# Secure State Estimation with Byzantine Sensors: A Probabilistic Approach

Xiaoqiang Ren[1], Yilin Mo[2], Jie Chen[3], and Karl H. Johansson[4]

***Abstract*—This paper studies static state estimation in multi-sensor settings, with a caveat that an unknown subset of the sensors are compromised by an adversary, whose measurements can be manipulated arbitrarily. The attacker is able to compromise $q$ out of $m$ sensors. A new performance metric, which quantifies the asymptotic decay rate for the probability of having an estimation error larger than $\delta$, is proposed. We develop an optimal estimator for the new performance metric with a fixed $\delta$, which is the Chebyshev center of a union of ellipsoids. We further provide an estimator that is optimal for every $\delta$, for the special case where the sensors are homogeneous. Numerical examples are given to elaborate the results.**

***Index Terms*—Security, Secure estimation, Byzantine attacks, Large deviation**

## I. Introduction

In cyber-physical systems, numerous sensors with limited capacity are spatially deployed and connected via ubiquitous wired and wireless communication networks. This makes it nearly impossible to guarantee the security of every single sensor or communication channel. Therefore, security problems of cyber-physical systems have attracted much attention recently, e.g., [1], [2].

Robust estimation has been studied over decades to deal with uncertainties of input data [3]–[5]. The robustness is usually measured by influence functions or breakdown point, and several celebrated estimators have been developed, such as M-, L-, and R-estimators. The limitation of this robustness theory is the assumption that the bad data are independent [5], which, however, is not the case in general for cyber attacks. The fact that compromised sensors may cooperate and the estimation is done sequentially makes the "bad" data correlated both spatially and temporarily.

1: School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, China. xqren@shu.edu.cn. This work was written when he was with School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Sweden.

2: Department of Automation and BNRist, Tsinghua University, China. ylmo@tsinghua.edu.cn. Corresponding Author

3: Department of Electronic Engineering, City University of Hong Kong, Hong Kong, China. jichen@cityu.edu.hk

4: School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden. kallej@kth.se

The work of X. Ren is funded in part by Shanghai Key Laboratory of Power Station Automation Technology, and by National Key R&D Program of China (No. 2018AAA0102800, No. 2018AAA0102804). The work of Y. Mo is supported by the National Key Research and Development Program of China under Grant 2018AAA0101601. The work of J. Chen is supported in part by Hong Kong RGC under Project CityU 11200415, and in part by City University of Hong Kong under the project 7004866. The work of K. H. Johansson is supported in part by the Knut and Alice Wallenberg Foundation, the Swedish Strategic Research Foundation, and by the Swedish Research Council.

Recently, dynamic state estimation with some Byzantine sensors has been discussed. Most approaches in the existing literature can be classified into two categories: stacked measurements [6]–[8] and Kalman filter decomposition [9], [10]. Fawzi *et al.* [6] used the stacked measurements from time $k$ to $k+T-1$ to estimate the state at time $k$ and provided $l_0$ and $l_1$-based state estimation procedures. Since deterministic systems are concerned, the $l_0$-based procedure can exactly recover the state. Pajic *et al.* [7] extended the deterministic systems in [6] to ones with bounded measurement noises and obtained upper bounds of estimation error for both $l_0$ and $l_1$-based estimators. Mishar *et al.* [8] studied stochastic systems with unbounded noises and proposed a notion of $\epsilon$-effective attack. The state estimation there is in essence an attack detection problem; a Chi-squared test is applied to the residues and the standard Kalman filter output based on the measurements from the largest set of sensors that are deemed $\epsilon$-effective attack-free is used as the state estimate. Notice that to detect the $\epsilon$-effective attack-free sensors correctly with high probability, the window size $T$ must be large enough. The authors did not provide estimators before detection decisions are made. The authors of [9], [10] used local estimators at each sensor and proposed a LASSO based fusion scheme. However, their approach imposes some strong constraints on the system dynamics. Furthermore, the estimate error of the proposed algorithm when there are indeed attacks is not specifically characterized.

In this paper, we deal with scenarios where noises are not necessarily bounded and give a different characterization of the estimator performance, i.e., the decaying rate of the worst-case probability that the estimation error is larger than some value $\delta$ rather than the worst-case error in [7], [9], [10] and estimation error covariance in [8]. This is partially motivated by the following three observations. Firstly, with unbounded noise, the worst-case estimation error might result in too conservative system designs. Notice also that even for the bounded noise cases studied in [7], the upper bound of the worst-case estimation error thereof increases with respect to (w.r.t.) the window size $T$, which counters intuition since more information should lead to better estimation accuracy. Secondly, to mitigate the bad effects caused by Byzantine sensors, one has to accumulate much enough information, i.e., the time window $T$ should be large enough. In this case, the decaying rate is able to characterize the probability well enough (just as, e.g., [8]). Lastly, the system operator may pre-define the error threshold $\delta$ according to the performance specification, which leads to a more flexible system design.

In the subsequent sections, we focus on the problem of

secure static state estimation with Byzantine sensors. A fusion center aims to estimate a vector state $x \in \mathbb{R}^n$ from measurements collected by $m$ sensors, among which $q$ sensors might be compromised. Without imposing any restrictions on the attacker's capabilities, we assume that the compromised sensors can send arbitrary messages. Static state estimation has a wide range of applications in power system, where the power network states (i.e., bus voltage phase angles and bus voltage magnitudes) are estimated from measurements collected by Supervisory Control And Data Acquisition (SCADA) systems (e.g., transmission line power flows, bus power injections, and part of the bus voltages) through remote terminal units (RTUs) [11], [12]. Considering the possibility that the RTUs are controlled and the communicated data from SCADA systems tampered with by malicious attackers, much work has devoted to security problems of power systems, e.g. [13]–[16]. The closest literature is [17], [18], which, however, both focused on the one-shot scenario, while in this work the observations are taken sequentially, the possible temporal correlations of which make the analysis more challenging. We should also note that both [17], [18] used the worst-case estimate error as the performance metric rather than the probabilistic approach in this paper. Moreover, the main results of this work provide fundamental insights on the counterpart for dynamical systems that we are still investigating.

The main contributions of this work are summarized as follows.

1) We propose a new metric to characterize the performance of an estimator when observation noise is not necessarily bounded and an attacker may be present.
2) We provide an optimal estimator for a given estimation error threshold $\delta$ (Theorem 2), which is the Chebyshev center of a union of ellipsoids. We then propose an algorithm to compute the optimal estimator (Algorithm 1 and Theorem 3).
3) When the sensors are homogeneous, we further provide a uniformly optimal estimator, i.e., simultaneously optimal for any error threshold $\delta$ (Theorem 4). The estimator is just the "trimmed mean" of the averaged observations.

A preliminary version of this paper was presented in [19]. The main difference is threefold. Firstly, new results have been provided in this paper, i.e., numerical implementation of our algorithm (Section III-C) and uniformly optimal estimator design (Section IV). Secondly, in [19], only proofs of Lemmas 8 and 9 were presented due to page limitation. Lastly, new simulations have been conducted in this paper for better illustration.

*Organization:* In Section II, we formulate the problem of static state estimation with Byzantine sensors, including the attack model and performance metric. The main results are presented in Section III. We first prove that one may only consider estimators with certain "nice" structures. Based on this, we then provide an optimal estimator for a given error threshold and propose an algorithm to compute the optimal estimator. Furthermore, a very simple yet uniformly optimal estimator when sensors are homogeneous is provided in Section IV. After showing numerical examples in Section V, we conclude the paper in Section VI. All proofs are reported in the appendix.

*Notations*: $\mathbb{R}$ ($\mathbb{R}_+$) is the set of (nonnegative) real numbers. $\mathbb{N}$ ($\mathbb{N}_+$) is the set of nonnegative (positive) integers. For a vector $x \in \mathbb{R}^n$, define $\|x\|_0$ as the "zero norm", i.e., the number of nonzero elements of the vector $x$. For a vector $x \in \mathbb{R}^n$, the support of $x$, denoted by $\mathrm{supp}(x)$, is the set of indices of nonzero elements:

$$\mathrm{supp}(x) \triangleq \{i \in \{1, 2, \ldots, n\} : x_i \neq 0\}.$$

Define $\mathbf{1}$ as the column vector of ones and the size clear from the context if without further notice. Let $\boldsymbol{I}_n$ be the identity matrix of size $n \times n$. For a matrix $\boldsymbol{M} \in \mathbb{R}^{m\times n}$, unless stated otherwise, $\boldsymbol{M}_i$ represents the $i$-th row, and $\boldsymbol{M}_{\mathcal{I}}$ the matrix obtained from $\boldsymbol{M}$ after removing all of the rows except those in the index set $\mathcal{I}$. For a set of matrices $\mathcal{A} \subseteq \mathbb{R}^{m\times n}$, we use $\mathcal{A}_{\mathcal{I}}$ to denote the set of matrices keeping rows indexed by $\mathcal{I}$, i.e.,

$$\mathcal{A}_{\mathcal{I}} \triangleq \{\boldsymbol{M}_{\mathcal{I}} : \boldsymbol{M} \in \mathcal{A}\}.$$

For a set $\mathcal{A}$, define the indicator function as $\mathbb{1}_{\mathcal{A}}(x) = 1$, if $x \in \mathcal{A}$; 0 otherwise. The cardinality of a set $\mathcal{A}$ is denoted as $|\mathcal{A}|$. Let $\boldsymbol{M}^\top$ denote the transpose of the matrix $\boldsymbol{M}$. We write $\boldsymbol{M} \succcurlyeq \boldsymbol{N}$ if $\boldsymbol{M} - \boldsymbol{N}$ is a positive semi-definite matrix.

## II. Problem Formulation

### A. System Model

Consider the problem of estimating the state $x \in \mathbb{R}^n$ using $m$ sensor measurements as depicted in Fig. 1. Let $\mathcal{M} \triangleq \{1, \ldots, m\}$ be the index set of all the sensors. The measurement equation for sensor $i \in \mathcal{M}$ is

$$z_i(k) = H_i x + w_i(k),$$

where $z_i(k) \in \mathbb{R}$ is the ("true") measurement collected by the sensor $i$ at time $k \in \mathbb{N}_+$, $H_i \in \mathbb{R}^{1\times n}$ is the output matrix associated with sensor $i$, $w_i(k) \in \mathbb{R}$ is the observation noise. It is assumed that $w_i(k)$ is Gaussian distributed with zero mean and variance $\mathbb{E}[(w_i(k))^2] = W_i > 0$ for any $i, k$[1]. Furthermore, $w_i(k)$ are independent across the sensors and over time, i.e., $\mathbb{E}[w_{i_1}(k_1)w_{i_2}(k_2)] = 0$ if $i_1 \neq i_2$ or $k_1 \neq k_2$.

In the presence of attacks, the measurement received by the fusion center is $y_i(k)$, with satisfies the following equation:

$$y_i(k) = z_i(k) + a_i(k),$$

where $a_i(k) \in \mathbb{R}$ is the bias injected by the attacker.

We assume the attacks are $q$-sparse:

**Assumption 1** ($q$-sparse attack)**.** *There exists an index set $\mathcal{C} \subseteq \mathcal{M}$ such that*

1) *for any sensor $i \in \mathcal{M} \setminus \mathcal{C}$, $a_i(k) = 0$ for any time $k$.*
2) *$|\mathcal{C}| = q$.*

[1]Actually, the main results in this paper hold for any noise distribution in the exponential family; the details are discussed in Remark 2

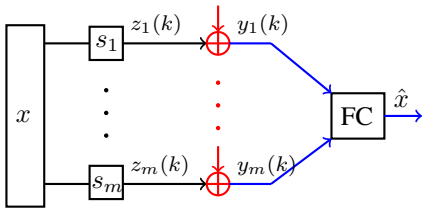



Fig. 1: The fusion center (FC) estimates the underlying state $x$ using sensor measurements that might be manipulated.

The sparse attack model, which is conventional in the literature [6]–[10], [18], [20], [21], says that the set of compromised sensors is somewhat "constant" over time. This is in essence the only restriction we impose on the attacker's capability. The bias $a_i(k)$ of a compromised sensor may take any value and might be correlated across sensors and over time. If the set of compromised sensors is time-varying, the estimators (or detectors) in *all* the aforementioned literature will be destroyed. That is, the estimators (or detectors) could not work at all or the error could be arbitrarily large. In this paper, without this constant property, even Lemma 1 provided later (in particular, e.g., (34) and (35)), which is the basis for Theorems 1 and 2, would not hold.

**Assumption 2** (System knowledge)**.** *The system designer knows the number $q$, but does not know the exact set of compromised sensors $\mathcal{C}$.*

The quantity $q$ might be determined by the *a priori* knowledge about the quality of each sensor. Alternatively, the quantity $q$ may be viewed as a design parameter, which indicates the resilience level that the designer is willing to pay for. One finds more comments about the above assumption in Remark 3.

Let $\boldsymbol{H} = [H_1^\top, H_2^\top, \dots, H_m^\top]^\top$ be the measurement matrix. We assume that the matrix $\boldsymbol{H}$ is $2q$-observable:

**Assumption 3.** *The measurement matrix $\boldsymbol{H}$ is $2q$-observable, i.e., for every set $\mathcal{I} \subseteq \mathcal{M}$ with $|\mathcal{I}| = m - 2q$, the matrix $\boldsymbol{H}_\mathcal{I}$ is of full column rank.*

It has been shown in [6] that $2q$-observability of the measurement matrix is a necessary and sufficient condition to recover the exact state under $q$-sparse attacks when there are no observation noises. One finds the results if Assumption 3 is violated in Lemma 3 later. Notice that in power systems, measurement redundancy is a common practice [12].

To introduce the knowledge available at the attacker, we need the following definitions. Define the measurement from all sensors at time $k$ to be a column vector:

$$\boldsymbol{y}(k) \triangleq \begin{bmatrix} y_1(k) & y_2(k) & \dots & y_m(k) \end{bmatrix}^\top \in \mathbb{R}^m. \tag{1}$$

We further define $\boldsymbol{Y}(k)$ as a matrix of all measurements from time 1 to time $k$:

$$\boldsymbol{Y}(k) \triangleq \begin{bmatrix} \boldsymbol{y}(1) & \boldsymbol{y}(2) & \dots & \boldsymbol{y}(k) \end{bmatrix} \in \mathbb{R}^{m \times k}. \tag{2}$$

The quantities $\boldsymbol{a}(k), \boldsymbol{A}(k)$ are defined in the same manner. At time $k$, given measurements from all the sensors $\boldsymbol{Y}(k)$, the fusion center generates a state estimate $\hat{x}_k$. The estimator $f$ might be random, i.e., given $\boldsymbol{Y}(k)$, $\hat{x}_k$ is a random variable governed by certain probability measure on $\mathbb{R}^n$ determined by $f$.

**Assumption 4** (Attacker's knowledge)**.** *It is assumed that*

1) *the attacker knows the true state $x$;*
2) *the attacker knows the estimator $f$, the system parameters (i.e., each $H_i$ and $W_i$), and can access the historical and current observations from the compromised sensors.*

The above assumption as a whole has been adopted in literature on sparse attack, e.g., [20]–[22], while the second bullet prevails in literature on data-injection attack, e.g., [8], [13], [23]. The parameters $H_i$ and $W_i$ might be developed by an attacker using the *a priori* knowledge of the underlying physical model. To obtain the true state, the attacker may deploy its own sensor network. Though it might be difficult in practice to obtain the accurate parameters and true state for an attacker, this assumption is de facto when dealing with potential worst-case attacks. We should note that this assumption is in accordance with the Kerckhoffs's principle [24], namely the security of a system should not rely on its obscurity. Interested readers are referred to [25] to see more attack models in cyber-physical systems. This assumption is leveraged later to define the performance metric in (3) and characterize the attack capacity in Theorems 1 and 2. In particular, one finds more on how Assumptions 1 and 4 are utilized to derive (34) in Remark 6 later.

### B. Performance Metric

At time $k$, given the measurements $\boldsymbol{Y}(k)_\mathcal{C}$, the bias $\boldsymbol{A}(k-1)$, the set of compromised sensors $\mathcal{C}$, and true state $x$, the bias $\boldsymbol{a}(k)$ is generated according to some probability measure on $\mathbb{R}^m$. This bias injection mechanism is denoted by $g$. Let $\mathcal{G}$ be the set of all attack strategies such that the generated bias $\boldsymbol{a}(k)$ satisfies the $q$-sparse attack model in Assumption 1.

In this paper, we are concerned with the worst-case scenario. Given an estimator $f$, we define

$$e(f,k,\delta) \triangleq \sup_{\mathcal{C}\subseteq\mathcal{M}, g\in\mathcal{G}, x\in\mathbb{R}^n} \mathbb{P}_{f,g,x,\mathcal{C}}\left(\|\hat{x}_k - x\|_2 > \delta\right) \tag{3}$$

as the worst-case probability that the distance between the estimate at time $k$ and the true state is larger than a certain value $\delta \in \mathbb{R}_+$ considering all possible attack strategies, the set of compromised sensors and the true state. We use $\mathbb{P}_{f,g,x,\mathcal{C}}$ to denote the probability measure governing $\hat{x}_k$ when the estimator $f$, attack strategy $g$, the true state $x$, and the set of compromised sensors $\mathcal{C}$ are given.

Ideally, one wants to design an estimator $f$ such that $e(f,k,\delta)$ is minimized at any time $k$ for any $\delta$. However, it is quite difficult to analyze $e(f,k,\delta)$ when $k$ takes finite values since computing the probability of error usually involves numerical integration. Therefore, we consider an asymptotic estimation performance, i.e., the exponential rate with which the worst-case probability goes to zero:

$$r(f,\delta) \triangleq \liminf_{k\to\infty} -\frac{\log e(f,k,\delta)}{k}. \tag{4}$$

Obviously, for any $\delta$, the system designer would like to maximize $r(f,\delta)$ by choosing a suitable estimator $f$.

The threshold $\delta$ is chosen by the designer in accordance with system accuracy requirement by noticing that a true state $x$ is perceived as the same with any point $x'$ lying inside its neighbourhood, i.e., $\|x' - x\|_2 \leq \delta$ by the above performance metric. However, in some cases (see Section IV), there is no need to determine $\delta$ since one can find an estimator that simultaneously maximizes $r(f, \delta)$ for all $\delta$.

### C. Problems of Interest

The following three problems are to be addressed.

1) *Performance limit.* For any $\delta$, what is the maximal rate $r(f, \delta)$ that can be achieved by all possible estimators?
2) *Optimal estimator.* Given $\delta$, what is the optimal estimator that maximizes $r(f, \delta)$?
3) *Uniform optimality.* Is there an estimator that simultaneously maximizes $r(f, \delta)$ for all $\delta > 0$?

## III. Optimal Estimator

In this section, the first two problems in Section II-C shall be addressed. We provide an estimator based on Chebyshev centers, prove its optimality, and further present a numerical algorithm to implement it.

### A. Compressed and Deterministic Estimator

A generic estimator $f_k$ might *randomly* generate an estimate $\hat{x}_k$ based on *all* the information contained in $\boldsymbol{Y}(k)$. In other words, given $\boldsymbol{Y}(k)$, the estimate $\hat{x}_k$ might be a random variable; and if any element (totally there are $m \times k$) of two observation matrices, say $\boldsymbol{Y}(k)$ and $\boldsymbol{Y}'(k)$, is different, the corresponding probability distributions of the estimate $\hat{x}_k$ might be different. In this subsection, however, we shall show that, without loss of optimality, one may only consider estimators with certain "nice" structure (i.e., the compressed and deterministic estimators defined in Definition 3 later).

Define an operator $\mathrm{avg}(\cdot)$ that averages each row of the inputed real-valued matrix, i.e., for any matrix $\boldsymbol{M} \in \mathbb{R}^{n_1 \times n_2}$,

$$\mathrm{avg}(\boldsymbol{M}) \triangleq \boldsymbol{M}\mathbf{1}/n_2.$$

Hence, $\mathrm{avg}(\boldsymbol{Y}(k))$ is a vector in $\mathbb{R}^m$ and the $i$-th element is the empirical mean of the observation from time 1 to $k$ available for sensor $i$.

We use $\mathbb{P}_f(\hat{x}_k|\boldsymbol{Y}(k))$ to denote the conditional probability measure of estimate $\hat{x}_k$ given any estimator $f$ and the information $\boldsymbol{Y}(k)$. Notice that an estimator $f$ can be completely characterized by the sequence of conditional probability measures from time 1 to $\infty$: $(\mathbb{P}_f(\hat{x}_1|\boldsymbol{Y}(1)), \mathbb{P}_f(\hat{x}_2|\boldsymbol{Y}(2)), \ldots)$.

**Definition 1.** *An estimator $f$ is said to be compressed if at each time $k$, it only utilizes the averaged information $\mathrm{avg}(\boldsymbol{Y}(k))$ to generate estimate $\hat{x}_k$, i.e., the conditional probability measures satisfy*

$$\mathbb{P}_f(\hat{x}_k \in \mathcal{A}|\boldsymbol{Y}(k)) = \mathbb{P}_f(\hat{x}_k \in \mathcal{A}|\boldsymbol{Y}'(k)) \tag{5}$$

*for any Borel set $\mathcal{A} \subseteq \mathbb{R}^n$ whenever $\mathrm{avg}(\boldsymbol{Y}(k)) = \mathrm{avg}(\boldsymbol{Y}'(k))$.*

Let $\mathcal{F}$ ($\mathcal{F}_\mathrm{c}$, resp.) be the set of all possible (compressed, resp.) estimators. In the following lemma, we show that it suffices to consider an estimator in $\mathcal{F}_\mathrm{c}$.

**Lemma 1.** *For any estimator $f \in \mathcal{F}$, there exists another compressed estimator $f' \in \mathcal{F}_\mathrm{c}$ such that for all $\delta > 0$,*

$$e(f', k, \delta) \leq e(f, k, \delta), \quad k = 1, 2, \ldots$$

*Proof.* See Appendix A. □

**Remark 1.** *Intuitively, only measurements from benign sensors provide "useful information" needed to estimate the underlying state, while under the most harmful attack, compromised sensors will merely generate disturbing noises. In our case, the averaged information $\mathrm{avg}(\mathbf{Y}(k))$ can* fully *summarize the information contained in measurements from benign sensors due to the fact that $\mathrm{avg}(\mathbf{Y}(k))$ is a sufficient statistic for the underlying state $x$ when there is no attacker. Therefore, it suffices to consider a compressed estimator that only utilizes the averaged information each time. This might be counterintuitive as one expects that with more information, i.e., using raw data $\mathbf{Y}(k)$, the compromised sensors could be detected more easily and, thus, better performance could be achieved. This, however, is not the case.*

**Remark 2.** *Lemma 1 says that $\mathrm{avg}(\mathbf{Y}(k))$ is a sufficient statistic for the underlying state $x$ whether or not the attacker is present. In fact, one may verify, using the same idea in Appendix A, in particular, the construction technique in* (30)*, that Lemma 1 holds if the distribution of $w_i(k)$ is in the exponential family and not necessarily Gaussian as we assume. This is mainly due to the fact that, if the distribution of a one-shot observation is in the exponential family, the sufficient statistic of a set of i.i.d. observations is simply the sum of individual sufficient statistics, the size of which will not increase as data accumulate.*

In the following, we refine the set $\mathcal{F}$ from another perspective.

**Definition 2.** *An estimator $f$ is said to be deterministic w.r.t $\boldsymbol{Y}(k)$ if for every time $k$ and observations $\boldsymbol{Y}(k)$, the estimate $f(\boldsymbol{Y}(k))$ is a single point in $\mathbb{R}^n$.*

Let $\mathcal{F}_\mathrm{d}$ be the set of all estimators that are deterministic w.r.t. $\boldsymbol{Y}(k)$. Then similar to the above lemma we have

**Lemma 2.** *For any estimator $f \in \mathcal{F}$, there exists another deterministic one $f' \in \mathcal{F}_\mathrm{d}$ such that for all $\delta > 0$*

$$r(f', \delta) \geq r(f, \delta).$$

*Proof.* See Appendix B. □

Based on the above two lemmas, we further refine $\mathcal{F}$.

**Definition 3.** *An estimator $f$ is said to be compressed and deterministic if it is deterministic w.r.t. $\mathrm{avg}(\boldsymbol{Y}(k))$, i.e., there exists a sequences of functions $\{\tilde{f}_k\}_{k=1,2,\ldots}$ with $\tilde{f}_k : \mathbb{R}^m \to \mathbb{R}^n$ such that the estimate at each time $k$*

$$f(\boldsymbol{Y}(k)) = \tilde{f}_k(\mathrm{avg}(\boldsymbol{Y}(k))).$$

Let $\mathcal{F}_{\rm cd}$ be the set of all compressed and deterministic estimators. Obviously, $\mathcal{F}_{\rm cd} \subseteq \mathcal{F}_{\rm c}, \mathcal{F}_{\rm cd} \subseteq \mathcal{F}_{\rm d}$. In the following theorem, we show that instead of $\mathcal{F}$, one may only consider the set $\mathcal{F}_{\rm cd}$ for our problem.

**Theorem 1.** *For any estimator $f \in \mathcal{F}$, there exists another compressed and deterministic estimator $f' \in \mathcal{F}_{\rm cd}$ such that*

$$r(f', \delta) \geq r(f, \delta), \quad \forall \delta > 0.$$

*Proof.* See Appendix C. □

### B. Optimal Estimator Based on Chebyshev Centers

In this subsection, we propose an optimal compressed and deterministic estimator. To this end, we need the following definitions: The distance of a point $x_0 \in \mathbb{R}^n$ to a bounded and non-empty set $\mathcal{A} \subseteq \mathbb{R}^n$ is defined as

$$\operatorname{dist}(x_0, \mathcal{A}) \triangleq \sup\{\|x - x_0\|_2 : x \in \mathcal{A}\}.$$

Moreover, the set's radius $\operatorname{rad}(\mathcal{A}) \in \mathbb{R}_+$ and Chebyshev center $\operatorname{chv}(\mathcal{A}) \in \mathbb{R}^n$ are defined by

$$\operatorname{rad}(\mathcal{A}) \triangleq \min_{x_0 \in \mathbb{R}^n} \operatorname{dist}(x_0, \mathcal{A}), \tag{6}$$

$$\operatorname{chv}(\mathcal{A}) \triangleq \arg\min_{x_0 \in \mathbb{R}^n} \operatorname{dist}(x_0, \mathcal{A}). \tag{7}$$

Notice that the Chebyshev center exists and is unique, since $\mathbb{R}^n$ is uniformly convex and $\mathcal{A}$ is bounded [26, Part 5, §33].

Given $y \in \mathbb{R}^m, x \in \mathbb{R}^n$, define their inconsistency $d_x(y)$ as the optimal value of the following optimization problem:

$$\begin{array}{ll} \underset{a \in \mathbb{R}^m}{\text{minimize}} & \dfrac{1}{2}\displaystyle\sum_{i=1}^{m}(y_i - H_i x + a_i)^2/W_i \\ \text{subject to} & \|a\|_0 \leq q. \end{array} \tag{8}$$

Further define the set $\mathcal{X}(y, \phi), \phi \geq 0$ as the set of $x$ such that the inconsistency with $y$ is upper bounded by $\phi$, i.e.,

$$\mathcal{X}(y, \phi) \triangleq \{x \in \mathbb{R}^n : d_x(y) \leq \phi\}. \tag{9}$$

Given $\delta \geq 0$, define $\mathbb{X}(y, \delta)$ as the biggest $\mathcal{X}(y, \phi)$ of which the radius is upper bounded by $\delta$:

$$\mathbb{X}(y, \delta) \triangleq \bigcup_{\operatorname{rad}(\mathcal{X}(y,\phi)) \leq \delta,\, \phi \geq 0} \mathcal{X}(y, \phi). \tag{10}$$

It is easy to see that $\mathcal{X}(y, \phi)$ is monotonically increasing w.r.t. $\phi$. As a result, its radius is also increasing. Notice also that given $y$, the radius $\operatorname{rad}(\mathcal{X}(y, \phi))$ is right-continuous with respect to $\phi$ (see details in Lemma 6 later). Therefore, it might happen that $\operatorname{rad}(\mathbb{X}(y, \delta)) < \delta$ for certain $\delta$, while in most cases $\operatorname{rad}(\mathbb{X}(y, \delta)) = \delta$ is achieved. Let $f_\delta^*$ be the estimator such that the estimate at time $k$ is the Chebyshev center of $\mathbb{X}(\operatorname{avg}(\boldsymbol{Y}(k)), \delta)$, i.e.,

$$f_\delta^*\left(\boldsymbol{Y}(k)\right) = \operatorname{chv}\left(\mathbb{X}(\operatorname{avg}(\boldsymbol{Y}(k)), \delta)\right). \tag{11}$$

For $y \in \mathbb{R}^m$ and $\delta > 0$, we define $u(y, \delta)$ as the upper bound of the inconsistency between $y$ and the elements in $\mathbb{X}(y, \delta)$:

$$u(y, \delta) \triangleq \sup_{x \in \mathbb{X}(y,\delta)} d_x(y). \tag{12}$$

With a slight abuse of notation, we define $u(\delta)$ as the lower bound of $u(y, \delta)$:

$$u(\delta) \triangleq \inf_{y \in \mathbb{R}^m} u(y, \delta). \tag{13}$$

We have our first main result about the estimator (11).

**Theorem 2.** *Given any $\delta > 0$, the estimator $f_\delta^*$ in* (11) *is optimal in the sense that it maximizes the rate* (4)*, i.e., for any estimator $f \in \mathcal{F}$,*

$$r(f, \delta) \leq r(f_\delta^*, \delta) = u(\delta). \tag{14}$$

*Proof.* See Appendix D. □

**Remark 3.** *Notice that our estimator involves $q$, as is the case in [8], where the estimator (i.e., Algorithm 2 thereof) depends on the perceived number of compromised sensors (or its upper bound) as well. On the contrary, estimators in [6], [9] do not. In practice, the number of actually compromised sensors, $q_0$, might be smaller or larger than the design parameter $q$. If $q_0 < q$, the performance of our estimator is lower bounded by $u(\delta)$ in* (13)*. The details are as follows. With a little abuse of notation, in this remark, we use $d_x(y, q)$ (instead of $d_x(y)$) to denote the optimal value of optimization problem in* (8)*, and rewrite $r(f_\delta^*, \delta)$ as $r_q(f_\delta^*, \delta)$. Then the performance of our estimator when the number of compromised sensors is $q_0 < q$ is:*

$$r_{q_0}(f_\delta^*, \delta) = \inf_{y \in \mathbb{R}^m} \sup_{x \in \mathbb{X}(y,\delta)} d_x(y, q_0) \geq u(\delta). \tag{15}$$

*We should admit that it is challenging to design an estimator that balances decently $r_q(f_\delta^*, \delta)$ and $r_{q_0}(f_\delta^*, \delta)$ in our case. Interested readers are referred to our previous work [21], where an detector that achieves the "best" trade-off among performances with different $q$'s in the binary hypothesis testing case was provided. While if $q_0 > q$, our estimator will be destroyed, i.e., $r(f_\delta^*, \delta) = 0$, as is the case in [8]. This is not desirable in practice. Our future work will investigate estimators independent of $q$.*

In the following lemma, we consider the case where Assumption 3 is violated.

**Lemma 3.** *If Assumption 3 is violated, the followings holds:*

1) *For any $\delta > 0$, there exists $y^*, x_1, x_2$ (dependent on $\delta$) such that $d_{x_1}(y^*) = d_{x_2}(y^*) = 0$ and $\|x_1 - x_2\|_2 > \delta$;*
2) *$r(f, \delta) = r(f_\delta^*, \delta) = 0$.*

*Proof.* See Appendix E □

The above first bullet yields that for any $\delta > 0$, there exists $y$ (dependent on $\delta$) such that $\mathbb{X}(y, \delta)$ is empty.

### C. Numerical Implementation

In this subsection, we provide an algorithm to compute the estimator $f_\delta^*$ proposed above. We shall first propose a method to compute the Chebyshev center and the radius of $\mathcal{X}(y, \phi)$ for a given $\phi$. This shares a similar spirit with [27]. We then consider how to derive the appropriate $\phi$ using a modified bisection method. To proceed, we need the following definition and lemmas.

A variation of $d_x(y)$, where the support of $a$ in the definition in (8) is given *a priori*, is defined as follows:

**Definition 4.** *Given $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, and index set $\mathcal{I} \subseteq \mathcal{M}$, the restricted inconsistency $d_x(y, \mathcal{I})$ is*

$$d_x(y, \mathcal{I}) \triangleq \frac{1}{2} \sum_{i \in \mathcal{I}} (y_i - H_i x)^2 / W_i. \tag{16}$$

It is clear that with a fixed set $\mathcal{I}$, $d_x(y, \mathcal{I})$ is continuous w.r.t. both $x$ and $y$. Furthermore,

$$d_x(y) = \min_{\mathcal{I} \subseteq \mathcal{M},\, |\mathcal{I}| = m-q} d_x(y, \mathcal{I}).$$

**Lemma 4.** *When $|\mathcal{I}| \geq m - 2q$, the restricted inconsistency $d_x(y, \mathcal{I})$ can be equivalently written as:*

$$d_x(y, \mathcal{I}) = (x - \kappa_{\mathcal{I}} y_{\mathcal{I}})^\top \operatorname{var}(\mathcal{I})(x - \kappa_{\mathcal{I}} y_{\mathcal{I}}) + \operatorname{res}(\mathcal{I}) \tag{17}$$

*where the "variance"*

$$\operatorname{var}(\mathcal{I}) = \frac{1}{2} \boldsymbol{H}_{\mathcal{I}}^\top \boldsymbol{W}_{\{\mathcal{I}\}}^{-1} \boldsymbol{H}_{\mathcal{I}} \tag{18}$$

*and the "residue"*

$$\operatorname{res}(\mathcal{I}) = \frac{1}{2} (y_{\mathcal{I}} - \boldsymbol{H}_{\mathcal{I}} \kappa_{\mathcal{I}} y_{\mathcal{I}})^\top \boldsymbol{W}_{\{\mathcal{I}\}}^{-1} (y_{\mathcal{I}} - \boldsymbol{H}_{\mathcal{I}} \kappa_{\mathcal{I}} y_{\mathcal{I}}) \tag{19}$$

*with $\boldsymbol{W}_{\{\mathcal{I}\}}$ (different from $\boldsymbol{W}_{\mathcal{I}}$) being the square matrix obtained from $\boldsymbol{W} = \operatorname{diag}(W_1, W_2, \ldots, W_m)$ after removing all of the rows and columns except those in the index set $\mathcal{I}$, and*

$$\kappa_{\mathcal{I}} = (\boldsymbol{H}_{\mathcal{I}}^\top \boldsymbol{W}_{\{\mathcal{I}\}}^{-1} \boldsymbol{H}_{\mathcal{I}})^{-1} \boldsymbol{H}_{\mathcal{I}}^\top \boldsymbol{W}_{\{\mathcal{I}\}}^{-1}. \tag{20}$$

*Proof.* See Appendix F. □

In the following, we show that computing the Chebyshev center and radius of the set $\mathcal{X}(y, \phi)$ introduced in (9) can be transferred to a convex optimization problem. Notice that one can rewrite $\mathcal{X}(y, \phi)$ as:

$$\mathcal{X}(y, \phi) = \bigcup \mathcal{X}(y, \phi, \mathcal{I}), \tag{21}$$

where

$$\mathcal{X}(y, \phi, \mathcal{I}) \triangleq \{x \in \mathbb{R}^n : d_x(y, \mathcal{I}) \leq \phi\}.$$

In other words, $\mathcal{X}(y, \phi)$ is a union of ellipsoids. It is worth pointing out that if $\operatorname{res}(\mathcal{I}) = \phi$, $\mathcal{X}(y, \phi, \mathcal{I})$ degenerates to a single point; and if $\operatorname{res}(\mathcal{I}) > \phi$, $\mathcal{X}(y, \phi, \mathcal{I})$ is empty. Therefore, to differentiate these cases, we define

$$\begin{aligned} \mathfrak{I}(\phi) &\triangleq \{\mathcal{I} \subseteq \mathcal{M} : \operatorname{res}(\mathcal{I}) \leq \phi \text{ and } |\mathcal{I}| = m - q\}, \\ \mathfrak{I}_+(\phi) &\triangleq \{\mathcal{I} \subseteq \mathcal{M} : \operatorname{res}(\mathcal{I}) < \phi \text{ and } |\mathcal{I}| = m - q\}, \\ \mathfrak{I}_0(\phi) &\triangleq \mathfrak{I}(\phi) \setminus \mathfrak{I}_+(\phi). \end{aligned} \tag{22}$$

**Lemma 5.** *Given $\phi$ such that $\mathfrak{I}(\phi)$ is not empty. Consider the following semidefinite programming problem:*

$$\begin{aligned} \underset{\tau \in \mathbb{R}^{|\mathfrak{I}_+(\phi)|}, c, \psi \in \mathbb{R}}{\text{minimize}} \quad & \psi \\ \text{subject to} \quad & \psi \geq 0, \\ & \tau_i \geq 0,\ \forall 1 \leq i \leq |\mathfrak{I}_+(\phi)|, \\ & \tau_{\mathrm{id}(\mathcal{I})} \Theta(\mathcal{I}, \phi) \succcurlyeq \theta(c, \psi),\ \forall \mathcal{I} \in \mathfrak{I}_+(\phi), \\ & \begin{bmatrix} \psi & (\kappa_{\mathcal{I}} y_{\mathcal{I}} - c)^\top \\ * & \boldsymbol{I}_n \end{bmatrix} \succcurlyeq 0,\ \forall \mathcal{I} \in \mathfrak{I}_0(\phi), \end{aligned}$$

*where*

$$\Theta(\mathcal{I}, \phi) \triangleq \begin{bmatrix} \operatorname{var}(\mathcal{I}) & -\operatorname{var}(\mathcal{I}) \kappa_{\mathcal{I}} y_{\mathcal{I}} & 0 \\ * & (\kappa_{\mathcal{I}} y_{\mathcal{I}})^\top \operatorname{var}(\mathcal{I}) \kappa_{\mathcal{I}} y_{\mathcal{I}} + \operatorname{res}(\mathcal{I}) - \phi & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

*with $*$ being recovered by symmetry,*

$$\theta(c, \psi) \triangleq \begin{bmatrix} \boldsymbol{I}_n & -c & 0 \\ -c^\top & -\psi & c^\top \\ 0 & c & -\boldsymbol{I}_n \end{bmatrix},$$

*and $\mathrm{id}(\cdot) : \mathfrak{I}_+(\phi) \mapsto \{1, 2, \ldots, |\mathfrak{I}_+(\phi)|\}$ is any one-to-one function. Then*

$$\operatorname{chv}(\mathcal{X}(y, \phi)) = c^*, \tag{23}$$

$$\operatorname{rad}(\mathcal{X}(y, \phi)) = \sqrt{\psi^*}, \tag{24}$$

*where $c^*$ and $\psi^*$ are the optimal solution of the semidefinite programming problem.*

*Proof.* See Appendix F. □

It follows from this lemma that, finding the Chebyshev center and radius of the set $\mathcal{X}(y, \phi)$ is a semidefinite programming problem when $y, \phi$ are given. However, we are interested in finding the optimal estimator that maximize the rate $r(f, \delta)$, for a given $\delta$. In the following lemma, we give how $\operatorname{rad}(\mathcal{X}(y, \phi))$ varies with $\phi$, the illustration of which is in Section V-A.

**Lemma 6.** *Given any $y \in \mathbb{R}^m$, the radius $\operatorname{rad}(\mathcal{X}(y, \phi))$ have the following properties:*

1) *$\operatorname{rad}(\mathcal{X}(y, \phi))$ is increasing, right-continuous w.r.t. $\phi$.*
2) *If $\operatorname{rad}(\mathcal{X}(y, \phi))$ is discontinuous at a point $\phi_0$, then there must exist a set $\mathcal{I} \subseteq \mathcal{M}$ with $|\mathcal{I}| = m - q$ such that $\operatorname{res}(\mathcal{I}) = \phi_0$.*
3) *When $\operatorname{rad}(\mathcal{X}(y, \phi)) > 0$, $\operatorname{rad}(\mathcal{X}(y, \phi))$ is strictly increasing w.r.t. $\phi$.*

*Proof.* See Appendix F. □

Given a predefined approximation bound $\varepsilon > 0$, we compute the corresponding estimate $\hat{x}$ for an averaged measurement $\operatorname{avg}(\boldsymbol{Y}(k)) \in \mathbb{R}^m$ in Algorithm 1. Denoted by $\hat{f}_\varepsilon$ the resulting estimator, and by $\hat{f}(y, \varepsilon)$ the output of Algorithm 1 (i.e., the estimate $\hat{x}$) when the inputs are $y, \varepsilon$.

Notice that Algorithm 1 is a slight variation of the classic bisection method. The distinguished part lies in (25), which together with Lemma 6 assures that for any $y \in \mathbb{R}^m$,

$$\inf_{x \notin \mathcal{B}_\delta(\hat{f}(y, \varepsilon))} d_x(y) \geq u(y, \delta) - \varepsilon,$$

where $u(y, \delta)$ is defined in (12). Therefore, the following theorem readily follows:

**Theorem 3.** *Let an estimator $\hat{f}_\varepsilon(\boldsymbol{Y}(k)) = \hat{f}(\operatorname{avg}(\boldsymbol{Y}(k)), \varepsilon)$ be computed by Algorithm 1 with $\operatorname{avg}(\boldsymbol{Y}(k)$ and $\varepsilon > 0$ as inputs, then for all $\delta > 0$ this estimator possesses the guaranteed performance:*

$$r(\hat{f}_\varepsilon, \delta) \geq r(f_\delta^*, \delta) - \varepsilon,$$

*where $f_\delta^*$ is the optimal estimator in* (11)

Clearly, a smaller $\varepsilon$ in Algorithm 1 leads to a better estimator, which, however, requires more iterations to run.

**Algorithm 1** Approximate Optimal Estimator $f_\delta^*$ in (11)

**Inputs:** averaged measurements $y \in \mathbb{R}^m$, performance error tolerance $\varepsilon > 0$.
**Output:** estimate $\hat{x} \in \mathbb{R}^n$
**Initialization:** Let

$$\underline{\phi} = \min\{\mathrm{res}(\mathcal{I}) : \mathcal{I} \subseteq \mathcal{M}, |\mathcal{I}| = m - q\} \triangleq \Upsilon,$$

and $\overline{\phi}$ be such that $\mathrm{rad}(\mathcal{X}(y, \overline{\phi})) > \delta$.
**Repeat:**
1. **If** $\overline{\phi} - \underline{\phi} < \varepsilon/2$ **then**

$$\phi = \max\{\Upsilon, \underline{\phi} - \varepsilon/2\}, \quad \hat{x} = \mathrm{chv}(\mathcal{X}(y, \phi)) \tag{25}$$

**Stop**
**EndIf**
2. Let $\phi = (\underline{\phi} + \overline{\phi})/2$.
3. **If** $\mathrm{rad}(\mathcal{X}(y, \phi)) = \delta$ **then**
$\hat{x} = \mathrm{chv}(\mathcal{X}(y, \phi))$
**Stop**
**ElseIf** $\mathrm{rad}(\mathcal{X}(y, \phi)) > \delta$ **then**
$\overline{\phi} = \phi$
**Else** $\underline{\phi} = \phi$
**EndIf**

**Remark 4.** *Though semidefinite programming problem can be (approximately) solved in a polynomial time of program size [28]. In our case, however, when $\phi$ is large enough, $|\mathfrak{I}_+(\phi)| = \binom{m}{q}$, where $\binom{m}{q}$ is the binomial coefficient, which renders the optimization problem in Lemma 5 rather computationally heavy when $\phi$ and $m$ are large. Nevertheless, we defend our estimator from the following two aspects. First, though efficient and optimal algorithms might exist in certain problems, see e.g., [21], the resilient information fusion under sparse attack is intrinsically of combinatorial nature, see e.g., [6], [8], [29], since we basically need to search over all combinations of possibly healthy sensors. Nevertheless, this work is just a starting point, and we are planning to investigate the approaches that could relieve the computational burden (in certain cases) just as in [30]–[32]. Second, in practice, a small $\delta$ would be usually chosen. Then the size of $\mathfrak{I}_+(\phi)$ will be small as well no matter how big $m$ is, and, therefore, the optimization problem in Lemma 5 could be efficiently solved. Though finding $\Upsilon$ of Algorithm 1 is of combinatorial nature, computing* $\mathrm{res}(\mathcal{I})$ *(given in* (19)*) for a given set $\mathcal{I}$ is light (notice that $\boldsymbol{W}_{\{\mathcal{I}\}}$ is a diagonal matrix and its inverse, therefore, is readily given). Therefore, the computation burden of Algorithm 1 could be tolerated for a large $m$.*

**Remark 5.** *The resilience of the proposed optimal estimator $f_\delta^*$ in* (11) *may not be that apparent since Chebyshev center itself is sensitive to noises, i.e., the Chebyshev center of a set $\mathcal{A}$ can be driven to anywhere even if only one point of $\mathcal{A}$ is allowed to be manipulated. Nevertheless, the resilience of the estimator $f_\delta^*$ can be heuristically explained by the following two factors. First, when the time $k$ is large enough, the measurements from benign sensors can lead to rather accurate estimate, i.e., the* $\mathrm{res}(\mathcal{I}_*)$ *will be quite small, where $\mathcal{I}_*$ is of size $m-q$ and contains no compromised sensors. Therefore, $\mathcal{I}_*$ would be in the collection $\mathfrak{I}(\phi)$ defined in* (22) *and somehow serves as an anchor when computing the estimate as in Algorithm 1 and Lemma 5. Second, when the injected bias of a compromised sensor is too large, the resulting* $\mathrm{res}(\mathcal{I})$ *for any $\mathcal{I}$ containing this compromised sensor will be quite large as well. Therefore, the set $\mathcal{I}$ will not be in the collection $\mathfrak{I}(\phi)$ and the measurement from this compromised sensor will be discarded when computing the estimate.*

## IV. Uniformly Optimal Estimator for Homogeneous Sensors

In this section we provide a simple yet uniformly optimal estimator $f$ such that $r(f, \delta)$ is simultaneously maximized for all $\delta > 0$ when the sensors are homogeneous, i.e., $H_1 = \cdots = H_m$ and $W_1 = \cdots = W_m$. Notice that when homogeneous sensors are considered, to satisfy the $2q$-observable assumption in Assumption 3, the state has to be scalar, i.e., $x \in \mathbb{R}$.

To proceed, we first provide an upper bound of the optimal performance proved in Theorem 2, $u(\delta)$, for any $\delta$ and *any* system models (instead of only homogeneous sensors).

**Lemma 7.** *The optimal performance $u(\delta)$ in* (14) *is upper bounded:*

$$u(\delta) \le \bar{u}(\delta),$$

*where $\bar{u}(\delta) = \delta^2 \bar{u}(1)$ with $\bar{u}(1)$ being the optimal value of the following optimization problem:*

$$\begin{aligned} \underset{x \in \mathbb{R}^n, s \in \mathbb{R}^m}{\text{minimize}} \quad & \frac{1}{2}\sum_{i=1}^m (H_i x + s_i)^2 / W_i \\ \text{subject to} \quad & \|s\|_0 \le 2q, \\ & \|x\|_2 = 1. \end{aligned} \tag{26}$$

*Proof.* See Appendix G. □

In the remainder of this section, we consider the case where sensors are homogeneous and the system is scalar. Then without loss of generality, we let $H_i = 1$ for any $1 \le i \le m$. We define the estimator $f^{\mathrm{trm}}$ be the "trimmed mean", i.e.,

$$f^{\mathrm{trm}}(\boldsymbol{Y}(k)) = \mathrm{trm}\left(\mathrm{avg}(\boldsymbol{Y}(k))\right), \tag{27}$$

where for any $y \in \mathbb{R}^m$,

$$\mathrm{trm}(y) \triangleq \frac{1}{m - 2q} \sum_{i=q+1}^{m-q} y_{[i]} \tag{28}$$

with $y_{[i]}$ being the $i$-th smallest element. In other words, $\mathrm{trm}(y)$ first discards the largest $q$ and smallest $q$ elements of $y$, and then averages over the remaining ones.

We show that the trimmed mean estimator $f^{\mathrm{trm}}$ is uniformly optimal in Theorem 4. The theorem is proved by showing that $f^{\mathrm{trm}}$ achieves the upper bound in Lemma 7 for every

$\delta$, which, in turn, means that the upper bound is tight when homogeneous sensors are considered.

**Theorem 4.** *When the sensors are homogeneous (and thus the system state is scalar), $f^{\mathrm{trm}}$ in (27) is uniformly optimal, i.e.,*

$$r(f^{\mathrm{trm}}, \delta) = u(\delta)$$

*holds for every $\delta$.*

*Proof.* See Appendix H. □

## V. Numerical Examples

### A. Illustration of $\mathcal{X}(y,\phi)$ and $\mathrm{rad}(\mathcal{X}(y,\phi))$

We illustrate how $\mathcal{X}(y,\phi)$ and $\mathrm{rad}(\mathcal{X}(y,\phi))$ vary with $\phi$ in Fig. 2 and Fig. 3, respectively. The parameters used are summarized as follows: $m = 4$ sensors used to estimate $x \in \mathbb{R}^2$, $q = 1$ sensor might be manipulated, measurement matrix $\boldsymbol{H} = [1,0;0,1;1,2;2,1]$, covariance matrix $\boldsymbol{W} = \mathrm{diag}(1,2,2,1)$, and observation $y = [4;-4;5;-5]$. Let $\mathrm{res}_{[i]}$ be the $i$-th item of the set $\{\mathrm{res}(\mathcal{I}) : \mathcal{I} \subseteq \mathcal{M}, |\mathcal{I}| = m - q\}$ sorted in an ascending order. Then we have, in our case, that $\mathrm{res}_{[1]} = 3.68182, \mathrm{res}_{[2]} = 5.78571, \mathrm{res}_{[3]} = 13.5, \mathrm{res}_{[4]} = 24.3$.

From Fig. 2, one sees that $\mathcal{X}(y,\phi)$ is indeed a union of several ellipses. One also sees in Fig. 3 that $\mathrm{rad}(\mathcal{X}(y,\phi))$ is strictly increasing w.r.t. $\phi$ when $\mathrm{rad}(\mathcal{X}(y,\phi)) > 0$, and discontinuous only at $\mathrm{res}_{[2]}$ and $\mathrm{res}_{[3]}$, which verifies Lemma 6. Notice also that as $\phi$ crosses $\mathrm{res}_{[4]}$ from below, the new ellipse, which is indicated by the red one in the right-below subfigure of Fig. 2, is inside the blue dashed circle that covers the previous three ellipses. Therefore, $\mathrm{rad}(\mathcal{X}(y,\phi))$ is continuous at $\phi = \mathrm{res}_{[4]}$.

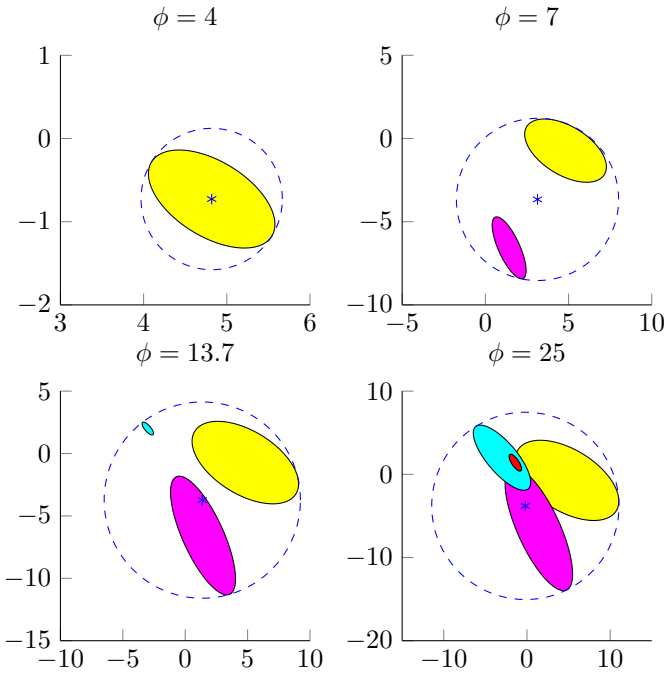



Fig. 2: The set $\mathcal{X}(y,\phi)$ with different $\phi$'s. In each of the four sub-figures, x-axis is $x_1$ and y-axis $x_2$. The filled area corresponds to $\mathcal{X}(y,\phi)$. The blue "*" is the Chebyshev center of $\mathcal{X}(y,\phi)$, and blue dashed line the circle centered at the Chebyshev center with radius $\mathrm{rad}(\mathcal{X}(y,\phi))$.

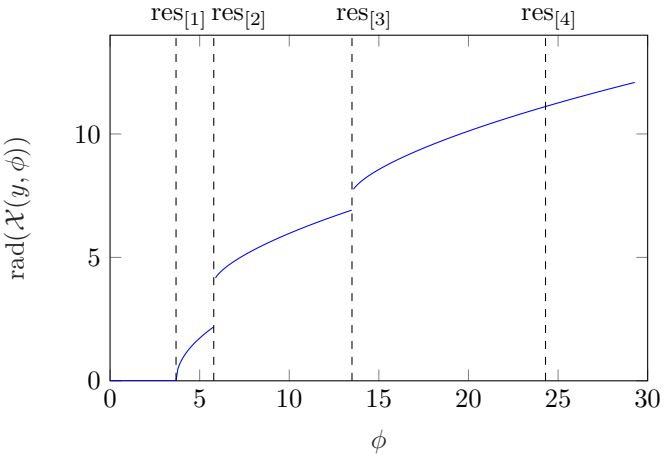



Fig. 3: Radius $\mathrm{rad}(\mathcal{X}(y,\phi))$ as a function of $\phi$.

### B. Resilience of the Proposed Estimator

In the following, in order to verify the intuitive comments of Remark 5 about the resilience of $f^*_\delta$ (11), we use a numerical example to show how the output of $f^*_\delta$ varies with the injected bias and $\delta$. The parameters used are summarized as follows: $m = 4$ sensors used to estimate $x \in \mathbb{R}^2$, measurement matrix $\boldsymbol{H} = [1,0;0,1;1,2;2,1]$, covariance matrix $\boldsymbol{W} = \mathrm{diag}(1,2,2,1)$, and observation $z = [1;1;3;3]$. We let the fourth sensor be attacked, i.e., the first three elements of $y$ are $[1;1;3]$ and $y_4 = z_4 + a$. In particular, we let $a$ vary from 0 to 15. We simulate our estimator $f^*_\delta$ for two different error thresholds $\delta = 1, 3$, and further compare it to the least squares estimator, which computes the estimate as $(\boldsymbol{H}^\top \boldsymbol{W}^{-1} \boldsymbol{H})^{-1} \boldsymbol{H}^\top \boldsymbol{W}^{-1} y$. When using Algorithm 1, we let the performance error tolerance $\varepsilon = 0.001$.

The result is illustrated in Fig. 4. One sees that when the bias injected $a$ is too large, the estimation error of our algorithm is zero, i.e., the attack effects are eliminated. This is consistent with intuitive comments in Remark 5. Furthermore, when using a smaller $\delta$ (i.e., $\delta = 1$ in our example), the estimator tends to discard the injected bias: the zero-error range is $a \in [3,\infty]$ for $\delta = 1$, which is contrasted with $[8,\infty]$ for $\delta = 3$. This is because given the same observation $y$, smaller $\delta$ is, smaller $\phi$ and, thus, the collection $\mathfrak{I}(\phi)$ are, which means that the "abnormal" data (with large $\mathrm{res}(\mathcal{I})$) will be more likely to be discarded. It is clear that the naive least squares estimator is not resilient to the attack.

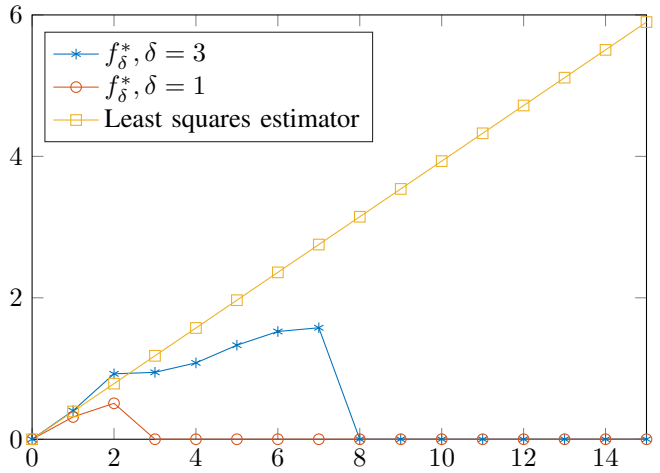



Fig. 4: 2-norm of estimation error as a function of the bias injected. Our algorithm with different $\delta$ and the least squares estimator are compared.

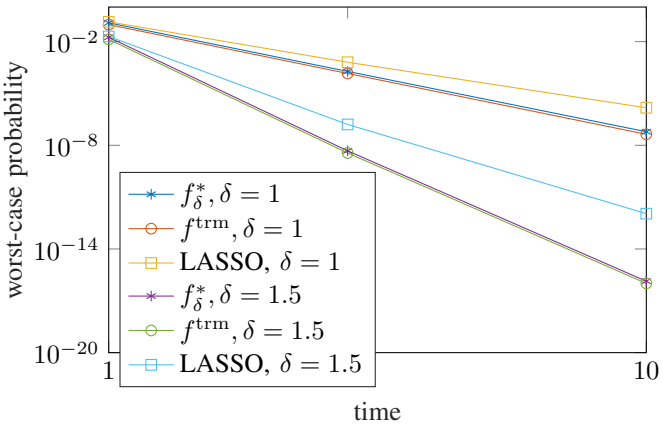



Fig. 5: Worst-case probability $e(f, k, \delta)$ in (3) as a function of estimator $f$ (our proposed estimator $f^*_\delta$ in Algorithm 1, trimmed mean estimator $f^{\mathrm{trm}}$ in (27), and LASSO in (29)), time $k$ $(1, 5, 10)$, and error threshold $\delta$ $(1, 1.5)$.

### C. Comparison with Other Estimators

In this section, we compare our estimator with the LASSO. In our case, given $\mathrm{avg}(\boldsymbol{Y}(k)) = y \in \mathbb{R}^m$, the LASSO reads

$$\underset{x\in\mathbb{R}^n, a\in\mathbb{R}^m}{\text{minimize}} \ \|(\boldsymbol{W}/k)^{-1/2}(y - \boldsymbol{H}x - a)\|^2 + \lambda\|a\|_1, \quad (29)$$

where $\lambda$ is predefined parameter and the optimal solution $x$ is the estimate. Basically, the smaller $\lambda$ is, the securer is LASSO. Therefore, in our simulation, we set $\lambda = 10^{-3}$. Notice that the $l_0$ and $l_1$-based state estimation procedures [6], [7] works in systems without noises or with (small) bounded measurement noises, while the estimator in [8] (i.e., Algorithm 2 thereof) is undecided for (many) certain observations, that is, it can happen that no subset of sensors are deemed as attack free and, therefore, no output will be generated. We should also note that while [9] proves the resilience of LASSO when each sensor is observable, i.e., $H_i$ is scalar in our case, the LASSO under sparse attack is not resilient in general; see [18]. Therefore, we consider scalar state in this simulation and for simplicity, we further assume the sensors are homogeneous.

We assume there are totally $m = 5$ sensors, of which $q = 1$ sensor is compromised. We let measurement matrix $\boldsymbol{H} = [1; 1; 1; 1; 1]$ and covariance matrix $\boldsymbol{W} = \mathrm{diag}(1, 1, 1, 1, 1)$. When computing the worst-case probability $e(f, k, \delta)$ in (3), we assume that, without loss of generality, the true state is $x = 0$ and the fifth sensor compromised. We then simulate the error probability for a fixed $y_5$ with $y = \mathrm{avg}(\boldsymbol{Y}(k))$ being the averaged measurement, the maximum of which is then regarded as the worst-case probability $e(f, k, \delta)$. From Fig. 5, one sees that for either $\delta = 1$ or $\delta = 1.5$, the performances of $f^{\mathrm{trm}}$ and $f^*_\delta$ are quite close, which is consistent with the uniform optimality of $f^{\mathrm{trm}}$ stated in Theorem 4. One should also note that both $f^{\mathrm{trm}}$ and $f^*_\delta$ outperform the LASSO.

## VI. Conclusion and Future Work

In this paper, we provided a different perspective on secure static state estimation with Byzantine sensors by introducing a new probabilistic performance metric, i.e., the decaying rate of the worst-case probability that the estimation error is larger than some value $\delta$ rather than the worst-case error or estimation error covariance in the existing literature. This new metric does not necessarily require bounded noise. With this metric, we gave an optimal estimator for any given error threshold $\delta$, which is the Chebyshev center of a certain set, and proposed an algorithm to compute it. A significant byproduct is that if distribution of the observation noise is in the exponential family, the sufficient statistic for the underlying state remains the same whether or not the attacker is present. When the sensors are homogeneous, we further derived a simple yet uniformly optimal estimator, which, to be specific, is the trimmed mean of the averaged observations and simultaneously optimal for every $\delta$.

For the future work, there are two interesting directions. One is to extend the existing results into dynamic systems, while the other one is to investigate the uniformly optimal estimator when sensors are heterogeneous.

## Appendix A
## Proof of Lemma 1

The proof is of constructive nature and mainly stems from the fact that $\mathrm{avg}(\boldsymbol{Z}(k))$ is a sufficient statistic for the underlying state $x$, where $\boldsymbol{Z}(k)$ is the "true" measurement matrix when there are no attacks and is defined in the same manner with $\boldsymbol{Y}(k)$.

In the following, for simplicity of presentation, we do not distinguish the probability density function (pdf) for a continuous random variable and probability mass function (pmf) for a discrete one. Therefore, in some cases the summation is actually needed though we use integration universally.

For any $f \in \mathcal{F}$, we let $f'$ satisfy (30) and (31). For any

$y \in \mathbb{R}^m$, Borel set $\mathcal{A} \subseteq \mathbb{R}^n$, and time $k$,

$$\begin{aligned}&\mathbb{P}_{f'}(\hat{x}_k \in \mathcal{A}|\operatorname{avg}(\boldsymbol{Y}(k)) = y)\\ =&\int_{Y\in\mathbb{R}^{m\times k}} \mathbb{P}_f(\hat{x}_k \in \mathcal{A}|\boldsymbol{Y}(k) = Y)\\ &\qquad \mathrm{d}\mathbb{P}(\boldsymbol{Z}(k) = Y|\operatorname{avg}(\boldsymbol{Z}(k)) = y). \end{aligned} \tag{30}$$

The above equation is the integral of $\mathbb{P}_f(\cdot)$ over $Y \in \mathbb{R}^{m\times k}$ with respect to the conditional probability measure $\mathbb{P}(\boldsymbol{Z}(k)|\operatorname{avg}(\boldsymbol{Z}(k)) = y)$. Notice that this conditional probability measure $\mathbb{P}(\boldsymbol{Z}(k)|\operatorname{avg}(\boldsymbol{Z}(k)) = y)$ is well-defined since $\operatorname{avg}(\boldsymbol{Z}(k))$ is a sufficient statistic of the "true" measurements $\boldsymbol{Z}(k)$ for the underlying state $x$, i.e., for any state $x$,

$$\mathbb{P}_x(\boldsymbol{Z}(k)|\operatorname{avg}(\boldsymbol{Z}(k)) = y) = \mathbb{P}(\boldsymbol{Z}(k)|\operatorname{avg}(\boldsymbol{Z}(k)) = y),$$

where $\mathbb{P}_x(\cdot)$ denotes the probability measure governing the original measurements $\boldsymbol{Z}(k)$ when the state $x$ is given. Notice that RHS of (30) can be interpreted as "taking expectation" of the conditional probability measure $\mathbb{P}_f(\hat{x}_k|\boldsymbol{Y}(k))$ given that $\operatorname{avg}(\boldsymbol{Y}(k)) = y$ and that $\boldsymbol{Y}(k)$ shares the same distribution with $\boldsymbol{Z}(k)$.

Furthermore, let $f'$ be in $\mathcal{F}_{\mathrm{c}}$, i.e.,

$$\mathbb{P}_{f'}(\hat{x}_k \in \mathcal{A}|\boldsymbol{Y}(k)) = \mathbb{P}_{f'}(\hat{x}_k \in \mathcal{A}|\boldsymbol{Y}'(k)) \tag{31}$$

for any Borel set $\mathcal{A} \subseteq \mathbb{R}^n$ whenever $\operatorname{avg}(\boldsymbol{Y}(k)) = \operatorname{avg}(\boldsymbol{Y}'(k))$.

Let $\mathcal{B}_\delta(x)$ denote the closed ball centered at $x \in \mathbb{R}^n$ with radius $\delta > 0$:

$$\mathcal{B}_\delta(x) \triangleq \{y \in \mathbb{R}^n : \|y - x\|_2 \le \delta\}. \tag{32}$$

Regarding with $f$ and $f'$, in the remainder of this proof we devote ourselves to showing that the following inequality holds for any state $x$, set $\mathcal{C}$, $\delta > 0$ and time $k$:

$$\sup_{g\in\mathcal{G}} \mathbb{P}_{f,g,x,\mathcal{C}}\left(\hat{x}_k \notin \mathcal{B}_\delta(x)\right) \ge \sup_{g\in\mathcal{G}} \mathbb{P}_{f',g,x,\mathcal{C}}\left(\hat{x}_k \notin \mathcal{B}_\delta(x)\right), \tag{33}$$

from which Lemma 1 follows straightforwardly.

We first identify the most harmful attack strategy for a generic $f$. Given state $x$, set $\mathcal{C}$, $\delta > 0$, time $k$, and estimator $f$, consider the following optimization problem:

$$\begin{aligned}\max_{Y_2\in\mathbb{R}^{q\times k}} \int_{Y_1\in\mathbb{R}^{(m-q)\times k}} &\mathbb{P}_f\left(\hat{x}_k \notin \mathcal{B}_\delta(x)\middle|\boldsymbol{Y}(k)_{\mathcal{M}\setminus\mathcal{C}} = Y_1,\right.\\ &\left.\boldsymbol{Y}(k)_{\mathcal{C}} = Y_2\right)\mathrm{d}\mathbb{P}_x(\boldsymbol{Z}(k)_{\mathcal{M}\setminus\mathcal{C}} = Y_1). \end{aligned} \tag{34}$$

Denote its optimal solution (i.e., the "manipulated matrix") as $\operatorname{mm}(f, x, \mathcal{C}, \delta, k)$. Then one may see that changing the measurements of the compromised sensors available at time $k$, $\boldsymbol{Y}(k)_{\mathcal{C}}$, to $\operatorname{mm}(f, x, \mathcal{C}, \delta, k)$ would maximize the error[2] probability under estimator $f$. The optimal value of the optimization problem (34) is just the worst-case error probability $\sup_{g\in\mathcal{G}} \mathbb{P}_{f,g,x,\mathcal{C}}\left(\hat{x}_k \notin \mathcal{B}_\delta(x)\right)$.

[2]For the sake of presentation, we call the event $\hat{x}_k \notin \mathcal{B}_\delta(x)$ an error.

We then identify the most harmful attack strategy for the compressed estimator $f'$. Given state $x$, set $\mathcal{C}$, $\delta > 0$, time $k$, and estimator $f'$, consider the following optimization problem:

$$\begin{aligned}\max_{y_2\in\mathbb{R}^q} \int_{y_1\in\mathbb{R}^{m-q}} &\mathbb{P}_{f'}\left(\hat{x}_k \notin \mathcal{B}_\delta(x)\middle|\operatorname{avg}(\boldsymbol{Y}(k))_{\mathcal{M}\setminus\mathcal{C}} = y_1,\right.\\ &\left.\operatorname{avg}(\boldsymbol{Y}(k))_{\mathcal{C}} = y_2\right)\mathrm{d}\mathbb{P}_x(\operatorname{avg}(\boldsymbol{Z}(k))_{\mathcal{M}\setminus\mathcal{C}} = y_1). \end{aligned} \tag{35}$$

Denote its optimal solution (i.e., the "manipulated vector") as $\operatorname{mv}(f', x, \mathcal{C}, \delta, k)$. One may verify that changing the measurements of the compromised sensors available at time $k$ such that $\operatorname{avg}(\boldsymbol{Z}(k))_{\mathcal{C}} = \operatorname{mv}(f', x, \mathcal{C}, \delta, k)$ would maximize the error probability under estimator $f'$. The optimal value of the optimization problem (35) is just the worst-case error probability $\sup_{g\in\mathcal{G}} \mathbb{P}_{f',g,x,\mathcal{C}}\left(\hat{x}_k \notin \mathcal{B}_\delta(x)\right)$.

For the sake of better presentation, in the remainder of this proof, for any matrix $\boldsymbol{M}$, we rewrite $\boldsymbol{M}_{\mathcal{M}\setminus\mathcal{C}}$ as $\boldsymbol{M}_{[1]}$ and $\boldsymbol{M}_{\mathcal{C}}$ as $\boldsymbol{M}_{[2]}$. We also omit the time index $k$ of $\boldsymbol{Z}(k)$ and $\boldsymbol{Y}(k)$. The set $\mathcal{B}_\delta(x)$ is denoted by $\mathcal{B}$. Notice that the "true" measurements $\boldsymbol{Z}$ are independent across sensors given the underlying state $x$. Therefore, we can rewrite (30) as follows:

$$\begin{aligned}&\mathbb{P}_{f'}(\hat{x}_k \in \mathcal{A}|\operatorname{avg}(\boldsymbol{Y}) = y)\\ =&\int_{\mathbb{R}^{q\times k}}\int_{\mathbb{R}^{(m-q)\times k}} \mathbb{P}_f(\hat{x}_k \in \mathcal{A}|\boldsymbol{Y}_{[1]} = Y_{[1]}, \boldsymbol{Y}_{[2]} = Y_{[2]})\\ &\qquad \mathrm{d}\mathbb{P}(\boldsymbol{Z}_{[1]} = Y_{[1]}|\operatorname{avg}(\boldsymbol{Z}_{[1]}) = y_{[1]})\\ &\qquad \mathrm{d}\mathbb{P}(\boldsymbol{Z}_{[2]} = Y_{[2]}|\operatorname{avg}(\boldsymbol{Z}_{[2]}) = y_{[2]}). \end{aligned}$$

Then one obtains that

$$\begin{aligned}&\sup_{g\in\mathcal{G}} \mathbb{P}_{f',g,x,\mathcal{C}}\left(\hat{x}_k \notin \mathcal{B}_\delta(x)\right)\\ =&\int_{\mathbb{R}^{m-q}}\int_{\mathbb{R}^{(m-q)\times k}}\int_{\mathbb{R}^{q\times k}} \mathbb{P}_f(\hat{x}_k \notin \mathcal{B}|\boldsymbol{Y}_{[1]} = Y_{[1]}, \boldsymbol{Y}_{[2]} = Y_{[2]})\\ &\qquad \mathrm{d}\mathbb{P}(\boldsymbol{Z}_{[2]} = Y_{[2]}|\operatorname{avg}(\boldsymbol{Z}_{[2]}) = \operatorname{mv}(f', x, \mathcal{C}, \delta, k))\\ &\qquad \mathrm{d}\mathbb{P}(\boldsymbol{Z}_{[1]} = Y_{[1]}|\operatorname{avg}(\boldsymbol{Z}_{[1]}) = z_{[1]})\\ &\qquad \mathrm{d}\mathbb{P}_x(\operatorname{avg}(\boldsymbol{Z}_{[1]}) = z_{[1]})\\ =&\int_{\mathbb{R}^{(m-q)\times k}}\int_{\mathbb{R}^{q\times k}} \mathbb{P}_f(\hat{x}_k \notin \mathcal{B}|\boldsymbol{Y}_{[1]} = Y_{[1]}, \boldsymbol{Y}_{[2]} = Y_{[2]})\\ &\qquad \mathrm{d}\mathbb{P}(\boldsymbol{Z}_{[2]} = Y_{[2]}|\operatorname{avg}(\boldsymbol{Z}_{[2]}) = \operatorname{mv}(f', x, \mathcal{C}, \delta, k))\\ &\qquad \mathrm{d}\mathbb{P}_x(\boldsymbol{Z}_{[1]} = Y_{[1]})\\ =&\int_{\mathbb{R}^{q\times k}}\int_{\mathbb{R}^{(m-q)\times k}} \mathbb{P}_f(\hat{x}_k \notin \mathcal{B}|\boldsymbol{Y}_{[1]} = Y_{[1]}, \boldsymbol{Y}_{[2]} = Y_{[2]})\\ &\qquad \mathrm{d}\mathbb{P}_x(\boldsymbol{Z}_{[1]} = Y_{[1]})\\ &\qquad \mathrm{d}\mathbb{P}(\boldsymbol{Z}_{[2]} = Y_{[2]}|\operatorname{avg}(\boldsymbol{Z}_{[2]}) = \operatorname{mv}(f', x, \mathcal{C}, \delta, k))\\ \le&\max_{Y_{[2]}\in\mathbb{R}^{q\times k}}\int_{\mathbb{R}^{(m-q)\times k}} \mathbb{P}_f(\hat{x}_k \notin \mathcal{B}|\boldsymbol{Y}_{[1]} = Y_{[1]}, \boldsymbol{Y}_{[2]} = Y_{[2]})\\ &\qquad \mathrm{d}\mathbb{P}_x(\boldsymbol{Z}_{[1]} = Y_{[1]})\\ =&\sup_{g\in\mathcal{G}} \mathbb{P}_{f,g,x,\mathcal{C}}\left(\hat{x}_k \notin \mathcal{B}_\delta(x)\right), \end{aligned}$$

where the second equality follows from the law of total probability, and the inequality holds because $\mathbb{P}(\boldsymbol{Z}_{[2]}|\operatorname{avg}(\boldsymbol{Z}_{[2]}) = y_{[2]})$ is a probability measure for any $y_{[2]}$, i.e.,

$$\int_{\mathbb{R}^{q\times k}} \mathrm{d}\mathbb{P}(\boldsymbol{Z}_{[2]} = Y_{[2]}|\operatorname{avg}(\boldsymbol{Z}_{[2]}) = y_{[2]}) = 1$$

for any $y_{[2]}$. The proof is thus complete.

In order for readers to have a better picture of relationship between the main results obtained in this paper (e.g., Theorems 1 and 2) and the assumptions posed in Section II, in particular, Assumptions 1 and 4, in the following remark, we explain in detail how these assumptions are utilized to derive (34).

**Remark 6.** *Due to Assumption 1, one could split the sensors into two groups: "good" ones in $\mathcal{M}\setminus\mathcal{C}$ and "bad" ones in $\mathcal{C}$. All the measurements from up to time $k$ from good sensors are not manipulated and, thus, denoted by $\boldsymbol{Z}(k)_{\mathcal{M}\setminus\mathcal{C}}$. The attacker can develop the term $\mathbb{P}_x(\boldsymbol{Z}(k)_{\mathcal{M}\setminus\mathcal{C}}=Y_1)$ since it knows the true state $x$ and system parameters by Assumption 4. Furthermore, since the attacker knows the estimator $f$ and has unlimited memory, it is proper to obtain the most harmful attack strategy by* (34)*.*

## APPENDIX B
## PROOF OF LEMMA 2

### A. Preliminaries

In the following lemma, we bound the area where a random variable has a high probability showing up.

Given any random variable $y\in\mathbb{R}^n$, we shall say that a point $x$ is $\delta$-typical, if

$$\mathbb{P}(y\in\mathcal{B}_\delta(x))>n/(n+1), \tag{36}$$

where $\mathcal{B}_\delta(x)$ is the closed ball defined in (32). In other words, $y$ has a high probability lying in the $\delta$-neighborhood of $x$.

Let $\mathcal{T}_\delta(y)$ denote the set of all $\delta$-typical point $x$ w.r.t. a random variable $y$, i.e.,

$$\mathcal{T}_\delta(y)\triangleq\{x\in\mathbb{R}^n : x \text{ is } \delta\text{-typical w.r.t. } y\}. \tag{37}$$

We have the following lemma to show that $\mathcal{T}_\delta(y)$ lies in a ball with radius $\delta$:

**Lemma 8.** *For any random variable $y\in\mathbb{R}^n$, there exists $x^*\in\mathbb{R}^n$ such that*

$$\mathcal{T}_\delta(y)\subseteq\mathcal{B}_\delta(x^*). \tag{38}$$

To proceed, we need the following lemma:

**Lemma 9.** *Let $\mathcal{A}_1,\dots,\mathcal{A}_n$ be $n$ random events with the same underlying probability space, then it holds that*

$$\mathbb{P}(\cap_{j=1}^n\mathcal{A}_j)\geq\sum_{j=1}^n\mathbb{P}(\mathcal{A}_j)-n+1. \tag{39}$$

*Proof of Lemma 9.*

$$\begin{aligned}\mathbb{P}(\cap_{j=1}^n\mathcal{A}_j)&=1-\mathbb{P}(\cup_{j=1}^n\mathcal{A}_j^c)\geq 1-\sum_{j=1}^n\mathbb{P}(\mathcal{A}_j^c)\\&=1-\sum_{j=1}^n(1-\mathbb{P}(\mathcal{A}_j))=\sum_{j=1}^n\mathbb{P}(\mathcal{A}_j)-n+1,\end{aligned}$$

where $\mathcal{A}^c$ is the complement of set $\mathcal{A}$. The proof is thus complete. □

*Proof of Lemma 8.* If $\mathcal{T}_\delta(y)$ is empty, then obviously $\mathcal{T}_\delta(y)=\emptyset\subset\mathcal{B}_\delta(0)$.

If $\mathcal{T}_\delta(y)$ only contains $j\leq(n+1)$ elements, say, $x_1,\dots,x_j$. Then Lemma 9 together with (36) yields that

$$\mathbb{P}(y\in\cap_{i=1}^j\mathcal{B}_\delta(x_i))>j\times\frac{n}{n+1}-j+1\geq 0,$$

which means that the set $\cap_{i=1}^j\mathcal{B}_\delta(x_i)$ is not empty. Then $\mathcal{T}_\delta(y)\subseteq\mathcal{B}_\delta(x^*)$ for some $x^*\in\cap_{i=1}^j\mathcal{B}_\delta(x_i)$.

If $\mathcal{T}_\delta(y)$ contains $j>(n+1)$ elements ($j$ might be infinite). Then again by Lemma 9, one obtains that for *any* $n+1$ elements, say, $x_1,\dots,x_{n+1}$, there holds

$$\mathbb{P}(y\in\cap_{i=1}^{n+1}\mathcal{B}_\delta(x_i))>0,$$

that is, $\cap_{i=1}^{n+1}\mathcal{B}_\delta(x_i)\neq\emptyset$. Since $\mathcal{B}_\delta(x)$ is compact and convex for any $x$, then Helly's theorem [33] means that

$$\cap_{x\in\mathcal{T}_\delta(y)}\mathcal{B}_\delta(x)\neq\emptyset.$$

Then $\mathcal{T}_\delta(y)\subseteq\mathcal{B}_\delta(x^*)$ for some $x^*\in\cap_{x\in\mathcal{T}_\delta(y)}\mathcal{B}_\delta(x)$. The proof is thus complete. □

**Definition 5.** *Any point $x^*\in\mathbb{R}^n$ is said to be a $\delta$-center of a random variable $y\in\mathbb{R}^n$ if it is such that* (38) *holds.*

The following follows readily from Lemma 8.

**Lemma 10.** *If $x^*\in\mathbb{R}^n$ is a $\delta$-center of a random variable $y\in\mathbb{R}^n$, then for any $x\in\mathbb{R}^n$:*

$$1-\mathbb{1}_{\mathcal{B}_\delta(x)}(x^*)\leq(n+1)\mathbb{P}(y\notin\mathcal{B}_\delta(x)). \tag{40}$$

### B. Main Body

Consider any estimator $f\in\mathcal{F}$, we construct a deterministic one $f'\in\mathcal{F}_{\rm d}$: for any time $k$ and observations $\boldsymbol{Y}(k)$, let $f'_k(\boldsymbol{Y}(k))$ be a $\delta$-center of the random variable $f_k(\boldsymbol{Y}(k))$, the existence of which is guaranteed by Lemma 8.

Hence, for any attack strategy $g$, true state $x$, set of compromised sensors $\mathcal{C}$, and time $k$, we have

$$\begin{aligned}&\mathbb{P}_{f',g,x,\mathcal{C}}(\|\hat{x}_k-x\|>\delta)\\=&\int_{Y\in\mathbb{R}^{m\times k}}1-\mathbb{1}_{\mathcal{B}_\delta(x)}(f'_k(Y))\,\mathrm{d}\mathbb{P}_{g,x,\mathcal{C}}(\boldsymbol{Y}(k)=Y)\\\leq&\int_{Y\in\mathbb{R}^{m\times k}}(n+1)\mathbb{P}(f_k(Y)\notin\mathcal{B}_\delta(x))\,\mathrm{d}\mathbb{P}_{g,x,\mathcal{C}}(\boldsymbol{Y}(k)=Y)\\=&(n+1)\mathbb{P}_{f,g,x,\mathcal{C}}(\|\hat{x}(k)-x\|>\delta),\end{aligned}$$

where the inequality follows from Lemma 10, and $\mathbb{P}_{g,x,\mathcal{C}}(\cdot)$ denotes the probability measure governing $\boldsymbol{Y}(k)$ when $g,x,\mathcal{C}$ are given. Then it is clear that

$$e(f,k,\delta)\geq e(f',k,\delta)/(n+1). \tag{41}$$

Recall that $e(f,k,\delta)$ is the worst-case probability defined in (3). Then it follows that for any $\delta>0$:

$$\begin{aligned}r(f,\delta)=&\liminf_{k\to\infty}-\frac{\log e(f,k,\delta)}{k}\\\leq&\liminf_{k\to\infty}-\frac{\log e(f',k,\delta)/(n+1)}{k}\\=&\liminf_{k\to\infty}-\frac{\log e(f',k,\delta)}{k}\\=&r(f',\delta).\end{aligned} \tag{42}$$

The proof is thus complete.

## APPENDIX C
## PROOF OF THEOREM 1

Consider a compressed but possibly random estimator $f \in \mathcal{F}_{\rm c}$, we construct a deterministic one $f' \in \mathcal{F}_{\rm cd}$ satisfying:

- for any time $k$ and observations $\boldsymbol{Y}(k)$, $f'(\boldsymbol{Y}(k))$ is a $\delta$-center of the random variable $f(\boldsymbol{Y}(k))$;
- $f'(\boldsymbol{Y}(k)) = f'(\boldsymbol{Y}'(k))$ if $\mathrm{avg}(\boldsymbol{Y}(k)) = \mathrm{avg}(\boldsymbol{Y}'(k))$.

The existence of $f'$ is guaranteed by Lemma 8 and the fact that random variables $f(\boldsymbol{Y}(k)), f(\boldsymbol{Y}'(k))$ have the same distribution if $\mathrm{avg}(\boldsymbol{Y}(k)) = \mathrm{avg}(\boldsymbol{Y}'(k))$.

Then as in Appendix B, one obtains $r(f,\delta) \leq r(f',\delta)$, which, together with Lemma 1, concludes the proof.

## APPENDIX D
## PROOF OF THEOREM 2

We first prove that $r(f,\delta)$ is upper-bounded by $u(\delta)$ for any $f \in \mathcal{F}$ in Lemma 12. We then show that $r(f^*_\delta,\delta) = u(\delta)$ in Lemma 13. Before proceeding, we need the following supporting definitions and lemmas.

### A. Supporting Definition and Lemmas

**Lemma 11.** *For any $\mathcal{A} \subseteq \mathbb{R}^n$, if $\mathrm{rad}(\mathcal{A}) > \gamma$, there exists a set $\mathcal{A}_0 \subseteq \mathcal{A}$ such that $|\mathcal{A}_0| \leq n+1$ and $\mathrm{rad}(\mathcal{A}_0) > \gamma$.*

*Proof.* If $|\mathcal{A}| \leq n+1$, then the lemma holds trivially by letting $\mathcal{A}_0 = \mathcal{A}$. When $|\mathcal{A}| > n+1$, we prove the lemma by contradiction. Suppose for every subset $\mathcal{A}_0 \subseteq \mathcal{A}$ with $|\mathcal{A}_0| \leq n+1$, there holds $\mathrm{rad}(\mathcal{A}_0) \leq \gamma$. Then we have

$$\cap_{x\in\mathcal{A}_0}\mathcal{B}_\gamma(x) \neq \emptyset$$

for every $\mathcal{A}_0 \subseteq \mathcal{A}$ with $|\mathcal{A}_0| = n+1$. Since $\mathcal{B}_\gamma(x)$ is compact and convex for every $x \in \mathcal{A}$, Helly's theorem [33] implies that

$$\cap_{x\in\mathcal{A}}\mathcal{B}_\gamma(x) \neq \emptyset.$$

Hence, for any $x_0 \in \cap_{x\in\mathcal{A}}\mathcal{B}_\gamma(x)$, $\mathcal{A} \subseteq \mathcal{B}_\gamma(x_0)$. Therefore $\mathrm{rad}(\mathcal{A}) \leq \gamma$, which contradicts the condition $\mathrm{rad}(\mathcal{A}) > \gamma$. □

With a slight abuse of notation, we use the sequence of functions $(f_1(\mathrm{avg}(\mathbf{Y}(1))), f_2(\mathrm{avg}(\mathbf{Y}(2))), \ldots)$ from time 1 to $\infty$ to denote a compressed and deterministic estimator $f \in \mathcal{F}_{\rm cd}$.

**Definition 6.** *Given a compressed and deterministic estimator $f \in \mathcal{F}_{\rm cd}$, $x \in \mathbb{R}^n$, $\delta > 0$, and time $k$, let $\mathcal{Y}(f,x,\delta,k)$ be the set of averaged measurements $\mathrm{avg}(\mathbf{Y}(k))$ such that the estimate $\hat{x}_k$ lies outside the ball $\mathcal{B}_\delta(x)$, i.e.,*

$$\mathcal{Y}(f,x,\delta,k) \triangleq \{y \in \mathbb{R}^m : f_k(y) \notin \mathcal{B}_\delta(x)\}. \tag{43}$$

### B. Upper Bound

**Lemma 12.** *For any estimator $f \in \mathcal{F}$, there holds*

$$r(f,\delta) \leq u(\delta). \tag{44}$$

*Proof.* We show that $r(f,\delta) < u(\delta)+\epsilon$ for any $\epsilon > 0$ and $f \in \mathcal{F}$.

Given $\epsilon > 0$, from the definition of $u(\delta)$, one obtains that there must exist $y^* \in \mathbb{R}^m$ and a set $\mathcal{A} \subseteq \mathbb{R}^n$ such that:

1) $d_x(y^*) \leq u(\delta) + \epsilon/2$ for all $x \in \mathcal{A}$;
2) $\mathrm{rad}(\mathcal{A}) > \delta$.

Notice that the above $y^*$ and $\mathcal{A}$ can be constructed as follows. By the definition of $u(\delta)$, there must exist a $y^*$ such that for every $x \in \mathbb{X}(y^*,\delta)$, $d_x(y^*) < u(\delta)+\epsilon/4$ holds. Then we construct $\mathcal{A}$ by cases. If $\mathrm{rad}(\mathbb{X}(y^*,\delta)) = \delta$, then there must exist $\phi^* < u(\delta)+\epsilon/4$ such that $\mathcal{X}(y^*,\phi^*) = \mathbb{X}(y^*,\delta)$. Let $\mathcal{A} = \mathcal{X}(y^*,\phi^*+\epsilon/4)$, and, therefore, $\mathcal{A} \subset \mathcal{X}(y^*,u(\delta)+\epsilon/2)$. Also, by the third bullet of Lemma 6, $\mathrm{rad}(\mathcal{A}) > \delta$ holds. If $\mathrm{rad}(\mathbb{X}(y^*,\delta)) < \delta$, then let $\phi^* = \min\{\phi : \mathbb{X}(y^*,\delta) \subseteq \mathcal{X}(y^*,\phi)\}$ and $\mathcal{A} = \mathcal{X}(y^*,\phi^*)$. Then by Lemma 6, $\mathrm{rad}(\mathcal{A}) > \delta$ and $d_x(y^*) \leq u(\delta)+\epsilon/4$ for all $x \in \mathcal{A}$.

Lemma 11 yields that there exists $\mathcal{A}_0 \subseteq \mathcal{A}$ such that $\mathrm{rad}(\mathcal{A}_0) > \delta$ and $|\mathcal{A}_0| \leq n+1$. Let $a^*(y,x)$ be the optimal solution to the optimization problem in (8) given $y \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$. Since $d_x(y,\mathcal{I})$ in (16) is continuous w.r.t. $y$ and $|\mathcal{A}_0| \leq n+1$, then one obtains that there exists a ball $\mathcal{B}_\beta(y^*)$, where $\beta > 0$ is dependent on $\epsilon$, such that $d_x(y, \mathcal{M} \setminus \mathrm{supp}(a^*(y^*,x))) < u(\delta)+\epsilon$ for every $x \in \mathcal{A}_0$ and every $y \in \mathcal{B}_\beta(y^*)$.

By Theorem 1, one suffices to consider a compressed and deterministic estimator $f \in \mathcal{F}_{\rm cd}$. Furthermore, since $\mathrm{rad}(\mathcal{A}_0) > \delta$, one concludes that for every time $k$ and $f \in \mathcal{F}_{\rm cd}$, there holds

$$\mathcal{B}_\beta(y^*) \subseteq \cup_{x\in\mathcal{A}_0}\mathcal{Y}(f,x,\delta,k). \tag{45}$$

Let $\mathfrak{L}_n(\cdot)$ denote the Lebesgue measure on $\mathbb{R}^n$. Because of countable additivity of Lebesgue measure [34], one obtains that there must exist a point $x^* \in \mathcal{A}_0$ such that

$$\mathfrak{L}_m(\mathcal{B}_\beta(y^*) \cap \mathcal{Y}(f,x^*,\delta,k)) \geq \mathfrak{L}_m(\mathcal{B}_\beta(y^*))/(n+1). \tag{46}$$

For the sake of simplicity, let $\mathcal{B}(x^*,k) \triangleq \mathcal{B}_\beta(y^*) \cap \mathcal{Y}(f,x^*,\delta,k)$ and $\mathcal{I}^* \triangleq \mathrm{supp}(a^*(y^*,x^*))$. Then it is clear that

$$\begin{aligned}
&e(f,k,\delta)\\
\geq&\sup_{g\in\mathcal{G}}\mathbb{P}_{g,x^*,\mathcal{I}^*}\left(\mathrm{avg}(\mathbf{Y}(k)) \in \mathcal{Y}(f,x^*,\delta,k)\right)\\
\geq&\sup_{g\in\mathcal{G}}\mathbb{P}_{g,x^*,\mathcal{I}^*}\left(\mathrm{avg}(\mathbf{Y}(k)) \in \mathcal{B}(x^*,k)\right)\\
=&\sup_{\mathrm{avg}(\mathbf{Z}(k))_{\mathcal{I}^*}\in\mathbb{R}^q}\mathbb{P}_{x^*}\left(\mathrm{avg}(\mathbf{Z}(k)) \in \mathcal{B}(x^*,k)\right)\\
=&\sup_{o\in\mathbb{R}^q}\mathbb{P}_{x^*}\left(\mathrm{avg}(\mathbf{Z}(k))_{\mathcal{M}\setminus\mathcal{I}^*} \in \mathbb{B}(o,k)\right),
\end{aligned} \tag{47}$$

where $\mathbb{B}(o,k)$ with $o \in \mathbb{R}^q$ is the projected set of $\mathcal{B}(x^*,k)$:

$$\mathbb{B}(o,k) \triangleq \{y_{\mathcal{M}\setminus\mathcal{I}^*} : y \in \mathcal{B}(x^*,k), y_{\mathcal{I}^*} = o\}.$$

Further let $\mathbb{B}(k)$ be a set that satisfies:

$$\mathfrak{L}_{m-q}(\mathbb{B}(k)) = \sup_{o\in\mathbb{R}^q}\mathfrak{L}_{m-q}(\mathbb{B}(o,k)),$$

and for any $\upsilon > 0$, there exists $o \in \mathbb{R}^q$ such that

$$\mathfrak{L}_{m-q}(\mathbb{B}(k) \setminus \mathbb{B}(o,k)) < \upsilon.$$

Roughly speaking, $\mathbb{B}(k)$ can be viewed as the supremum set. Then one obtains that

$$\begin{aligned}
&\sup_{o\in\mathbb{R}^q}\mathbb{P}_{x^*}\left(\mathrm{avg}(\mathbf{Z}(k))_{\mathcal{M}\setminus\mathcal{I}^*} \in \mathbb{B}(o,k)\right)\\
&\geq \mathbb{P}_{x^*}\left(\mathrm{avg}(\mathbf{Z}(k))_{\mathcal{M}\setminus\mathcal{I}^*} \in \mathbb{B}(k)\right)
\end{aligned} \tag{48}$$

In the following, we focus on characterizing the term in (48). Let $p_{x^*}(\cdot) : \mathbb{R}^{m-q} \mapsto \mathbb{R}_+$ be the probability density of $\mathrm{avg}(\mathbf{Z}(k))_{\mathcal{M}\setminus\mathcal{I}^*}$ conditioned on the underlying state $x^*$, i.e.,

$$p_{x^*}(z) = \mathcal{N}(\boldsymbol{H}_{\mathcal{M}\setminus\mathcal{I}^*}, \boldsymbol{W}_{\{\mathcal{M}\setminus\mathcal{I}^*\}}/k, z),$$

where $\boldsymbol{W} = \mathrm{diag}(W_i, W_2, \ldots, W_m)$ is the diagonal matrix, $\boldsymbol{W}_{\{\mathcal{M}\setminus\mathcal{I}^*\}}$ (different from $\boldsymbol{W}_{\mathcal{M}\setminus\mathcal{I}^*}$) the square matrix obtained from $\boldsymbol{W}$ after removing all of the rows and columns except those in the index set $\mathcal{M}\setminus\mathcal{I}^*$, and $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{x})$ the probability density function of a Gaussian random variable with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$ taking value at $\boldsymbol{x}$. Then one obtains that

$$\begin{aligned}&\mathbb{P}_{x^*}\left(\mathrm{avg}(\mathbf{Z}(k))_{\mathcal{M}\setminus\mathcal{I}^*} \in \mathbb{B}(k)\right)\\ =&\int_{\mathbb{R}^{m-q}} \mathbb{1}_{\mathbb{B}(k)}(z) p_{x^*}(z)\mathrm{d}z.\end{aligned}$$

From (46), some basic arguments mainly involving the regularity theorem for Lebesgue measure and the Heine–Borel theorem [34] give that there exists $\gamma > 0$ such that

$$\mathfrak{L}_{m-q}(\mathbb{B}(k)) > \gamma \mathfrak{L}_{m-q}(\mathcal{B}_\beta(y^*)_{\mathcal{M}\setminus\mathcal{I}^*}). \tag{49}$$

Furthermore, $\gamma$ can be determined by $m, n, q$, and $\beta$, which is, in particular, irrelevant to time $k$. Let $\mathcal{Z}(x^*, k) \subseteq \mathcal{B}_\beta(y^*)_{\mathcal{M}\setminus\mathcal{I}^*}$ be the pre-image of $(\underline{p}, \bar{p})$ under the function $p_{x^*}(\cdot)$, where $\underline{p} \triangleq \min_{z\in\mathcal{B}_\beta(y^*)_{\mathcal{M}\setminus\mathcal{I}^*}} p_{x^*}(z)$ is the minimum value[3] and $\bar{p}$ is such that

$$\mathfrak{L}_{m-q}(\mathcal{Z}(x^*, k)) = \gamma \mathfrak{L}_{m-q}(\mathcal{B}_\beta(y^*)_{\mathcal{M}\setminus\mathcal{I}^*}). \tag{50}$$

Notice that $\bar{p}$ exists since $\mathfrak{L}_{m-q}(\{z : p_{x^*}(z) = p\}) = 0$ for any $p$. Then one obtains that

$$\begin{aligned}&\mathbb{P}_{x^*}\left(\mathrm{avg}(\mathbf{Z}(k))_{\mathcal{M}\setminus\mathcal{I}^*} \in \mathbb{B}(k)\right)\\ \geq&\mathbb{P}_{x^*}\left(\mathrm{avg}(\mathbf{Z}(k))_{\mathcal{M}\setminus\mathcal{I}^*} \in \mathcal{Z}(x^*, k)\right).\end{aligned} \tag{51}$$

Notice that the pre-image of an open set under a continuous function is also open, $\mathcal{Z}(x^*, k)$ is thus open. Furthermore, since both $\gamma$ and $\mathcal{B}_\beta(y^*)$ are independent of time $k$, $\mathcal{Z}(x^*, k)$ will be an nonempty set whatever $k$ is. Therefore, the following holds:

$$\begin{aligned}&\limsup_{k\to\infty} \frac{1}{k} \log \mathbb{P}_{x^*}\left(\mathrm{avg}(\mathbf{Z}(k))_{\mathcal{M}\setminus\mathcal{I}^*} \in \mathcal{Z}(x^*, k)\right)\\ \leq&\inf_{z\in\mathcal{Z}(x^*,k)} \frac{1}{2}(z - \boldsymbol{H}_{\mathcal{M}\setminus\mathcal{I}^*}x^*)^\top \left(\boldsymbol{W}_{\mathcal{M}\setminus\mathcal{I}^*}\right)^{-1}(z - \boldsymbol{H}_{\mathcal{M}\setminus\mathcal{I}^*}x^*)\\ =&\inf_{z\in\mathbb{R}^m, z_{\mathcal{M}\setminus\mathcal{I}^*}\in\mathcal{Z}(x^*,k)} d_{x^*}(y, \mathcal{M}\setminus\mathcal{I}^*)\\ <&u(\delta) + \epsilon,\end{aligned} \tag{52}$$

where first inequality is due to the Cramér's Theorem [35] and the fact that $\mathcal{Z}(x^*, k)$ is open and $d_{x^*}(\cdot, \mathcal{M}\setminus\mathcal{I}^*)$ is the corresponding rate function since the observation noise $w_i(k)$ is i.i.d. across time and independent across the sensors; the last inequality follows from the definitions of $\mathcal{Z}(x^*, k)$ and $\mathcal{B}_\beta(y^*)$. Then, combining with (47), (48) and (51), one concludes the proof. □

[3]Notice that this minimum can be attained since $p_{x^*}(z)$ is a continuous function and $\mathcal{B}_\beta(y^*)_{\mathcal{M}\setminus\mathcal{I}^*}$ is compact.

### C. Achievability

About the estimator $f^*_\delta$ defined in (11), we have the following lemma:

**Lemma 13.** *There holds $r(f^*_\delta, \delta) = u(\delta)$.*

*Proof.* Notice that, by the definition of $u(\delta)$, for any $x$, $\delta$ and $k$, if $y \in \mathcal{Y}(f^*_\delta, x, \delta, k)$, then $d_x(y) \geq u(\delta)$. Recall that $\mathcal{Y}(\cdot,\cdot,\cdot,\cdot)$ is introduced in Definition 6. Let

$$\mathcal{Y}^*(x) \triangleq \{y : d_x(y) \geq u(\delta)\}.$$

Then $\mathcal{Y}(f^*_\delta, x, \delta, k) \subseteq \mathcal{Y}^*(x)$ holds. Therefore, for any $k$, $x$ and $\mathcal{I}$:

$$\begin{aligned}&\sup_{g\in\mathcal{G}} \mathbb{P}_{g,x,\mathcal{I}}\left(\mathrm{avg}(\mathbf{Y}(k)) \in \mathcal{Y}(f^*_\delta, x, \delta, k)\right)\\ \leq&\sup_{g\in\mathcal{G}} \mathbb{P}_{g,x,\mathcal{I}}\left(\mathrm{avg}(\mathbf{Y}(k)) \in \mathcal{Y}^*(x)\right)\\ \leq&\mathbb{P}_x\left(\mathrm{avg}(\mathbf{Z}(k))_{\mathcal{M}\setminus\mathcal{I}} \in \mathcal{Y}^*(x)_{\mathcal{M}\setminus\mathcal{I}}\right).\end{aligned}$$

Then similar to (52), by the Cramér's Theorem [35] and the fact that $\mathcal{Y}^*(x)_{\mathcal{M}\setminus\mathcal{I}}$ is closed, one obtains that

$$\begin{aligned}&\liminf_{k\to\infty} \frac{1}{k} \log \mathbb{P}_x\left(\mathrm{avg}(\mathbf{Z}(k))_{\mathcal{M}\setminus\mathcal{I}} \in \mathcal{Y}^*(x)_{\mathcal{M}\setminus\mathcal{I}}\right)\\ \geq&\inf_{z\in\mathbb{R}^m, z_{\mathcal{M}\setminus\mathcal{I}}\in\mathcal{Y}^*(x)_{\mathcal{M}\setminus\mathcal{I}}} d_x(z, \mathcal{M}\setminus\mathcal{I})\\ =&\inf_{z\in\mathcal{Y}^*(x)} d_x(z, \mathcal{M}\setminus\mathcal{I})\\ \geq&\inf_{z\in\mathcal{Y}^*(x)} d_x(z)\\ \geq&u(\delta).\end{aligned}$$

Since the above argument holds for any $x$ and $\mathcal{I}$, one concludes that $r(f^*_\delta, \delta) \geq u(\delta)$. Furthermore, $r(f^*_\delta, \delta)$ is upper bounded by $u(\delta)$ due to Lemma 12, the proof is thus complete. □

## APPENDIX E
## PROOF OF LEMMA 3

Using the same argument as in the proof of Lemma 12, one readily obtains the second bullet of Lemma 3 from the first one. Therefore, we focus on the first bullet in the sequel.

The proof is of constructive nature. Since $\boldsymbol{H}$ is not $2q$-observable, without loss of generality, we let $H_{\mathcal{I}^*}$ is not of full column rank with $\mathcal{I}^* = \{2q+1, \ldots, m\}$. For any $\delta$, let $x_1, x_2 \in \mathbb{R}^n$ be any two vectors such that

$$H_{\mathcal{I}^*}(x_2 - x_1) = 0, \text{ and } \|x_2 - x_1\| > \delta.$$

Let $\mathcal{I}_1 = \{1, \ldots, q\}$ and $\mathcal{I}_2 = \{q+1, \ldots, 2q\}$. We then construct $y^*$ as follows:

$$\begin{aligned}y^*_{\mathcal{I}^*} &= H_{\mathcal{I}^*}x_1,\\ y^*_{\mathcal{I}_1} &= H_{\mathcal{I}_1}x_1,\\ y^*_{\mathcal{I}_2} &= H_{\mathcal{I}_2}x_2.\end{aligned}$$

Then it is easy to verify that $d_{x_1}(y^*) = d_{x_2}(y^*) = 0$. The proof is thus complete.

## APPENDIX F
## PROOFS OF LEMMAS IN SECTION III-C

*Proof of Lemma 4.* For any index set $\mathcal{I}$, there holds

$$\begin{aligned} d_x(y,\mathcal{I}) =& \frac{1}{2}(y_{\mathcal{I}} - \boldsymbol{H}_{\mathcal{I}}x)^\top \boldsymbol{W}_{\{\mathcal{I}\}}^{-1}(y_{\mathcal{I}} - \boldsymbol{H}_{\mathcal{I}}x) \\ =& \frac{1}{2}(\sqrt{\boldsymbol{W}_{\{\mathcal{I}\}}^{-1}}y_{\mathcal{I}} - \sqrt{\boldsymbol{W}_{\{\mathcal{I}\}}^{-1}}\boldsymbol{H}_{\mathcal{I}}x)^\top \\ & (\sqrt{\boldsymbol{W}_{\{\mathcal{I}\}}^{-1}}y_{\mathcal{I}} - \sqrt{\boldsymbol{W}_{\{\mathcal{I}\}}^{-1}}\boldsymbol{H}_{\mathcal{I}}x), \end{aligned}$$

which holds since $\boldsymbol{W}_{\{\mathcal{I}\}}^{-1}$ is a diagonal matrix. For simplicity of notation, in the remainder of this proof, we let $y_w = \sqrt{\boldsymbol{W}_{\{\mathcal{I}\}}^{-1}}y_{\mathcal{I}}$ and $H_w = \sqrt{\boldsymbol{W}_{\{\mathcal{I}\}}^{-1}}\boldsymbol{H}_{\mathcal{I}}$. By orthogonally projecting $y_w$ onto the range of $H_w$ using $H_wH_w^+$, where $H_w^+$ is the pseudo-inverse of $H_w$, one obtains

$$\begin{aligned} d_x(y,\mathcal{I}) =& \frac{1}{2}(y_w - H_wH_w^+y_w + H_wH_w^+y_w - H_wx)^\top \\ & (y_w - H_wH_w^+y_w + H_wH_w^+y_w - H_wx). \end{aligned} \quad (53)$$

Notice that $(y_w - H_wH_w^+y_w)$ is orthogonal to $(H_wH_w^+y_w - H_wx)$. Furthermore, $W_i > 0$ for each $i$, and by Assumption 3, $\boldsymbol{H}_{\mathcal{I}}$ is of full column rank for any $\mathcal{I}$ with $|\mathcal{I}| \geq m - 2q$, $H_w$ is thus full column rank and $H_w^+ = (H_w^\top H_w)^{-1}H_w^\top$. One then obtains (17) from (53). The proof is thus complete. □

*Proof of Lemma 5.* It follows readily from [36, Lemma 2.8] that a ball $\mathcal{B}_\upsilon(c) \subseteq \mathbb{R}^n$ covers a full dimensional ellipsoid $\{x : (x-x_0)^\top Q(x-x_0) \leq \delta^2\}$, where $Q \in \mathbb{R}^{n\times n}$ is positive definite and $\delta > 0$, if and only if there exists $\tau \geq 0$ such that

$$\tau \begin{bmatrix} Q & -Qx_0 & 0 \\ * & x_0^\top Qx_0 - \delta^2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \succcurlyeq \begin{bmatrix} \boldsymbol{I}_n & -c & 0 \\ -c^\top & -\upsilon^2 & c^\top \\ 0 & c & -\boldsymbol{I}_n \end{bmatrix}. \quad (54)$$

Also, it is clear that a ball $\mathcal{B}_\upsilon(c) \subseteq \mathbb{R}^n$ covers a point $x \in \mathbb{R}^n$ if and only if

$$(x-c)^\top(x-c) \leq \upsilon^2,$$

which, by Schur complement, is equivalent to

$$\begin{bmatrix} \upsilon^2 & (x-c)^\top \\ * & \boldsymbol{I}_n \end{bmatrix} \succcurlyeq 0. \quad (55)$$

Furthermore, for any $\phi$ such that $\mathfrak{I}(\phi)$ is not empty, the set $\mathcal{X}(y,\phi)$ is a union of some full dimensional ellipsoids (when the set $\mathfrak{I}_+(\phi)$ is not empty) and some single points (when the set $\mathfrak{I}_0(\phi)$ is not empty). Therefore, one can conclude Lemma 5. □

*Proof of Lemma 6.* It holds that $\mathcal{X}(y,\phi_0) \subseteq \mathcal{X}(y,\phi_1)$ for any $y$ and $\phi_0 \leq \phi_1$. Therefore, $\mathrm{rad}(\mathcal{X}(y,\phi))$ is monotonically increasing w.r.t. $\phi$.

Let $\mathrm{res}_{[i]}$ be the $i$-th item of the set $\{\mathrm{res}(\mathcal{I}) : \mathcal{I} \subseteq \mathcal{M}, |\mathcal{I}| = m-q\}$ sorted in an ascending order. Then by viewing $\mathcal{X}(y,\phi)$ as a union of ellipsoids as in (21), one obtains that for any $\phi_0 \in \bigcup_{i=1}^{\binom{m}{q}-1}\left(\mathrm{res}_{[i]}, \mathrm{res}_{[i+1]}\right) \bigcup \left(\mathrm{res}_{[\binom{m}{q}]}, \infty\right)$, where $\binom{m}{q}$ is the binomial coefficient, the following holds:

$$\lim_{\phi\to\phi_0} \mathcal{X}(y,\phi) = \mathcal{X}(y,\phi_0),$$

and for any $\phi_0 \in \bigcup_{i=1}^{\binom{m}{q}} \mathrm{res}_{[i]}$,

$$\lim_{\phi\to\phi_0^+} \mathcal{X}(y,\phi) = \mathcal{X}(y,\phi_0)$$

holds, where $\phi \to \phi_0^+$ means that $|\phi - \phi_0| \to 0$ and $\phi - \phi_0 > 0$. Notice also that $\mathrm{rad}(\mathcal{X}(y,\phi)) = 0$ for all $\phi \leq \mathrm{res}_{[1]}$. Therefore, one can conclude the first two bullets of Lemma 6.

By the first two bullets, in order to obtain the third one, it suffices to show that when $\phi$ is in any of the $\binom{m}{q}$ intervals $\bigcup_{i=1}^{\binom{m}{q}-1}\left(\mathrm{res}_{[i]}, \mathrm{res}_{[i+1]}\right) \bigcup \left(\mathrm{res}_{[\binom{m}{q}]}, \infty\right)$, $\mathrm{rad}(\mathcal{X}(y,\phi))$ is strictly increasing w.r.t. $\phi$. Notice that when $\phi$ is in any of these intervals, $\mathfrak{I}_0(\phi)$ is empty and $\mathfrak{I}_+(\phi)$ remains the same, and, therefore, the optimal solution $\psi^*$ (i.e., the square of $\mathrm{rad}(\mathcal{X}(y,\phi))$) to the optimization problem in Lemma 5 is strictly increasing w.r.t. $\phi$. The third bullet is thus concluded and, therefore, the proof is complete. □

## APPENDIX G
## PROOF OF LEMMA 7

Let $x^*, s^*$ be the optimal solution to the optimization problem (26). Further let $x \in \mathbb{R}^n$ be any vector and $\mathcal{I}_0, \mathcal{I}_1$ the two index sets such that $\mathcal{I}_0 \cup \mathcal{I}_1 = \mathrm{supp}(s^*)$, $|\mathcal{I}_0| \leq q$, and $|\mathcal{I}_1| \leq q$. We then construct the following three quantities $x_0, x_1 \in \mathbb{R}^n$ and $y^* \in \mathbb{R}^m$:

$$x_0 = x - \delta x^*, \;\; x_1 = x + \delta x^*, \quad (56)$$

$$y^*_{\mathcal{M}\setminus\mathrm{supp}(s^*)} = (\boldsymbol{H}x)_{\mathcal{M}\setminus\mathrm{supp}(s^*)}, \quad (57)$$

$$y^*_{\mathcal{I}_0} = (\boldsymbol{H}x_0)_{\mathcal{I}_0}, \;\; y^*_{\mathcal{I}_1} = (\boldsymbol{H}x_1)_{\mathcal{I}_1}. \quad (58)$$

Then one verifies that

$$\|x_0 - x_1\|_2 = 2\delta, \quad (59)$$

$$d_{x_0}(y^*) \leq \bar{u}(\delta), \quad (60)$$

$$d_{x_1}(y^*) \leq \bar{u}(\delta). \quad (61)$$

Notice that (60) holds because by the definition of $d_{x_0}(y^*)$ (i.e., the optimal value of (8)), we have

$$\begin{aligned} d_{x_0}(y^*) \leq& \frac{1}{2}\sum_{i=1}^m (y_i^* - H_ix_0 + a_i)^2/W_i, \\ =& \bar{u}(\delta) \end{aligned}$$

where $a_i = -H_i(x_1 - x_0)$ for $i \in \mathcal{I}_1$ and $a_i = 0$ for $i \in \mathcal{M}\setminus\mathcal{I}_1$. The equation (61) can be obtained in the same manner.

Therefore, $x_0, x_1 \in \mathcal{X}(y^*, \bar{u}(\delta))$ by (60) and (61). Combining (59), one then obtains that

$$\mathrm{rad}(\mathcal{X}(y^*, \bar{u}(\delta))) \geq \delta. \quad (62)$$

Notice that since $\bar{u}(\delta) = \delta^2\bar{u}(1)$, we have for any $\epsilon, \delta > 0$,

$$\bar{u}(\delta) + \epsilon = \bar{u}(\delta'),$$

where $\delta' = \delta\sqrt{(\bar{u}(\delta)+\epsilon)/\bar{u}(\delta)} > \delta$. Then using the same construction technique as in (56)-(58), one concludes that, by (62), for any $\epsilon > 0$, there exists $y \in \mathbb{R}^m$ such that

$$\begin{aligned} \mathrm{rad}(\mathcal{X}(y, \bar{u}(\delta)+\epsilon)) =& \mathrm{rad}(\mathcal{X}(y, \bar{u}(\delta'))) \\ \geq& \delta' > \delta. \end{aligned}$$

Therefore, from the definition of $u(\delta)$, one obtains that $u(\delta) \leq \bar{u}(\delta)$ for any $\delta$. The proof is thus complete.

## APPENDIX H
## PROOF OF THEOREM 4

The proof is divided into two parts.

*Part I.* We show that for every $y \in \mathbb{R}^m$,

$$d_x(y) \geq \bar{u}(|x - \mathrm{trm}(y)|) \tag{63}$$

holds for every $x \in \mathbb{R}$, where recall that $\bar{u}(\delta)$ is the upper bound in Lemma 7.

Since the sensors are homogeneous, then without loss of generality, we let $W_i = W/2$ for any $1 \leq i \leq m$. Then one obtains that

$$\bar{u}(\delta) = (m - 2q)\delta^2/W.$$

One further obtains that for any $y \in \mathbb{R}^m$ and $x \in \mathbb{R}$,

$$\begin{aligned} d_x(y) &\geq \sum_{i=q+1}^{m-q} (y_{[i]} - x)^2/W \\ &\geq (m-2q)(\frac{1}{m-2q}\sum_{i=q+1}^{m-q} y_{[i]} - x)^2/W \\ &= \bar{u}(|x - \mathrm{trm}(y)|). \end{aligned}$$

Therefore, (63) holds.

*Part II.* Notice that, for any $x, \delta$ and $k$, if $y \in \mathcal{Y}(f^{\mathrm{trm}}, x, \delta, k)$, $|x - \mathrm{trm}(y)| \geq \delta$ holds, where $\mathcal{Y}(\cdot,\cdot,\cdot,\cdot)$ is introduced in Definition 6. Then (63) yields that $d_x(y) \geq \bar{u}(\delta)$ for every $y \in \mathcal{Y}(f^{\mathrm{trm}}, x, \delta, k)$.

Using the same argument as in the proof of Lemma 13, one obtains that for any $\delta$,

$$r(f^{\mathrm{trm}}, \delta) \geq \bar{u}(\delta) \geq u(\delta).$$

Due to the optimality of $u(\delta)$, the above equation holds as equality. The proof is thus complete.

## REFERENCES


[1] Y. Mo, T. H.-J. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli, "Cyber–physical security of a smart grid infrastructure," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 195–209, 2012.

[2] A. Teixeira, K. C. Sou, H. Sandberg, and K. H. Johansson, "Secure control systems: A quantitative risk management approach," *IEEE Control Systems*, vol. 35, no. 1, pp. 24–45, 2015.

[3] F. R. Hampel, "The influence curve and its role in robust estimation," *Journal of the American Statistical Association*, vol. 69, no. 346, pp. 383–393, 1974.

[4] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," *Proceedings of the IEEE*, vol. 73, no. 3, pp. 433–481, 1985.

[5] P. J. Huber, *Robust statistics*. Springer, 2011.

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

[7] M. Pajic, I. Lee, and G. J. Pappas, "Attack-resilient state estimation for noisy dynamical systems," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 82–92, 2017.

[8] S. Mishra, Y. Shoukry, N. Karamchandani, S. N. Diggavi, and P. Tabuada, "Secure state estimation against sensor attacks in the presence of noise," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 49–59, 2017.

[9] Y. Mo and E. Garone, "Secure dynamic state estimation via local estimators," in *IEEE 55th Conference on Decision and Control (CDC)*. IEEE, 2016, pp. 5073–5078.

[10] X. Liu, Y. Mo, and E. Garone, "Secure dynamic state estimation by decomposing kalman filter," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 7351–7356, 2017.

[11] F. C. Schweppe and E. J. Handschin, "Static state estimation in electric power systems," *Proceedings of the IEEE*, vol. 62, no. 7, pp. 972–982, 1974.

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

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

[14] Y. Mo, T. H.-J. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli, "Cyber–physical security of a smart grid infrastructure," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 195–209, 2011.

[15] J. M. Hendrickx, K. H. Johansson, R. M. Jungers, H. Sandberg, and K. C. Sou, "Efficient computations of a security index for false data attacks in power networks," *IEEE Transactions on Automatic Control*, vol. 59, no. 12, pp. 3194–3208, 2014.

[16] K. C. Sou, H. Sandberg, and K. H. Johansson, "Data attack isolation in power networks using secure voltage magnitude measurements," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 14–28, 2014.

[17] Y. Mo and B. Sinopoli, "Secure estimation in the presence of integrity attacks," *IEEE Transactions on Automatic Control*, vol. 60, no. 4, pp. 1145–1151, 2015.

[18] D. Han, Y. Mo, and L. Xie, "Convex optimization based state estimation against sparse integrity attacks," *IEEE Transactions on Automatic Control*, vol. 64, no. 6, pp. 2383–2395, 2019.

[19] X. Ren, Y. Mo, and K. H. Johansson, "Secure static state estimation: A large deviation approach," *IFAC-PapersOnLine*, vol. 51, no. 23, pp. 289–294, 2018.

[20] G. Fellouris, E. Bayraktar, and L. Lai, "Efficient byzantine sequential change detection," *IEEE Transactions on Information Theory*, vol. 64, no. 5, pp. 3346–3360, 2017.

[21] X. Ren, J. Yan, and Y. Mo, "Binary hypothesis testing with Byzantine sensors: Fundamental trade-off between security and efficiency," *IEEE Transactions on Signal Processing*, vol. 66, no. 6, pp. 1454–1468, March 2018.

[22] S. Marano, V. Matta, and L. Tong, "Distributed detection in the presence of byzantine attacks," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 16–29, 2008.

[23] R. S. Smith, "A decoupled feedback structure for covertly appropriating networked control systems," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 90–95, 2011.

[24] C. E. Shannon, "Communication theory of secrecy systems," *Bell Labs Technical Journal*, vol. 28, no. 4, pp. 656–715, 1949.

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

[26] R. B. Holmes, *A course on optimization and best approximation*. Springer, 2006, vol. 257.

[27] Y. Mo and R. M. Murray, "Multi-dimensional state estimation in adversarial environment," in *Control Conference (CCC), 2015 34th Chinese*. IEEE, 2015, pp. 4761–4766.

[28] E. De Klerk, *Aspects of semidefinite programming: interior point algorithms and selected applications*. Springer Science & Business Media, 2006, vol. 65.

[29] M. S. Chong, M. Wakaiki, and J. P. Hespanha, "Observability of linear systems under adversarial attacks," in *2015 American Control Conference (ACC)*. IEEE, 2015, pp. 2439–2444.

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

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

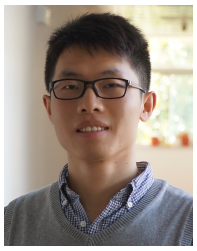
[32] Y. Nakahira and Y. Mo, "Attack-resilient $h_2, h_\infty$, and $l_1$ state estimator," *IEEE Transactions on Automatic Control*, vol. 63, no. 12, pp. 4353–4360, 2018.

[33] L. Danzer and V. Klee, *Helly's theorem and its relatives*. American Mathematical Society Providence, RI, 1963.

[34] W. Rudin, *Principles of mathematical analysis*. McGraw-hill New York, 1964, vol. 3.

[35] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*. Springer Science & Business Media, 2009, vol. 38.

[36] E. A. Yildirim, "On the minimum volume covering ellipsoid of ellipsoids," *SIAM Journal on Optimization*, vol. 17, no. 3, pp. 621–641, 2006.



**Xiaoqiang Ren** is a professor at the School of Mechatronic Engineering and Automation, Shanghai University, China. He received the B.E. degree in Automation from Zhejiang University, Hangzhou, China, in 2012 and the Ph.D. degree in control and dynamic systems from Hong Kong University of Science and Technology in 2016. Prior to his current position, he was a postdoctoral researcher in the Hong Kong University of Science and Technology in 2016, Nanyang Technological University from 2016 to 2018, and KTH Royal Institute of Technology from 2018 to 2019. His research interests include security of cyber-physical systems, sequential decision, and networked estimation and control.

**Yilin Mo** is an Associate Professor in the Department of Automation, Tsinghua University. He received his Ph.D. In Electrical and Computer Engineering from Carnegie Mellon University in 2012 and his Bachelor of Engineering degree from Department of Automation, Tsinghua University in 2007. Prior to his current position, he was a postdoctoral scholar at Carnegie Mellon University in 2013 and California Institute of Technology from 2013 to 2015. He held an assistant professor position in the School of Electrical and Electronic Engineering at Nanyang Technological University from 2015 to 2018. His research interests include secure control systems and networked control systems, with applications in sensor networks and power grids.

**Jie Chen (S'87-M'89-SM'98-F'07)** is a Chair Professor in the Department of Electronic Engineering, City University of Hong Kong, Hong Kong, China. He received the B.S. degree in aerospace engineering from Northwestern Polytechnic University, Xian, China in 1982, the M.S.E. degree in electrical engineering, the M.A. degree in mathematics, and the Ph.D. degree in electrical engineering, all from The University of Michigan, Ann Arbor, Michigan, in 1985, 1987, and 1990, respectively. Prior to joining City University, he was University of California, Riverside, California, from 1994 to 2014, where he was a Professor and served as Professor and Chair for the Department of Electrical Engineering from 2001 to 2006. His main research interests are in the areas of linear multivariable systems theory, system identification, robust control, optimization, time-delay systems, networked control, and multi-agent systems. He is a Fellow of IEEE, a Fellow of AAAS, a Fellow of IFAC, and an IEEE Distinguished Lecturer. He currently serves as an Associate Editor for *SIAM Journal on Control and Optimization*, and *International Journal of Robust and Nonlinear Control*.

**Karl Henrik Johansson** is Professor at the School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Sweden. He received MSc and PhD degrees from Lund University. He has held visiting positions at UC Berkeley, Caltech, NTU, HKUST Institute of Advanced Studies, and NTNU. His research interests are in networked control systems, cyber-physical systems, and applications in transportation, energy, and automation. He is a member of the Swedish Research Council's Scientific Council for Natural Sciences and Engineering Sciences. He has served on the IEEE Control Systems Society Board of Governors, the IFAC Executive Board, and is currently Vice-President of the European Control Association Council. He has received several best paper awards and other distinctions from IEEE, IFAC, and ACM. He has been awarded Distinguished Professor with the Swedish Research Council and Wallenberg Scholar with the Knut and Alice Wallenberg Foundation. He has received the Future Research Leader Award from the Swedish Foundation for Strategic Research and the triennial Young Author Prize from IFAC. He is Fellow of the IEEE and the Royal Swedish Academy of Engineering Sciences, and he is IEEE Control Systems Society Distinguished Lecturer.