics Group, as a Research Associate with Imperial College London, London, U.K., for an academic year and was a Visiting Lecturer with the Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus. He was a Postdoc with the Department of Automatic Control of the School of Electrical Engineering, KTH and Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden. Since September 2018, he has been nominated Research Fellow of the Academy of Finland and since July 2020 he has been a tenured Associate Professor of electrical engineering with Aalto University, Espoo, Finland.