Resilient Observer Design for Cyber-Physical Systems with Data-driven Measurement Pruning

Yu Zheng, Olugbenga Moses Anubi

Abstract Resilient observer design for Cyber-Physical Systems (CPS) in the presence of adversarial false data injection attacks (FDIA) is an active area of research. Existing state-of-the-art algorithms tend to breakdown as more and more knowledge of the system is built into the attack model; also as the percentage of attacked nodes increases. From the view of optimization theory, the problem is often cast as a classical error correction problem for which a theoretical limit of 50% has been established as the maximum percentage attacked nodes for which state recovery is guaranteed. Beyond this limit, the performance of ℓ_1-minimization based schemes, for instance, deteriorates rapidly. Similar performance degradation occurs for other types of resilient observers beyond certain percentages of attacked nodes.

In order to increase the corresponding percentage attacked nodes for which state recoveries can be guaranteed, researchers have begun to incorporate prior information into the underlying resilient observer design framework. For the most pragmatic cases, this prior information is often obtained through a data-driven Machine Learning process. Existing results have shown strong positive correlation between the maximum attacked percentages that can be tolerated and the accuracy of the data-driven model. Motivated by these results, this chapter examines the case for *pruning algorithms* designed to improve the *Positive Prediction Value* (PPV) of the resulting prior information, given a stochastic uncertainty characteristics of the underlying Machine Learning model. Theoretical quantification of the achievable improvement is given. Simulation results show that the pruning algorithm significantly increases the maximum correctable percentage of attacked nodes, even for Machine Learning model whose prediction power is comparable to random flip of a coin.

Yu Zheng
FAMU-FSU College of Engineering, Tallahassee, FL 32310, USA, e-mail: yz19b@fsu.edu

Olugbenga Moses Anubi
FAMU-FSU College of Engineering, Tallahassee, FL 32310, USA, e-mail: oanubi@fsu.edu

1 Notation

The following notations and definitions are used throughout the whole paper: $\mathbb{R}, \mathbb{R}^n, \mathbb{R}^{n\times m}$ denote the space of real numbers, real vectors of length n and real matrices of n rows and m columns respectively. $\mathbb{R}_+$ denotes the space of positive real numbers. Normal-face lower-case letters ($e.g.\ x \in \mathbb{R}$) are used to represent real scalars, bold-face lower-case letters ($e.g.\ \mathbf{x} \in \mathbb{R}^n$) represent vectors, while normal-face upper-case letters ($e.g.\ X \in \mathbb{R}^{n\times m}$) represent matrices. $X^\top$ denotes the transpose of matrix X. $\mathbf{1}_n$ and I_n denote vector of ones and identity matrix of size n respectively. Let $\mathscr{T} \subseteq \{1,\dots,n\}$, then for a matrix $X \in \mathbb{R}^{m\times n}$, $X_{\mathscr{T}} \in \mathbb{R}^{|\mathscr{T}|\times n}$ is the sub-matrix obtained by extracting the rows of X corresponding to the indices in $\mathscr{T}$. $\mathscr{T}^c$ denotes the complement of a set $\mathscr{T}$, and the universal set on which it is defined will be clear from the context. The support of a vector $\mathbf{x} \in \mathbb{R}^n$, a set of the indices of nonzero entries, is denoted by:

$$\mathsf{supp}(\mathbf{x}) \triangleq \{i \subseteq \{1,\dots,n\} | \mathbf{x}_i \neq 0\}.$$

If $|\mathsf{supp}(\mathbf{x})| = k$, we say $\mathbf{x}$ is a k-sparse vector. Moreover, $\Sigma_k \subset \mathbb{R}^n$ denotes the set of all k-sparse vectors in $\mathbb{R}^n$. The operator $\mathsf{argsort}\downarrow(\mathbf{x})$ denotes a function that returns the sorted indices of vector $\mathbf{x}$ in descending order of the magnitude of $\mathbf{x}_i$. The symbol & denotes logical "AND" operator. The symbol $*$ denotes the convolution operator for vectors. The symbol $\circ$ denotes element-wise multiplication of two vectors, $\mathbf{z} = \mathbf{x}\circ\mathbf{y} \Rightarrow \mathbf{z}_i = \mathbf{x}_i\mathbf{y}_i$. The expression $x \sim \mathscr{B}(1,p)$ means that random variable x follows the Bernoulli distribution with $\mathsf{Pr}\{x=1\} = p$. The weighted 1-norm of a vector $\mathbf{z} \in \mathbb{R}^n$ with the weight vector $\mathbf{w} \in \mathbb{R}^n$ is given by:

$$\|\mathbf{z}\|_{1,\mathbf{w}} \triangleq \sum_{i=1}^{n} \mathbf{w}_i\mathbf{z}_i.$$

2 Introduction

As the backbone of future critical infrastructures, Cyber-physical Systems (CPS) are complicated integration of computation, communication, and physical components. Security, within the context of CPSs, poses more challenges compared to both traditional information technology (IT) security and operational technology (OT) security due to the temporal dynamics brought by physical environment and the heterogeneous nature of operation of CPSs [1]. In the context of CPS, failures induced by malicious attacks are beyond well studied random failures in reliability engineering or well-defined uncertainty classes in robust control. Moreover, the coupling of computation and communication with distributed sensing and actuation components increases the vulnerability to attacks [2, 3, 4, 5].

The control design for CPSs usually comprises of an observer to estimate the states of the physical system and a controller to compute control commands based

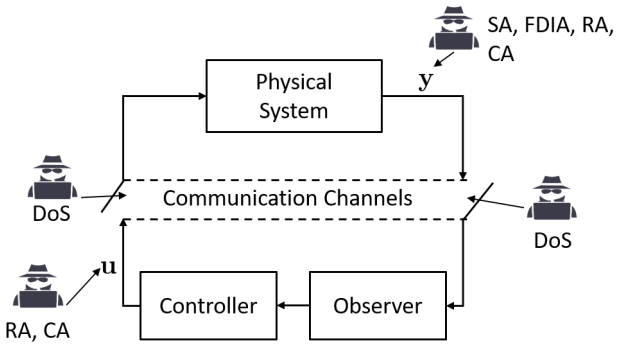

Fig. 1 Locations of Attacks in CPS in the context of security-control (SA: stealth attack, CA: covert attack, RA: replay attack, FDIA: false data injection attack, DoS: detial of service)

on the state estimation. Thus, the control system receives diverse information from measurement substations and distribute the computed control commands to a number of actuators through a communication network [6]. Thus, an elaborate attack on a CPS can be designed by considering the networked closed-loop interaction between the cyber and physical agents. Furthermore, the disperse geographical distribution and abundance of unmanned facilities also provide malicious attackers the opportunity to construct coordinated attacks. These attacks, studied extensively in literature, either targets the system integrity [7], such as stealth attacks [12], replay attacks [14], covert attacks [15] and false data injection attacks (FDIA) [16] or the availability [7], such as dential of serives (DoS) [8]. The locations of those attacks are shown in Fig. 1. It was shown in [9, 10, 11], that if FDIA is defined properly, it can exploit certain underlying vulnerabilities of bad data detection (BDD) schemes in order to force an erroneous state estimation using sparse measurement corruption. Consequently, in this chapter, we consider the resiliency of a class of observers against FDIA. If the observer estimates, using compromised measurements, are close to the true states, then control performance can be guaranteed with any control design which is robust to estimation error.

One of the pioneering works on resilient observer was presented in [17], where an unconstrained ℓ_1 observer was proposed to achieve exact state recovery. A necessary and sufficient condition for exact recovery is that less than half of the systems measurements be compromised. The authors in [18] proved this condition from an interesting aspect of s-sparse observability and proposed a event-triggered Luenberger observer against FDIA. In [19], the authors presented a more systematical work on the observability of the linear system under attacks and proposed a Gramian-based estimator. The authors in [20] and the authors in [21] both considered resilient estimation in the presence of noise and attacks at the same time and constructed ℓ_1-ℓ_2 observers. The authors in [22] considered robust estimation scheme against FDIA, in which local robust estimators and global fusion are combined to achieve resilient-robust estimation. Readers can also refer to [23] for feasible resilient estimation

methods by Satisfiability Modulo Theory (SMT) solvers. However, all the above observers would not achieve successful resilient estimation when 50%, or more, of system measurements are attacked. Equivalently, the system is not $2k$-detectable, where k is the number of attacks. This is a significant limitation, since it requires that there be twice as much as needed measurement stations installed for a CPS and the system has to be observable for every combination of 50% of the total sensors. This is a property that is currently not achieved by most critical cyber-physical critical infrastructures like the power grid.

In order to increase the corresponding percentage of attacked nodes for which state recoveries can be guaranteed, researchers have begun to incorporate prior information into the underlying resilient observer design framework. There are mainly three kinds of prior information considered in literature: state prior [24], measurement prior [26, 25], support prior [27, 16]. In [24], three types of state prior was discussed: sparsity information of the estimated states, $(\alpha, \overline{n}_0)$ sparsity information, where the estimated states are assumed to have α instead of 0 in the sparsity form, and side information which is the knowledge of the initial states from the physical attribution of the system and cannot be manipulated by malicious agents. Although the resiliency of the observer can be improved with such knowledge of the states of system, it is very difficult to obtain such information in practice. This will require prior determination of the state distribution for all operating conditions of an uncertain, large-scale nonlinear, and sometimes hybrid system!

Support prior is the estimated information of attack locations, which can be given by some data-driven localization algorithm or learning-based anomaly detection methods, such as watermark-based methods [28], moving-target based approach [29], distributed support vector machine [30], deep learning neural network [31], and many more [32, 33, 34, 35]. Although the localization algorithms can be readily defined and are very useful for monitoring purposes, using this kind of support prior for resilient estimation has two main drawbacks; imprecise classification and high training price. This limits their applicability in piratical purposes. In this chapter, we examine a class of pruning methods to generate a feasible pruned support prior with predetermined precision guarantees. Coupling the pruning algorithm with any localization algorithm can significantly improve the resulting precision, which directly improves the resiliency of the underlying resilient estimation process. This means less precise localization algorithm can be tolerated, thus slashing the required training price. The initial pruning idea was introduced in [27], analyzed and improved in [36]. In this chapter, a more detailed mathematical foundation is given, in addition to an improved implementation.

Measurement prior is a collection of additional auxiliary information about system measurements that is unknown to the malicious attackers. A direct use of measurement prior in resilient observer design was shown in [25, 26, 37] to improve the limit of the percentage of compromised measurement for which exact recovery is guaranteed from 50% to 80%. Also, the watermark-based detection approaches [28] and moving-average detection approaches [29] both use the additional information in an authentication layer in order to detect the attacks. Thus, against measurement attacks, the measurement prior and support prior are related. An advantage of mea-

surement prior is its expansibility to the authentication layer. The more measurement priors that can be constructed usually provides better detection precision. In this chapter, we will utilize a measurement prior constructed by using data-driven auxiliary model between auxiliary variables and the system measurements. The attacked measurements will then be detected if they cannot be explained by both the system dynamics and the measurement model prior with high likelihood, thus reducing the resulting attack surface.

The reminder of this chapter is organized as follows: In Section 3, concurrent models of CPS, including physical model, monitoring model, thread model, prior model and pruning algorithm are given; in Section 4, the resilient observer design with data-driven measurement pruning is given; in Section 5, numerical simulation and application examples are given to demonstrate the performance of the designed observer compared to other resilient observers in severe adversarial environment; concluding remarks follow in Section 6.

3 Concurrent Models

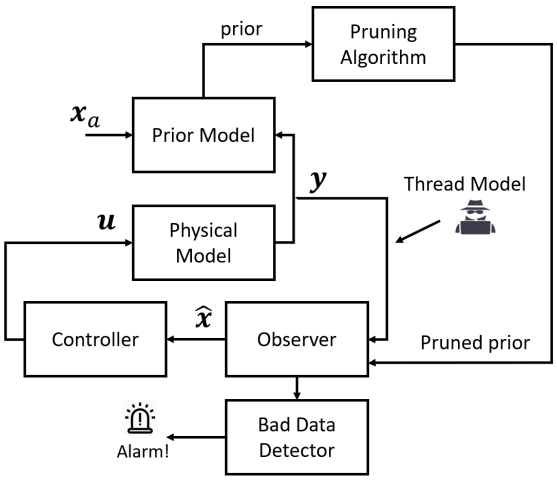

Fig. 2 Concurrent model on CPS ($\mathbf{x}_a$ is an auxiliary state used in the prior model)

To discuss the resilient observer design, relevant model developments are discussed in this section. Since CPS is a seamless integration of computational, physical and communication network systems, a single-layer model cannot sufficiently describe the complex characteristics of CPS. Also, as a closed-loop system, separately and independently modeling the seperate layers cannot capture the tight interaction between the cyber and physical layers [38]. Concurrent modeling has been used a good way to describe the complex operation on CPS [39], where different

models in different hierarchy work concurrently. As shown in Fig. 2, this small version of CPS has four concurrent loops; the physical dynamical loop, prior generation loop, monitoring path, and attack injection.

The rest of the subsections are dedicated to discuss, in more details, the modeling aspects for each layer, the underlying assumptions and connections with the subsequent resilient observer design.

3.1 Physical Model and Monitor

A linear time invariant (LTI) model is considered to describe the physical behaviour of the CPS in Fig. 2.

$$\begin{aligned} \mathbf{x}_{i+1} &= A\mathbf{x}_i \\ \mathbf{y}_i &= C\mathbf{x}_i + \mathbf{e}_i, \end{aligned} \tag{1}$$

where $\mathbf{x}_i \in \mathbb{R}^n$ is the state vector at time i, $\mathbf{y}_i \in \mathbb{R}^m$ is the measurement vector, $\mathbf{e}_i \in \mathbb{R}^m$ is the time-varying attack-noise vector. The measurement attacks and noises are modelled as additional error signals. Control inputs may be included in the model above. However, since control inputs are generally irrelevant to state estimation problems, we suppress it in the model considered here.

The following assumptions are used in subsequent developments:

1. The pair (A,C) is observable.
2. The measurements are redundant ($m > n$).
3. The attack signal is possibly unbounded and sparse, $\mathbf{e}_i \in \Sigma_k$ for some $k < m$.
4. The attack-free part of $\mathbf{e}_i$ is bounded, $\sum_{i \in \mathscr{T}^c} |\mathbf{e}_i| < \varepsilon$, for some $\varepsilon > 0$.

By iterating the system model (1) T time steps backwards, the T-horizon observation model is given by

$$\mathbf{y}_T = H\mathbf{x}_{i-T+1} + \mathbf{e}_T, \tag{2}$$

where $\mathbf{y}_T = [\mathbf{y}_i^\top \;\; \mathbf{y}_{i-1}^\top \cdots \mathbf{y}_{i-T+1}^\top]^\top \in \mathbb{R}^{Tm}$ is a sequence of observation in the moving window $[i-T+1 \;\; i]$, $\mathbf{x}_{i-T+1} \in \mathbb{R}^n$ is the state vector at time $i-T+1$, $\mathbf{e}_T = [\mathbf{e}_i^\top \;\; \mathbf{e}_{i-1}^\top \cdots \mathbf{e}_{i-T+1}^\top]^\top$ is the sequence of attack-noise vectors in the same moving window, $H \in \mathbb{R}^{Tm \times n}$ is the observation matrix such that

$$H = \begin{bmatrix} CA^{T-1} \\ \vdots \\ CA \\ C \end{bmatrix}.$$

The following definitions formalize the notions of a decoder and a detector which are used subsequently.

Definition 1 (Decoder). Given an observable pair (A,C) and a horizon parameter T, a decoder $\mathscr{D} : \mathbb{R}^{Tm} \mapsto \mathbb{R}^n$ is an operator given by

$$\hat{\mathbf{x}} = \mathscr{D}(\mathbf{y}_T|H) = \arg\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{y}_T - H\mathbf{x}\|_1, \tag{3}$$

where $\mathbf{y}_T = \{\mathbf{y}_i, \mathbf{y}_{i-1}, \cdots, \mathbf{y}_{i-T+1}\}$ is a moving-windowed measurement vector history and $\hat{\mathbf{x}} \in \mathbb{R}^n$ is the resulting estimated initial state vector $\mathbf{x}_{i-T+1}$. When the parameter is clear from context, they are dropped from the argument list for clarity.

Definition 2 (Detector). Given the measurements $\mathbf{y}_T \in \mathbb{R}^{Tm}$ taken in the moving window $[i-T+1 \;\; i]$, a detector based on the ℓ_1 decoder is mapping of the form:

$$\Psi_T : \{\mathbf{y}_T\} \mapsto \{\Psi_1, \Psi_2\}$$

where, $\Psi_1 \in \{0,1\}$[1] is the first output argument indicating whether or not the measurement $\mathbf{y}_T$ is attacked, $\Psi_2 \in 2^{\{1,2,\cdots,m\}}$ is the second output argument indicating the support of attack locations.

The decoder-detector pair constitutes a monitor scheme for the system (1), as shown below:

Remark 1 (Residual-based Monitor mechanism). Given a threshold value $\varepsilon_0 > 0$, the monitor returns $\Psi_1 = \{0\}$ in the first output argument for a given measurement vector history $\mathbf{y}_T = \{\mathbf{y}_i, \mathbf{y}_{i-1}, \cdots, \mathbf{y}_{i-T+1}\}$ if there exists a corresponding state trajectory $\hat{X}_T = \{\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_{i-1}, \cdots, \hat{\mathbf{x}}_{i-T}\}$ such that

$$\begin{aligned} \|\hat{\mathbf{x}}_{j+1} - A\hat{\mathbf{x}}_j\| &\le \varepsilon_0, \quad j = i-T, \cdots, i-1 \\ \|\mathbf{y}_j - C\hat{\mathbf{x}}_j\| &\le \varepsilon_0, \quad j = i-T+1, \cdots, i. \end{aligned}$$

Otherwise, the monitor returns $\Psi_1 = \{1\}$ in the first output argument and also the support of the sparsest attack trajectory $\mathbf{e}_T = \{\mathbf{e}_i, \mathbf{e}_{i-1}, \cdots, \mathbf{e}_{i-T+1}\}$ such that

$$\begin{aligned} \|\hat{\mathbf{x}}_{j+1} - A\hat{\mathbf{x}}_j\| &\le \varepsilon_0, \quad j = i-T, \cdots, i-1 \\ \|\mathbf{y}_j - C\hat{\mathbf{x}}_j - \mathbf{e}_j\| &\le \varepsilon_0, \quad j = i-T+1, \cdots, i. \end{aligned}$$

3.2 Threat Model

Following the setup above, we give a formal definition of successful false data injection attack (FDIA) and prescribe conditions under which a FDIA will successfully corrupt a decoder while evading detection by the residual-based monitor. To design a successful FDIA, the following assumptions are made, which are widely used in literature [11, 40]:

1. The attacker has perfect knowledge of the system dynamics in (1)
2. The attacker can inject arbitrary bias at the compromised nodes $\mathscr{T} \subset \{1, \cdots, m\}$.

[1] 0: safe, 1: unsafe

3. The number of nodes the attacker can simultaneously compromise at any given time is bounded. In other words, the attackers has limited resources.

Definition 3 (Successful FDIA [11]). Consider the CPS in (1) and the corresponding measurement model (2). Given a positive integer $k < m$, the attack sequence $\mathbf{e}_T \in \Sigma_{Tk}$ is said to be (ε, α)-successful against the decoder-detection pair described above if

$$\|\mathbf{x}^\star - \mathscr{D}(\mathbf{y}_T)\|_2 \geq \alpha, \text{ and } \|\mathbf{y}_T - H\mathscr{D}(\mathbf{y}_T)\|_2 \leq \varepsilon, \tag{4}$$

where $\mathbf{y}_T = \mathbf{y}_T^\star + \mathbf{e}_T$ with $\mathbf{y}_T^* \in \mathbb{R}^{Tm}$ being the true measurement vector, and $\mathbf{x}^\star$ is the true state vector.

In the above definition, k quantifies the attack sparsity level per time. Specially, it is the maximum number of attacks at each time index. Given the support sequence $\mathscr{T} = \{\mathscr{T}_i \ \ \mathscr{T}_{i-1} \cdots \mathscr{T}_{i-T+1}\}$ with $|\mathscr{T}_i| \leq k$. Let $\mathbf{x}_e$ be an optimal solution of the optimization program

$$\begin{aligned} \text{Maxmize}: &\quad \|H_{\mathscr{T}}\mathbf{x}\|_1, \\ \text{Subject to}: &\quad \|H_{\mathscr{T}^c}\mathbf{x}\|_1 \leq \varepsilon. \end{aligned} \tag{5}$$

Then a FDIA can be defined as

$$\mathbf{e}_{\mathscr{T}} = H_{\mathscr{T}}\mathbf{x}_e, \ \ \mathbf{e}_{\mathscr{T}^c} = \mathbf{0}. \tag{6}$$

The following theorem shows the condition under which the defined FDIA above is (ε, α)-successful for the given attack support $\mathscr{T}$.

Theorem 1. *Suppose there exists a vector* $\mathbf{w} \in \mathsf{range}(H)$ *such that*

$$\|\mathbf{w}_{\mathscr{T}}\|_1 > \|\mathbf{w}_{\mathscr{T}^c}\|_1, \tag{7}$$

then the FDIA in (6) *is* (ε, α)*-successful against the decoder-detector pair in Definition 1 and Remark 1 for all*

$$\alpha \leq \frac{\sigma_1 - 1}{\sqrt{|\mathscr{T}|}\overline{\sigma}_{\mathscr{T}} - \underline{\sigma}_{\mathscr{T}^c}}\varepsilon,$$

with $|\mathscr{T}| > \frac{\underline{\sigma}^2_{\mathscr{T}^c}}{\overline{\sigma}^2_{\mathscr{T}}}$*, where* $\overline{\sigma}_{\mathscr{T}}$ *and* $\underline{\sigma}_{\mathscr{T}^c}$ *are the largest and smallest non-zero singular values of* $H_{\mathscr{T}}$ *and* $H_{\mathscr{T}^c}$ *respectively, and*

$$\sigma_1 = \max_{\mathbf{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \frac{\|H_{\mathscr{T}}\mathbf{v}\|_1}{\|H_{\mathscr{T}^c}\mathbf{v}\|_1}$$

Proof. For the optimization problem in (5), let $\mathbf{x} = \alpha\mathbf{v}$, where $\mathbf{v} \in \mathbb{R}^n$ is arbitrary. Then,

$$\|H_{\mathscr{T}}\mathbf{x}_e\|_1 \geq \max_{|\alpha|\|H_{\mathscr{T}^c}\mathbf{v}\|_1 \leq \varepsilon} |\alpha|\|H_{\mathscr{T}}\mathbf{v}\|_1 \geq \frac{\|H_{\mathscr{T}}\mathbf{v}\|_1}{\|H_{\mathscr{T}^c}\mathbf{v}\|_1}\varepsilon.$$

Since $\mathbf{v}$ is arbitrary, it follows that $\|H_{\mathscr{T}}\mathbf{x}_e\|_1 \geq \sigma_1\varepsilon$.

Next, the attacked measurement $\mathbf{y}_T$ can be written as

$$\mathbf{y}_T = H\mathbf{x}^\star + P\begin{bmatrix} H_{\mathscr{T}} \\ 0 \end{bmatrix}\mathbf{x}_e,$$

with an appropriate permutation matrix P satisfying $H = P\begin{bmatrix} H_{\mathscr{T}} \\ H_{\mathscr{T}^c} \end{bmatrix}$, where $\mathbf{x}^\star \in \mathbb{R}^n$ is the unknown true state vector.
Consider a projection

$$\mathbf{x}_1 = \arg\min_{\mathbf{x}\in\mathbb{R}^n} \left\| H\mathbf{x} - P\begin{bmatrix} H_{\mathscr{T}} \\ 0 \end{bmatrix}\mathbf{x}_e \right\|_1.$$

It follows, by the optimality of $\mathbf{x}_1$, that

$$\|H_{\mathscr{T}}(\mathbf{x}_1 - \mathbf{x}_e)\|_1 + \|H_{\mathscr{T}^c}\mathbf{x}_1\|_1 \leq \left\| H\mathbf{x}_e - P\begin{bmatrix} H_{\mathscr{T}} \\ 0 \end{bmatrix}\mathbf{x}_e \right\|_1 \leq \|H_{\mathscr{T}^c}\mathbf{x}_e\|_1 \leq \varepsilon, \tag{8}$$

Using the reverse triangle inequality on $\|H_{\mathscr{T}}(\mathbf{x}_1 - \mathbf{x}_e)\|_1$ yields

$$\|H_{\mathscr{T}}\mathbf{x}_e\|_1 - \|H_{\mathscr{T}}\mathbf{x}_1\|_1 + \|H_{\mathscr{T}^c}\mathbf{x}_1\|_1 \leq \varepsilon,$$

which implies that

$$\begin{aligned} \|H_{\mathscr{T}}\mathbf{x}_1\|_1 - \|H_{\mathscr{T}^c}\mathbf{x}_1\|_1 &\geq \|H_{\mathscr{T}}\mathbf{x}_e\|_1 - \varepsilon \\ \sqrt{|\mathscr{T}|}\|H_{\mathscr{T}}\mathbf{x}_1\|_2 - \|H_{\mathscr{T}^c}\mathbf{x}_1\|_2 &\geq (\sigma_1 - 1)\varepsilon \\ (\sqrt{|\mathscr{T}|}\overline{\sigma}_{\mathscr{T}} - \underline{\sigma}_{\mathscr{T}^c})\|\mathbf{x}_1\|_2 &\geq (\sigma_1 - 1)\varepsilon \\ \|\mathbf{x}_1\|_2 &\geq \frac{(\sigma_1 - 1)\varepsilon}{\sqrt{|\mathscr{T}|}\overline{\sigma}_{\mathscr{T}} - \underline{\sigma}_{\mathscr{T}^c}}. \end{aligned}$$

Since $|\mathscr{T}| > \frac{\sigma^2_{\mathscr{T}^c}}{\overline{\sigma}^2_{\mathscr{T}}}$, the denominator of the right-hand side is positive. Moreover, we can choose a $\mathbf{v}_w \in \mathbb{R}^n$ such that $H\mathbf{v}_w = \mathbf{w}$, where $\mathbf{w}$ satisfies the condition in (7). This implies that there exist a $\mathbf{v}_w \in \mathbb{R}^n$ such that $\sigma_1 \geq \frac{\|H_{\mathscr{T}}\mathbf{v}_w\|_1}{\|H_{\mathscr{T}^c}\mathbf{v}_w\|_1} > 1$. Thus, the given upper bound on $\|\mathbf{x}_1\|_2$ is always positive.

Furthermore, by the decoder in (3),

$$\begin{aligned} \hat{\mathbf{x}} &= \arg\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{y}_T - H\mathbf{x}\|_1 \\ &= \arg\min_{\mathbf{x}\in\mathbb{R}^n} \left\| H(\mathbf{x}^\star - \mathbf{x}) + P\begin{bmatrix} H_{\mathscr{T}} \\ 0 \end{bmatrix}\mathbf{x}_e \right\|_1 = \mathbf{x}^\star + \mathbf{x}_1. \end{aligned}$$

Thus,

$$\|\mathbf{x}^\star - \mathscr{D}(\mathbf{y}_T)\|_2 = \|\mathbf{x}_1\|_2 \geq \frac{(\sigma_1 - 1)\varepsilon}{\sqrt{|\mathscr{T}|}\overline{\sigma}_{\mathscr{T}} - \underline{\sigma}_{\mathscr{T}^c}}.$$

Moreover, the residual is given by

$$\begin{aligned}\|\mathbf{y}_T - H\mathscr{D}(\mathbf{y}_T)\|_2 &= \left\| P\left(\begin{bmatrix} H_{\mathscr{T}} \\ H_{\mathscr{T}^c} \end{bmatrix} \mathbf{x}_1 - \begin{bmatrix} H_{\mathscr{T}} \\ 0 \end{bmatrix} \mathbf{x}_e \right) \right\|_2 \\ &\leq \left\| \begin{bmatrix} H_{\mathscr{T}} \\ H_{\mathscr{T}^c} \end{bmatrix} \mathbf{x}_1 - \begin{bmatrix} H_{\mathscr{T}} \\ 0 \end{bmatrix} \mathbf{x}_e \right\|_1 \\ &\leq \|H_{\mathscr{T}}(\mathbf{x}_1 - \mathbf{x}_e)\|_1 + \|H_{\mathscr{T}^c}\mathbf{x}_1\|_1 \leq \varepsilon.\end{aligned}$$

Thus, by Definition 3, the FDIA in (6) is (ε, α)-successful against the decoder-detector pair in Definition 1 and Remark 1 for all

$$\alpha \leq \frac{\sigma_1 - 1}{\sqrt{|\mathscr{T}|}\overline{\sigma}_{\mathscr{T}} - \underline{\sigma}_{\mathscr{T}^c}}\varepsilon.$$

Remark 2. If, in addition, $\mathsf{null}(H_{\mathscr{T}^c}) \backslash \mathsf{null}(H_{\mathscr{T}}) \neq \emptyset$, let $\mathbf{v}_n \in \mathsf{null}(H_{\mathscr{T}^c}) \backslash \mathsf{null}(H_{\mathscr{T}})$, then $\|H_{\mathscr{T}^c}\mathbf{v}_n\|_1 = 0$ but $\|H_{\mathscr{T}}\mathbf{v}_n\|_1 > 0$. Thus, $\sigma_1 \geq \frac{\|H_{\mathscr{T}}\mathbf{v}\|_1}{\|H_{\mathscr{T}^c}\mathbf{v}\|_1}$ is infinite, which implies that the FDIA in (6) is (ε, α)-successful for all $\varepsilon, \alpha \in \mathbb{R}_+$.

3.3 Data-driven Auxiliary Measurement Prior

In this subsection, we present a data-driven auxiliary measurement prior based on a generative probabilistic regression model constructed using Gaussian process (GP). This prior model is a mapping from the chosen auxiliary variables to the observed measurements, which plays a role of additional authentication layer.

Given a dataset Z, Y, where $Z \in \mathbb{R}^{p\times N}$ is the matrix collecting the auxiliary states columnwise, $Y \in \mathbb{R}^{m\times N}$ is the matrix of the corresponding observed measurements, the goal is to learn the underlying function $f : \mathbb{R}^p \to \mathbb{R}^m$ such that

$$\mathbf{y}_i = f(\mathbf{z}_i) + \varepsilon, \quad i = 1, \cdots, N, \tag{9}$$

where $\varepsilon \sim \mathscr{N}(0, \sigma^2)$. To achieve this goal, certain restrictions have to be made on the properties of the underlying function. Otherwise, all potential functions fitting the training dataset would be equally valid. As a means of regularization, we assume that the underlying function f is restricted to a class defined by a given Gaussian process. A Gaussian process (GP) is a generalization of Gaussian probabilistic distribution [41]. It is a collection of random variables, every finite subset of which are jointly Gaussian [42]. Gaussian process regression (GPR) uses GPs to encode prior distribution over functions f. Thus, suppose $f \in \mathbf{GP}$, then it satisfies the following distribution point-wise:

$$f(\mathbf{z}) \sim \mathscr{N}(m(\mathbf{z}), k(\mathbf{z}, \mathbf{z}')), \tag{10}$$

where $m(\mathbf{z}) = \mathbb{E}[f(\mathbf{z})]$ is the mean function and $k(\mathbf{z},\mathbf{z}') = \mathbb{E}[(f(\mathbf{z}) - m(\mathbf{z}))(f(\mathbf{z}') - m(\mathbf{z}'))^\top]$ is the covariance function encoded, apriori, by the kernel function k. The model of GP contains two parts: a joint distribution model and a kernel function. Kernel functions capture the similarity between the function's (or model's) outputs, for given inputs. The design of kernel function depends on the prior knowledge of the process that generated the data in question. For example, suppose we know that the output of the process changes slowly with respect to change in input, the smoothness prior knowledge can be modeled in the kernel function used by the GP. One of commonly used kernel function is the square exponential covariance function (also called RBF), given by [43]

$$k\left(\mathbf{z},\mathbf{z}'\right) = A\exp\left\{-\frac{\|\mathbf{z}-\mathbf{z}'\|^2}{2l}\right\}, \tag{11}$$

where, the hyperparameters A and l are amplitude coefficient and describing single scaling factor on the influence of nearby observations respectively. For a comprehensive summary of kernel functions, the readers are directed to [43].

Given a query point $\mathbf{z}_\star \in \mathbb{R}^p$ for the auxiliary measurement, by applying Bayes's rule, the posterior distribution for j-th observed measurement $y_j = f_j(\mathbf{z})$ is given by

$$p(y_j|\mathbf{z},\mathscr{D}) = \mathscr{N}(\mu_j(\mathbf{z}),\Sigma_j(\mathbf{z})), \tag{12}$$

where,

$$\begin{aligned}
&\mu_j(\mathbf{z}) = \mathbf{k}(\mathbf{z})^\top (K+\sigma_j^2 I)^{-1} Y_j^\top \\
&\Sigma_j(\mathbf{z}) = k(\mathbf{z}_\star,\mathbf{z}_\star) - \mathbf{k}(\mathbf{z})^\top (K+\sigma_j^2 I)^{-1}\mathbf{k}(\mathbf{z}), \quad j = 1,2,\cdots,m,
\end{aligned} \tag{13}$$

and

$$K = \begin{bmatrix} k(\mathbf{z}_1,\mathbf{z}_1) & \cdots & k(\mathbf{z}_1,\mathbf{z}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{z}_N,\mathbf{z}_1) & \cdots & k(\mathbf{z}_N,\mathbf{z}_N) \end{bmatrix} \in \mathbb{R}^{N\times N}, \quad \mathbf{k}(\mathbf{z}) = \begin{bmatrix} k(\mathbf{z}_1,\mathbf{z}_\star) \\ \vdots \\ k(\mathbf{z}_N,\mathbf{z}_\star) \end{bmatrix} \in \mathbb{R}^N$$

are covariance matrix on training dataset, and covariance vector between training auxiliary states $\mathbf{z}_i, i = 1,2,\cdots,N$ and the query point $\mathbf{z}_\star$ respectively.

The overall observed measurements' posterior distribution is then given by

$$p(\mathbf{y} \mid \mathbf{z},\mathscr{D}) = \prod_{j=1}^{m} \mathscr{N}\left(\mu_j(\mathbf{z}),\Sigma_j(\mathbf{z})\right) = \mathscr{N}(\mu(\mathbf{z}),\Sigma(\mathbf{z})), \tag{14}$$

where,

$$\mu(\mathbf{z}) = \begin{bmatrix} \mu_1(\mathbf{z}) \\ \vdots \\ \mu_m(\mathbf{z}) \end{bmatrix}, \quad \Sigma(\mathbf{z}) = \begin{bmatrix} \Sigma_1(\mathbf{z}) & & \\ & \ddots & \\ & & \Sigma_m(\mathbf{z}) \end{bmatrix}.$$

Next, the localization algorithm based on the trained GPRs in (13), (14) is given in Algorithm 1. Based on the localization algorithm, if a measurement cannot be explained by the trained prior model, it will be recognized as being attacked. In other words, the prior model provides an additional layer of security by: 1) requiring the attacker to have knowledge of the auxiliary model and the parameters, and 2) limiting the magnitude of possible state corruption.

Algorithm 1: Localization Algorithm with Measurement Prior

I. **Inputs**: $\mathbf{y} \in \mathbb{R}^m$ (real measurement), $\mathbf{z} \in \mathbb{R}^p$ (auxiliary variables)

II. **Parameters**: m trained GPR models **GP**

III. **Posterior distribution**:

$$\mathbf{GP}_j(\mathbf{z}) \to \{\mu_j, \Sigma_j\} \quad \forall j = 1, 2, \cdots, m$$

IV. **Calculate Z-score**:

$$\mathbf{z}_j = \frac{\mathbf{y}_j - \mu_j}{\Sigma_j}$$

V. **Calculate probability**:

$$\mathbf{p}_j = 1 - P_X(|x| \leq |\mathbf{z}_j|) = 1 - \int_{|\mathbf{z}_j|} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \quad \forall j = 1, 2, \cdots, m$$

VI. **Attack support prior**:

$$\mathscr{T} = \mathbf{0}_m; \quad \mathscr{T}_j = 1 \quad \text{if} \quad \mathbf{p}_j \leq 0.5 \quad \forall j = 1, 2, \cdots, m$$

VII. **Outputs**: $\mathscr{T} \in \mathbb{R}^m$, (support prior), $\mathbf{p} \in \mathbb{R}^m$ (confidence)

3.4 Prior Pruning

As shown in previous subsection, an estimated support prior $\hat{\mathscr{T}}$ can be generated by some machine learning localization algorithms. However, there are major limitations preventing their direct usage as the prior information in resilient observer design. One is the huge amount of training often needed for high enough precision will prevent such prior from being deployed for a dynamic observer, where real-time update is paramount. Another limitation is that the precision of data-driven results cannot be guaranteed due to their inherent uncertainties. Consequently, several fundamental questions emerge, that require significant research efforts to address. For example, what is the quantitative relation between the resulting resilient estimation error bound and the auxiliary model uncertainty? In this subsection, a relationship is derived or such connection and a prior pruning method is considered to mend some deficiencies in order to improve the degradation due to the uncertainty of prior model in the final estimation error bound.

Let $\mathscr{T} = \mathsf{supp}(\mathbf{e})$ be the unknown actual support of attacked channels. Let the vector $\mathbf{q} \in \{0\ \ 1\}^{Tm}$ be an indicator of $\mathscr{T}$ defined element-wise as:

$$\mathbf{q}_i = \begin{cases} 0 & \text{if } \ i \in \mathscr{T} \\ 1 & \text{otherwise.} \end{cases} \tag{15}$$

Thus, the output of the localization algorithm $\hat{\mathscr{T}} \subseteq \{1,2,\cdots,Tm\}$ is actually an estimate of $\mathscr{T}$, and its corresponding indicator $\hat{\mathbf{q}} \in \{0\ \ 1\}^{Tm}$ is defined similarly to (15). Consequently, the precision of the support prior is evaluated using positive prediction value [44] instead of true positive rate, F1 score or other evaluation metrics. This is because the only factor affecting the resilient estimation performance is the error in the estimated prior support of safe nodes $\hat{\mathscr{T}}^c$, which is directly used in observer.

Definition 4 (Positive Prediction Value, Precision, PPV [44]). Given an estimate $\hat{\mathbf{q}} \in \{0,1\}^{Tm}$ of an unknown attack support indicator $\mathbf{q} \in \{0,1\}^{Tm}$, PPV is the proportion of $\mathbf{q}$ that is correctly identified in $\hat{\mathbf{q}}$. It is given by

$$\mathsf{PPV} = \frac{\|\mathbf{q} \circ \hat{\mathbf{q}}\|_{\ell_0}}{\|\hat{\mathbf{q}}\|_{\ell_0}}. \tag{16}$$

As will be shown in subsequent sections, the precision PPV is positively correlated to the performance of resilient estimation.

The agreement between $\hat{\mathscr{T}}$ and $\mathscr{T}$ can be described using a Bernoulli uncertainty model since $\hat{\mathscr{T}}$ can be seen as an output of binary classifier. Thus, the following uncertainty model is considered:

$$\mathbf{q}_i = \varepsilon_i \hat{\mathbf{q}}_i + (1-\varepsilon_i)(1-\hat{\mathbf{q}}_i), \tag{17}$$

where $\varepsilon_i \sim \mathscr{B}(1,\mathbf{p}_i)$, with known $\mathbf{p}_i \in \mathbb{R}_+$ generated by Algorithm 1. Here $\mathbf{p}_i = E[\varepsilon_i] = \mathsf{Pr}\{\varepsilon_i = 1\}$. Next, some initial results are given to aid in the subsequent observer development.

Lemma 1. *With respect to the uncertainty model in* (17)*, the PPV defined in* (16) *can be expressed as:*

$$PPV = \frac{1}{|\hat{\mathscr{T}}^c|} \sum_{i \in \hat{\mathscr{T}}^c} \varepsilon_i. \tag{18}$$

Proof. From (17), it follows that

$$\mathbf{q}_i \hat{\mathbf{q}}_i = \varepsilon_i \hat{\mathbf{q}}_i.$$

This implies that

$$\mathsf{PPV} = \frac{\|\mathbf{q} \circ \hat{\mathbf{q}}\|_{\ell_0}}{\|\hat{\mathbf{q}}\|_{\ell_0}} = \frac{\sum_{i=1}^{Tm} \mathbf{q}_i \hat{\mathbf{q}}_i}{\sum_{i=1}^{Tm} \hat{\mathbf{q}}_i} = \frac{1}{|\hat{\mathscr{T}}^c|} \sum_{i=1}^{Tm} \varepsilon_i \hat{\mathbf{q}}_i = \frac{1}{|\hat{\mathscr{T}}^c|} \sum_{i \in \hat{\mathscr{T}}^c} \varepsilon_i.$$

Theorem 2. *The support estimate is better than random flip of a fair coin if and only if*

$$\sum_{i=1}^{Tm} \mathbf{p}_i > Tmp_A \tag{19}$$

where, $p_A \in (0,1)$ is the expected fraction of attacked nodes. Moreover, if p_A is the maximum fraction of attacked nodes, then the conclusion is sufficient, but not necessary.

Proof. Expending (17) yields

$$\begin{aligned}\mathbf{q}_i &= 2\varepsilon_i\hat{\mathbf{q}}_i + 1 - \hat{\mathbf{q}}_i - \varepsilon_i \\ &= 1 - \hat{\mathbf{q}}_i - \varepsilon_i + 2\mathbf{q}_i\hat{\mathbf{q}}_i.\end{aligned}$$

This implies that

$$\varepsilon_i - 1 + \mathbf{q}_i = 2(\mathbf{q}_i\hat{\mathbf{q}}_i - \frac{1}{2}\hat{\mathbf{q}}_i).$$

Summing over $i = 1, \cdots, Tm$ and taking expected value of both sides yield

$$\sum_{i=1}^{Tm} \mathbf{p}_i - Tm + E[\|\mathbf{q}\|_{\ell_0}] = 2E\left[\|\mathbf{q}\circ\hat{\mathbf{q}}\|_{\ell_0} - \frac{1}{2}\|\hat{\mathbf{q}}\|_{\ell_0}\right].$$

thus,

$$E\left[\|\mathbf{q}\circ\hat{\mathbf{q}}\|_{\ell_0} - \frac{1}{2}\|\hat{\mathbf{q}}\|_{\ell_0}\right] > 0 \text{ iff } \sum_{i=1}^{Tm} \mathbf{p}_i > Tm - E[\|\mathbf{q}\|_{\ell_0}] \triangleq Tmp_A.$$

Moreover, if $Tm - E[\|\mathbf{q}\|_{\ell_0}] \leq Tmp_A$, it is straightforward to see that $\sum_{i=1}^{Tm} \mathbf{p}_i > Tmp_A$ is only sufficient for $E\left[\|\mathbf{q}\circ\hat{\mathbf{q}}\|_{\ell_0} + \frac{1}{2}\|\hat{\mathbf{q}}\|_{\ell_0}\right] > 0$.

Lemma 2. *Given mutually independent Bernoulli random variables $\varepsilon_i \sim \mathscr{B}(1,\mathbf{p}_i)$, $i = 1, \cdots, N$, the following holds:*

$$Pr\left\{\sum_{i=1}^{N} \varepsilon_i = k-1\right\} = \mathbf{r}(k), \quad k = 1, \cdots, N+1, \tag{20}$$

where,

$$\mathbf{r} = \beta \cdot \begin{bmatrix}-\mathbf{s}_1\\ 1\end{bmatrix} * \begin{bmatrix}-\mathbf{s}_2\\ 1\end{bmatrix} * \cdots * \begin{bmatrix}-\mathbf{s}_m\\ 1\end{bmatrix}, \quad \textit{with} \quad \beta = \prod_{i=1}^{N} \mathbf{p}_i \quad \textit{and} \quad \mathbf{s}_i = -\frac{1-\mathbf{p}_i}{\mathbf{p}_i}.$$

Proof.

$$\begin{bmatrix}\mathsf{Pr}(\sum_{i=1}^{N} \varepsilon_i = 0)\\ \mathsf{Pr}(\sum_{i=1}^{N} \varepsilon_i = 1)\\ \vdots \\ \mathsf{Pr}(\sum_{i=1}^{N} \varepsilon_i = N)\end{bmatrix} = \begin{bmatrix}1-\mathbf{p}_1\\ \mathbf{p}_1\end{bmatrix} * \begin{bmatrix}1-\mathbf{p}_2\\ \mathbf{p}_2\end{bmatrix} * \cdots * \begin{bmatrix}1-\mathbf{p}_N\\ \mathbf{p}_N\end{bmatrix}$$

Let $\mathsf{Pr}_k = \mathsf{Pr}(\sum_{i=1}^{N} \varepsilon_i = k)$, taking z-transform yields

$$\begin{aligned}\mathsf{Pr}_0 + \mathsf{Pr}_1 z + \mathsf{Pr}_2 z^2 + \cdots + \mathsf{Pr}_N z^N &= (1-\mathbf{p}_1+\mathbf{p}_1 z)(1-\mathbf{p}_2+\mathbf{p}_2 z)\cdots(1-\mathbf{p}_N+\mathbf{p}_N z) \\ &= \mathbf{p}_1\mathbf{p}_2\cdots\mathbf{p}_N \cdot \left(z+\frac{1-\mathbf{p}_1}{\mathbf{p}_1}\right)\cdots\left(z+\frac{1-\mathbf{p}_N}{\mathbf{p}_N}\right)\end{aligned}$$

By doing reverse z-transform, it follows that

$$\begin{bmatrix} \mathsf{Pr}(\sum_{i=1}^{N} \varepsilon_i = 0) \\ \mathsf{Pr}(\sum_{i=1}^{N} \varepsilon_i = 1) \\ \vdots \\ \mathsf{Pr}(\sum_{i=1}^{N} \varepsilon_i = N) \end{bmatrix} = \prod_{i=1}^{N} \mathbf{p}_i \cdot \begin{bmatrix} \frac{1-\mathbf{p}_1}{\mathbf{p}_1} \\ 1 \end{bmatrix} * \cdots * \begin{bmatrix} \frac{1-\mathbf{p}_N}{\mathbf{p}_N} \\ 1 \end{bmatrix}$$

Now, we are ready to introduce the pruning method. The central idea is: if we could identify the errors in the prior information, then the precision of prior can be improved. In fact, the precision of prior will be improved by choosing an appropriate subset. However, how to achieve the best pruning performance, quantify the precision improvement, and improve resukting estimation resiliency are all essential but open questions. Here, we will give a formal definition of pruning operation, then provide some answers and give a simple algorithm to achieve sub-optimal pruning goal.

Definition 5 (Pruning, Pruning operation, PPV_η). Given a prior support estimate $\hat{\mathscr{T}}$, pruning operation, with parameter η, is any operation, or sequence of operations, which returns an updated estimated support prior $\hat{\mathscr{T}}_\eta \subset \{1,\cdots,Tm\}$ such that

$$\hat{\mathscr{T}}_\eta^c \subseteq \hat{\mathscr{T}}^c.$$

Also the precision of pruned support prior $\hat{\mathscr{T}}_\eta$ is given by

$$\mathsf{PPV}_\eta = \frac{1}{|\hat{\mathscr{T}}_\eta^c|} \sum_{i \in \hat{\mathscr{T}}_\eta^c} \varepsilon_i. \tag{21}$$

The following theorem quantifies the resulting precision improvement through the defined pruning operation.

Theorem 3. *Given an estimated attack support $\hat{\mathscr{T}} \subseteq \{1,2,\cdots,Tm\}$ with the uncertainty characteristic described in* (17)*. Let $\hat{\mathscr{T}}_\eta$ be a pruned support estimate satisfying $\hat{\mathscr{T}}_\eta^c \subseteq \hat{\mathscr{T}}^c$, then, for any $\gamma \in (01,1)$,*

$$Pr\big\{PPV_\eta - \gamma PPV \geq 0\big\} \geq \sum_{j=1}^{|\hat{\mathscr{T}}_\eta^c|+1} \left(\mathbf{r}_\eta(j) \sum_{i=1}^{\Phi_{j-1}+1} \tilde{\mathbf{r}}(i)\right), \tag{22}$$

where,

$$\mathbf{r}_\eta = \left(\prod_{i\in\hat{\mathscr{T}}_\eta^c} \mathbf{p}_i\right)\begin{bmatrix}-\mathbf{s}_{\eta,1}\\1\end{bmatrix} * \begin{bmatrix}-\mathbf{s}_{\eta,2}\\1\end{bmatrix} * \cdots * \begin{bmatrix}-\mathbf{s}_{\eta,|\hat{\mathscr{T}}_\eta^c|}\\1\end{bmatrix},$$

$$\tilde{\mathbf{r}} = \left(\prod_{i\in\hat{\mathscr{T}}^c\setminus\hat{\mathscr{T}}_\eta^c} \mathbf{p}_i\right)\begin{bmatrix}-\tilde{\mathbf{s}}_1\\1\end{bmatrix} * \begin{bmatrix}-\tilde{\mathbf{s}}_2\\1\end{bmatrix} * \cdots * \begin{bmatrix}-\tilde{\mathbf{s}}_{|\hat{\mathscr{T}}^c\setminus\hat{\mathscr{T}}_\eta^c|}\\1\end{bmatrix},$$

and $\Phi_k = \min\left\{\left\lceil\frac{|\hat{\mathscr{T}}^c|}{\gamma|\hat{\mathscr{T}}_\eta^c|}-1\right\rceil k, |\hat{\mathscr{T}}^c|-|\hat{\mathscr{T}}_\eta^c|\right\}$, $\mathbf{s}_{\eta,i} = -\dfrac{1-\mathbf{p}_{\hat{\mathscr{T}}_\eta^c,i}}{\mathbf{p}_{\hat{\mathscr{T}}_\eta^c,i}}, \tilde{\mathbf{s}}_i = -\dfrac{1-\mathbf{p}_{\hat{\mathscr{T}}^c\setminus\hat{\mathscr{T}}_\eta^c,i}}{\mathbf{p}_{\hat{\mathscr{T}}^c\setminus\hat{\mathscr{T}}_\eta^c,i}}$.

Proof.

$$\begin{aligned}
&\mathsf{Pr}\left\{\mathsf{PPV}_\eta - \gamma\mathsf{PPV} \geq 0\right\}\\
=&\mathsf{Pr}\left\{\left(\frac{1}{|\hat{\mathscr{T}}_\eta^c|}-\frac{\gamma}{|\hat{\mathscr{T}}^c|}\right)\sum_{i\in\hat{\mathscr{T}}_\eta^c}\varepsilon_i - \frac{\gamma}{|\hat{\mathscr{T}}^c|}\sum_{j\in\hat{\mathscr{T}}^c\setminus\hat{\mathscr{T}}_\eta^c}\varepsilon_j \geq 0\right\}\\
\geq&\mathsf{Pr}\left\{\left(\sum_{i\in\hat{\mathscr{T}}_\eta^c}\varepsilon_i = k\right)\ \&\ \left(\frac{\gamma}{|\hat{\mathscr{T}}^c|}\sum_{j\in\hat{\mathscr{T}}^c\setminus\hat{\mathscr{T}}_\eta^c}\varepsilon_j \leq \left(\frac{1}{|\hat{\mathscr{T}}_\eta^c|}-\frac{\gamma}{|\hat{\mathscr{T}}^c|}\right)k\right)\right\}\\
\geq&\sum_{k=0}^{|\hat{\mathscr{T}}_\eta^c|}\mathsf{Pr}\left\{\left[\sum_{i\in\hat{\mathscr{T}}_\eta^c}\varepsilon_i = k\right]\ \&\ \left[\sum_{j\in\hat{\mathscr{T}}^c\setminus\hat{\mathscr{T}}_\eta^c}\varepsilon_j \leq \frac{|\hat{\mathscr{T}}^c|k}{\gamma|\hat{\mathscr{T}}_\eta^c|}-k\right]\right\}.
\end{aligned}$$

Since $\sum_{i\in\hat{\mathscr{T}}_\eta^c}\varepsilon_i$ and $\sum_{j\in\hat{\mathscr{T}}^c\setminus\hat{\mathscr{T}}_\eta^c}\varepsilon_j$ are independent, it follows that

$$\begin{aligned}
&\mathsf{Pr}\left\{\mathsf{PPV}_\eta - \gamma\mathsf{PPV} > 0\right\}\\
\geq&\sum_{k=0}^{|\hat{\mathscr{T}}_\eta^c|}\mathsf{Pr}\left\{\sum_{i\in\hat{\mathscr{T}}_\eta^c}\varepsilon_i = k\right\}\mathsf{Pr}\left\{\sum_{j\in\hat{\mathscr{T}}^c\setminus\hat{\mathscr{T}}_\eta^c}\varepsilon_j \leq \frac{|\hat{\mathscr{T}}^c|k}{\gamma|\hat{\mathscr{T}}_\eta^c|}-k\right\}\\
\geq&\sum_{k=0}^{|\hat{\mathscr{T}}_\eta^c|}\mathsf{Pr}\left\{\sum_{i\in\hat{\mathscr{T}}_\eta^c}\varepsilon_i = k\right\}\mathsf{Pr}\left\{\sum_{j\in\hat{\mathscr{T}}^c\setminus\hat{\mathscr{T}}_\eta^c}\varepsilon_j \leq \Phi_k\right\}.
\end{aligned}$$

Next, by using Lemma 2, the lower bound in (22) is obtained.

The lower bound given by Theorem 3 can be expressed as $\mathbf{r}_\eta^\top\mathbf{r}_\Phi$ where, $\mathbf{r}_\Phi \in [0,1]^{|\hat{\mathscr{T}}_\eta^c|}$ is a vector whose entries are functions of $|\hat{\mathscr{T}}^c|$, $|\hat{\mathscr{T}}_\eta^c|$, γ and $\tilde{\mathbf{r}}$. Thus, given $\mathbf{p}_i$, $\hat{\mathscr{T}}$, γ and a fixed integer $l_\eta \leq |\hat{\mathscr{T}}^c|$, the pruned support $\hat{\mathscr{T}}_\eta$ can be chosen to maximize $\mathbf{r}_\eta^\top\mathbf{r}_\Phi$. However, such optimization problem is challenging and potentially NP-hard due to the index searching operation involved. But a simple heuristic of returning the indices of the channels with largest $\mathbf{p}_i$ in $\hat{\mathscr{T}}_\eta^c$ can provide a very good sub-optimal estimation. This idea is central to the pruning algorithm considered in this chapter. Fig. 3 shows the comparison of the *ordered* pruning idea vs. randomly selecting a subset of $\hat{\mathscr{T}}$. This illustrative example clearly demonstrates that ordered

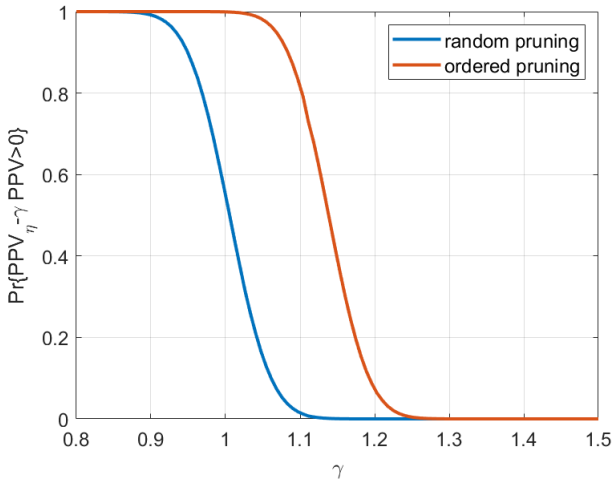

Fig. 3 A comparison between random pruning operation and ordered pruning operation

operation can offer some advantage. Next, one of ordered pruning algorithm is given in Algorithm 2.

Algorithm 2: Support Prior Pruning Algorithm

I. **Obtaining reliable trust parameter**

Given reliability level $\eta \in (0,1)$, return the maximum size l_η such that l_η safe nodes are correctly localized with a probability of at least η:

$$l_\eta = \max\left\{ |\mathscr{I}| \middle| \prod_{i\in\mathscr{I}} \mathbf{p}_i \geq \eta, \quad \mathscr{I} \in \hat{\mathscr{T}}^c \right\}. \tag{23}$$

II. **Pruning**

A pruned support prior is obtained through a robust extraction:

$$\hat{\mathscr{T}}_\eta^c = \left\{ \mathsf{argsort}\downarrow (\mathbf{p} \circ \hat{\mathbf{q}}) \right\}_1^{l_\eta}. \tag{24}$$

where, $\{\cdot\}_1^{l_\eta}$ is an index extraction from the first elements to l_η elements.

For pragmatic reasons, it is important to ensure that $l_\eta > 0$ in (23). This is guaranteed if η is chosen such that at least one node is selected into the pruned set. Formally, this condition is given by:

$$\eta \leq \max_{i\in\hat{\mathscr{T}}^c} (\mathbf{p}_i) \tag{25}$$

Definition 6 (η-successful pruning algorithm). A η-successful pruning algorithm is any pruning operation, as defined in Definition 5, that achieves:

$$\mathsf{Pr}\left\{\mathsf{PPV}_\eta = 1\right\} \geq \eta.$$

Proposition 1. *Given a support prior estimate $\hat{\mathscr{T}}$ generated by an underlying localization algorithm with associated uncertainty model in* (17)*, the pruning algorithm in Algorithm 2 is η-successful.*

Proof. According to (21),

$$\mathsf{Pr}\left\{\mathsf{PPV}_\eta = 1\right\} = \mathsf{Pr}\left\{\sum_{i\in\hat{\mathscr{T}}_\eta^c} \varepsilon_i = |\hat{\mathscr{T}}_\eta^c|\right\} = \prod_{i\in\hat{\mathscr{T}}_\eta^c} \mathbf{p}_i.$$

Based on (23), and (24), we have $|\hat{\mathscr{T}}_\eta^c| = |\mathscr{I}| = l_\eta$, and $\mathbf{p}_i$'s in $\hat{\mathscr{T}}_\eta^c$ are the l_η largest values. Thus,

$$\prod_{i\in\hat{\mathscr{T}}_\eta^c} \mathbf{p}_i \geq \prod_{i\in\mathscr{I}} \mathbf{p}_i \geq \eta.$$

4 Pruning-based Resilient Estimation

In this section, we will go through resilient observer designs using $\ell_0\backslash\ell_1$ minimization schemes. Firstly, the unconstrained ℓ_1 observer will be stated. Then, we will give a weighted ℓ_1 observer design and state the condition for resilient estimation with the pruned prior support. Furthermore, the quantified relationship between the precision of prior support and the resilient estimation performance will be clarified.

Researchers in compressed sensing have paid much attentions to the recoverability of $\ell_0\backslash\ell_1$ minimization program in last decade. Most of the efforts focused on finding well-defined compressed matrix satisfying null space property (NSP) or restricted isometry property (RIP). Then, the complete information can be reconstructed by $\ell_0\backslash\ell_1$ minimization program from the compressed measurements. From mathematical aspect, the decoding process is to solve an under-determined set of equations which does generally not have unique solutions. However, if the required solution is sparse, it can be recovered completely via ℓ_0 minimization. The condition on sparsity for exact unique recovery is also well known. However, the ℓ_0 minimization program is a NP-hard optimization problem. However, NSP or RIP pave way for a convex relaxation via ℓ_1 minimization program. Interested readers are directed to [45, 46, 47, 49] for more comprehensive treatment of compressed sensing, and [48, 50] for several extension cases.

The basic motivation for using ℓ_1 minimization for attack-resilient estimation is because the attack is possibly unbounded but is necessarily sparse. Consider the measurement model in (2), if a coding matrix F can be found that satisfies $FH = 0$, then a new under-determined equation $F\mathbf{y} = F\mathbf{e}$ is obtained. If the sparse attack vector is recovered, the resilient estimation goal is easily achieved. In this section, instead of finding a coding matrix F directly, we would formulate the problem within

the familiar framework of linear systems theory and prove results similarly to compressed sensing literature.

4.1 Unconstrained ℓ_1 Observer

In this subsection, we discuss the uniqueness of resilient estimation solution in the presence of measurement attacks, and introduce the concept of column space property (CSP). Furthermore, the estimation error bound is given using CSP.

Consider the system model in (1) and the unconstrained ℓ_1 decoder in (3), a formal notion of attack recovery is given as following:

Definition 7 (Resilient Recovery). k sensor attacks are correctable after T steps by $\mathscr{D}:(\mathbb{R}^m)^T \to \mathbb{R}^n$ if for any $\mathbf{x}_0 \in \mathbb{R}^n$ and any sequence of attack vectors $\mathbf{e}_0, \mathbf{e}_1, \cdots, \mathbf{e}_{T-1} \in \mathbb{R}^m$ with $\mathsf{supp}(e_t) \leq k$, we have $\mathscr{D}(\mathbf{y}_0, \cdots, \mathbf{y}_{T-1}) = \mathbf{x}_0$.

The following theorem states the uniqueness of resilient estimation solution:

Theorem 4. *Given attack support $\mathscr{T} = \{\mathscr{T}_i, \mathscr{T}_{i-1}, \cdots, \mathscr{T}_{i-T+1}\}$ with $|\mathscr{T}_i| \leq k$. Consider the noise-free version of the measurement model in* (2)*. If, for any $\mathbf{h} \in \mathsf{range}(H)$, it is true that*

$$\|\mathbf{h}_{\mathscr{S}}\|_1 \leq \|\mathbf{h}_{\mathscr{S}^c}\|_1, \quad \forall \mathscr{S} \subset \{1,2,\cdots,Tm\}, |\mathscr{S}| \leq Tk \tag{26}$$

then, for each attacked measurement $\mathbf{y}_T \in \mathbb{R}^{Tm}$, there exists an unique state vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ and Tk-sparse attack vector $\hat{\mathbf{e}}$ which satisfy (2)*.*

Proof. Let $(\mathbf{z}_1, \mathbf{e}_1), (\mathbf{z}_2, \mathbf{e}_2) \in \mathbb{R}^n \times \Sigma_{Tk}$ such that

$$\mathbf{y}_T = H\mathbf{z}_1 + \mathbf{e}_1 = H\mathbf{z}_2 + \mathbf{e}_2,$$

then

$$H(\mathbf{z}_1 - \mathbf{z}_2) = \mathbf{e}_2 - \mathbf{e}_1.$$

Thus, the uniqueness condition holds iff

$$\mathsf{range}(H) \cap \Sigma_{2Tk} = \{\mathbf{0}\}.$$

Now, given $\mathbf{h} \in \mathsf{range}(H)$ which satisfies (26), it suffices to show that $\|\mathbf{h}\|_0 > 2Tk$.

Suppose, for the sake of contradiction, that $\|\mathbf{h}\|_0 \leq 2Tk$. Choose $\overline{\mathscr{S}} \in \{1,2,\cdots,Tm\}$, $|\overline{\mathscr{S}}| = Tk$ to be the indices of the largest components of $\mathbf{h}$ in absolute value.
Then, it must be that

$$\|\mathbf{h}_{\overline{\mathscr{S}}}\|_0 > \|\mathbf{h}_{\overline{\mathscr{S}}^c}\|_0 \Rightarrow \|\mathbf{h}_{\overline{\mathscr{S}}}\|_1 > \|\mathbf{h}_{\overline{\mathscr{S}}^c}\|_1,$$

which is a contradiction. Thus, (26) implies that $\|\mathbf{h}\|_0 > 2Tk$.

Consequently, a formal definition of column space property is given as follows.

Definition 8 (Column space property (CSP)). A matrix $H \in \mathbb{R}^{m\times n}$ has a column space property of order $s < m$ (denoted as $H \triangleright \mathsf{CSP}(s)$) if there exists $\beta \in (0,1)$ such that, for every $\mathbf{h} \in \mathsf{range}(H)$,

$$\|\mathbf{h}_{\mathscr{S}}\|_1 \leq \beta \|\mathbf{h}_{\mathscr{S}^c}\|_1, \quad \forall \mathscr{S} \subset \{1,2,\cdots,m\}, |\mathscr{S}| \leq s. \tag{27}$$

The above definition is similar to the well-known *Null Space Property* but defined on the range space instead. For dynamic system (1), the unconstrained ℓ_1 observer is defined as a moving-horizon unconstrained ℓ_1 minimization program:

$$\begin{aligned} \mathsf{Minimize} \quad & \sum_{j=i-T+1}^{i} \|\mathbf{y}_j - C\mathbf{x}_j\|_1 \\ \mathsf{Subject\ to} \quad & \mathbf{x}_{j+1} - A\mathbf{x}_j = 0, \quad j = i-T+1,\cdots,i-1 \end{aligned} \tag{28}$$

An equivalent optimization program of (28) is given by

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\mathsf{Minimize}} \ \|\mathbf{y}_T - H\mathbf{x}\|_1 \tag{29}$$

The following theorem gives the conditions for resilient recovery of the state vector obtained by the above observer.

Theorem 5 (Resilient Recovery with CSP). *Consider the measurement model in* (2)*, let* $\mathscr{T} = \{\mathscr{T}_i, \mathscr{T}_{i-1}, \cdots, \mathscr{T}_{i-T+1}\}$*, with* $|\mathscr{T}_i| \leq k$*, be the unknown sequence of the attack support. If* $H \triangleright \mathsf{CSP}(Tk)$*, the estimation error due to the decoder in* (29) *can be upper bounded as:*

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq \frac{2(1+\beta)}{\underline{\sigma}(1-\beta)}\varepsilon, \tag{30}$$

for some $\beta \in (0,1)$*, and* $\underline{\sigma}$ *is the smallest singular value of* H*.*

Proof. Let $\hat{\mathbf{x}}$ be the optimal solution of (29), then its optimality yields

$$\begin{aligned} \|\mathbf{y} - H\hat{\mathbf{x}}\|_1 \leq \|\mathbf{y} - H\mathbf{x}\|_1 = \|\mathbf{e}\|_1 \\ \|\mathbf{y} - H\mathbf{x} + H(\mathbf{x} - \hat{\mathbf{x}})\|_1 \leq \|\mathbf{e}\|_1 \end{aligned}$$

Let $\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}$, and since 1-norm is decomposable for disjoint sets, then

$$\begin{aligned} \|\mathbf{e} + H\tilde{\mathbf{x}}\|_1 \leq \|\mathbf{e}\|_1 \\ \|\mathbf{e}_{\mathscr{T}} + H_{\mathscr{T}}\tilde{\mathbf{x}}\|_1 + \|\mathbf{e}_{\mathscr{T}^c} + H_{\mathscr{T}^c}\tilde{\mathbf{x}}\|_1 \leq \|\mathbf{e}_{\mathscr{T}}\|_1 + \|\mathbf{e}_{\mathscr{T}^c}\|_1 \\ \|\mathbf{e}_{\mathscr{T}}\|_1 - \|H_{\mathscr{T}}\tilde{\mathbf{x}}\|_1 - \|\mathbf{e}_{\mathscr{T}^c}\|_1 + \|H_{\mathscr{T}^c}\tilde{\mathbf{x}}\|_1 \leq \|\mathbf{e}_{\mathscr{T}}\|_1 + \|\mathbf{e}_{\mathscr{T}^c}\|_1 \end{aligned}$$

And let $\mathbf{h} = H\tilde{\mathbf{x}}$, it follows

$$\|\mathbf{h}_{\mathscr{T}^c}\|_1 \leq \|\mathbf{h}_{\mathscr{T}}\|_1 + 2\varepsilon. \tag{31}$$

Since $H \triangleright \mathsf{CSP}(Tk)$, there exist $\beta \in (0,1)$ such that $\|\mathbf{h}_{\mathscr{T}}\|_1 \leq \beta\|\mathbf{h}_{\mathscr{T}^c}\|_1$. Thus

$$\|\mathbf{h}_{\mathscr{T}}\|_1 \leq \frac{2\beta}{1-\beta}\varepsilon$$

Then,

$$\|\mathbf{h}\|_2 \leq \|\mathbf{h}_{\mathscr{T}}\|_1 + \|\mathbf{h}_{\mathscr{T}^c}\|_1 \leq 2\|\mathbf{h}_{\mathscr{T}}\|_1 + 2\varepsilon \leq \frac{2(1+\beta)}{1-\beta}\varepsilon$$

Finally, combining with $\underline{\sigma}\|\tilde{\mathbf{x}}\|_2 \leq \|\mathbf{h}\|_2$ yields the error bound in (30).

Notice that the CSP condition with $\beta \in (0,1)$ is a violation of the condition stated in (7), which is a guarantee of successful FDIA. And the CSP condition is relevant to the sparsity of attack vector. As shown in literature [17], the number of attacks is one of the most important factor deciding if successful resilient estimation would be achieved. With increasing power of FDIA, it is more likely that the CSP condition would be violated. This is one of the motivations for finding an improved resilient estimation method in worst environment.

4.2 Resilient Pruning Observer

In this subsection, we incorporate prior information into the resilient observer design. First, a support prior $\hat{\mathscr{T}}$ is generated by the localization algorithm in Algorithm 1. Then the pruning algorithm in Algorithm 2 is used to improve the precision of the support prior. Finally, a weighted ℓ_1 observer scheme is proposed to utilize the pruned support prior $\hat{\mathscr{T}}_\eta$. This process is summarized in Fig. 4.

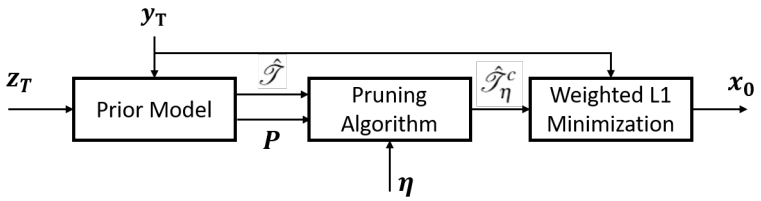

Fig. 4 Schematic depiction of resilient observer design with prior pruning

Consider a time horizon T and a set of attack support prior obtained by Algorithm 1: $\hat{\mathscr{T}} = \{\hat{\mathscr{T}}_i, \hat{\mathscr{T}}_{i-1}, \cdots, \hat{\mathscr{T}}_{i-T+1}\}$. The following weighted ℓ_1 observer is considered:

$$\begin{aligned} \text{Minimize} \quad & \sum_{j=i-T+1}^{i} \|\mathbf{y}_j - C\mathbf{x}_j\|_{1,\mathbf{w}(\hat{\mathscr{T}}_j,\omega)} \\ \text{Subject to} \quad & \mathbf{x}_{j+1} - A\mathbf{x}_j = 0, \quad j = i-T+1, \cdots, i-1, \end{aligned} \tag{32}$$

where, for $\omega \in (0,\ \ 1)$, the weight vector $\mathbf{w}(\hat{\mathscr{T}}_j, \omega) \in \mathbb{R}^m$ is defined element-wise as

$$\mathbf{w}(\hat{\mathscr{T}}_j,\omega)_l = \begin{cases} \omega & \text{if } l \in \hat{\mathscr{T}}_j \\ 1 & \text{otherwise} \end{cases} \tag{33}$$

The optimization problem in (32) is equivalent to

$$\underset{\mathbf{z}\in\mathbb{R}^n}{\text{Minimize}} \ \|\mathbf{y}_T - H\mathbf{z}\|_{1,\mathbf{w}(\hat{\mathscr{T}},\omega)}, \tag{34}$$

where, $\mathbf{w}(\hat{\mathscr{T}},\omega) = \begin{bmatrix} \mathbf{w}(\hat{\mathscr{T}}_i,\omega) \\ \vdots \\ \mathbf{w}(\hat{\mathscr{T}}_{i-T+1},\omega) \end{bmatrix} \in \mathbb{R}^{Tm}$.

Theorem 6 (Resilient Recovery with support prior $\hat{\mathscr{T}}$). *Consider the measurement model in* (2)*, let $\mathscr{T} = \{\mathscr{T}_i, \mathscr{T}_{i-1}, \cdots, \mathscr{T}_{i-T+1}\}$, with $|\mathscr{T}_i| \leq k$, be the unknown support sequence of the attack vector such that $\sum_{i\in\mathscr{T}^c} |\mathbf{e}_i| < \varepsilon$. Let $\hat{\mathscr{T}} = \{\hat{\mathscr{T}}_i, \hat{\mathscr{T}}_{i-1}, \cdots, \hat{\mathscr{T}}_{i-T+1}\}$ be a support prior estimate satisfying*

$$|\hat{\mathscr{T}}| = \rho|\mathscr{T}| \text{ and } |\mathscr{T}\cap\hat{\mathscr{T}}| = \alpha|\hat{\mathscr{T}}|. \tag{35}$$

If $H \triangleright \mathit{CSP}(\kappa Tk)$, where $\kappa = \rho + 1 - 2\alpha\rho$ with $\rho > 0, \alpha \in (0,\ 1)$, then the estimation error due to the decoder in (32) *can be upper bounded as:*

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq \frac{2(1+\beta)}{\underline{\sigma}(1-\beta)}\varepsilon, \tag{36}$$

for some $\beta \in (0,1)$, where $\underline{\sigma}$ is the smallest singular value of H.

Proof. Let $\hat{\mathbf{x}}$ be the optimal solution of (34), and define $\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}$, $\mathbf{h} = H\tilde{\mathbf{x}}$. Similar to the proof of Theorem 5, the optimality of $\hat{\mathbf{x}}$ yields

$$\|\mathbf{e} + \mathbf{h}\|_{1,\mathbf{w}(\hat{\mathscr{T}},\omega)} \leq \|\mathbf{e}\|_{1,\mathbf{w}(\hat{\mathscr{T}},\omega)}. \tag{37}$$

By the definition of weighted 1-norm, it follows that

$$\omega\|\mathbf{e}_{\hat{\mathscr{T}}} + \mathbf{h}_{\hat{\mathscr{T}}}\|_1 + \|\mathbf{e}_{\hat{\mathscr{T}}^c} + \mathbf{h}_{\hat{\mathscr{T}}^c}\|_1 \leq \omega\|\mathbf{e}_{\hat{\mathscr{T}}}\|_1 + \|\mathbf{e}_{\hat{\mathscr{T}}^c}\|_1,$$

then

$$\begin{aligned} &\omega\|\mathbf{e}_{\hat{\mathscr{T}}\cap\mathscr{T}} + \mathbf{h}_{\hat{\mathscr{T}}\cap\mathscr{T}}\|_1 + \omega\|\mathbf{e}_{\hat{\mathscr{T}}\cap\mathscr{T}^c} + \mathbf{h}_{\hat{\mathscr{T}}\cap\mathscr{T}^c}\|_1 + \|\mathbf{e}_{\hat{\mathscr{T}}^c\cap\mathscr{T}} + \mathbf{h}_{\hat{\mathscr{T}}^c\cap\mathscr{T}}\|_1 \\ &+ \|\mathbf{e}_{\hat{\mathscr{T}}^c\cap\mathscr{T}^c} + \mathbf{h}^c_{\hat{\mathscr{T}}^c\cap\mathscr{T}}\|_1 \leq \omega\|\mathbf{e}_{\hat{\mathscr{T}}\cap\mathscr{T}}\|_1 + \|\mathbf{e}_{\hat{\mathscr{T}}\cap\mathscr{T}^c}\|_1 + \|\mathbf{e}_{\hat{\mathscr{T}}^c\cap\mathscr{T}}\|_1 + \|\mathbf{e}_{\hat{\mathscr{T}}^c\cap\mathscr{T}^c}\|_1. \end{aligned}$$

Using the reverse triangle inequality yields

$$\omega\|\mathbf{h}_{\hat{\mathscr{T}}\cap\mathscr{T}^c}\|_1 + \|\mathbf{h}_{\hat{\mathscr{T}}^c\cap\mathscr{T}^c}\|_1 \leq \|\mathbf{h}_{\hat{\mathscr{T}}^c\cap\mathscr{T}}\|_1 + \omega\|\mathbf{h}_{\hat{\mathscr{T}}\cap\mathscr{T}}\|_1 + 2(\|\mathbf{e}_{\hat{\mathscr{T}}^c\cap\mathscr{T}^c}\|_1 + \|\mathbf{e}_{\hat{\mathscr{T}}\cap\mathscr{T}^c}\|_1)$$

Adding and subtracting $\omega\|\mathbf{h}_{\hat{\mathscr{T}}^c\cap\mathscr{T}^c}\|_1$ on the left, and $\omega\|\mathbf{h}_{\hat{\mathscr{T}}^c\cap\mathscr{T}}\|_1$, $\omega\|\mathbf{e}_{\hat{\mathscr{T}}^c\cap\mathscr{T}^c}\|_1$ on the right yields:

$$\omega\|\mathbf{h}_{\mathscr{T}^c}\|_1+(1-\omega)\|\mathbf{h}_{\hat{\mathscr{T}}^c\cap\mathscr{T}^c}\|_1\leq(1-\omega)\|\mathbf{h}_{\hat{\mathscr{T}}^c\cap\mathscr{T}}\|_1+\omega\|\mathbf{h}_{\mathscr{T}}\|_1 \\ +2(\omega\|\mathbf{e}_{\mathscr{T}^c}\|_1+(1-\omega)\|\mathbf{e}_{\hat{\mathscr{T}}^c\cap\mathscr{T}^c}\|_1).$$

Again, adding and subtracting $(1-\omega)\|\mathbf{h}_{\hat{\mathscr{T}}\cap\mathscr{T}^c}\|_1$ on the left and substituting $\sum_{i\in\mathscr{T}^c}|\mathbf{e}_i|<\varepsilon$ yields:

$$\|\mathbf{h}_{\mathscr{T}^c}\|_1\leq\omega\|\mathbf{h}_{\mathscr{T}}\|_1+(1-\omega)(\|\mathbf{h}_{\hat{\mathscr{T}}^c\cap\mathscr{T}}\|_1+\|\mathbf{h}_{\hat{\mathscr{T}}\cap\mathscr{T}^c}\|_1)+2\varepsilon.$$

Let $\mathscr{T}_\alpha\triangleq(\hat{\mathscr{T}}^c\cap\mathscr{T})\cup(\hat{\mathscr{T}}\cap\mathscr{T}^c)=\hat{\mathscr{T}}\cup\mathscr{T}\setminus\hat{\mathscr{T}}\cap\mathscr{T}$. It follows that $|\mathscr{T}_\alpha|=\kappa|\mathscr{T}|\leq\kappa Tk$. Also, since $\hat{\mathscr{T}}^c\cap\mathscr{T}$ and $\hat{\mathscr{T}}\cap\mathscr{T}^c$ are disjoint, the inequality above becomes

$$\|\mathbf{h}_{\mathscr{T}^c}\|_1\leq\omega\|\mathbf{h}_{\mathscr{T}}\|_1+(1-\omega)\|\mathbf{h}_{\mathscr{T}_\alpha}\|_1+2\varepsilon. \tag{38}$$

Since $H\triangleright\mathsf{CSP}(\kappa Tk)$, we have

$$\|\mathbf{h}_{\mathscr{T}}\|_1\leq\beta\|\mathbf{h}_{\mathscr{T}^c}\|_1 \tag{39}$$
$$\|\mathbf{h}_{\mathscr{T}_\alpha}\|_1\leq\beta\|\mathbf{h}_{\mathscr{T}_\alpha^c}\|_1 \tag{40}$$

And using (40) and property of 1-norm yields:

$$\begin{aligned}\|\mathbf{h}_{\mathscr{T}_\alpha}\|_1+\|\mathbf{h}_{\mathscr{T}_\alpha^c}\|_1&=\|\mathbf{h}\|_1\\ \|\mathbf{h}_{\mathscr{T}_\alpha}\|_1&\leq\frac{\beta}{1+\beta}\|\mathbf{h}\|_1\end{aligned} \tag{41}$$

Then, substituting (39) and (41) into (38) yields

$$(1-\beta\omega)\|\mathbf{h}_{\mathscr{T}^c}\|_1\leq\frac{\beta(1-\omega)}{1+\beta}\|\mathbf{h}\|_1+2\varepsilon \tag{42}$$

Next,

$$\|\mathbf{h}\|_1=\|\mathbf{h}_{\mathscr{T}}\|_1+\|\mathbf{h}_{\mathscr{T}^c}\|_1\leq(1+\beta)\|\mathbf{h}_{\mathscr{T}^c}\|_1\leq\frac{\beta(1-\omega)}{1-\beta\omega}\|\mathbf{h}\|_1+\frac{2(1+\beta)}{1-\beta\omega}\varepsilon$$

then

$$\|\mathbf{h}\|_2\leq\|\mathbf{h}\|_1\leq\frac{2(1+\beta)}{1-\beta}\varepsilon$$

Finally, combining with $\underline{\sigma}\|\tilde{\mathbf{x}}\|_2\leq\|\mathbf{h}\|_2$ yields the error bound in (36).

The estimation error bound in Theorem 6 is the same as the one in Theorem 5. The only difference is that the upper bound of the number of attacks which can be corrected by the underlying observer is governed by κ. If $\kappa<1$, then the weighted ℓ_1 observer with prior (32) has better attack-resiliency compared to the unconstrained ℓ_1 observer (28). Furthermore, the size of κ is actually the relative size of the disagreement set $\mathscr{T}_\alpha=\hat{\mathscr{T}}\cup\mathscr{T}\setminus\hat{\mathscr{T}}\cap\mathscr{T}$ between $\mathscr{T}$ and $\hat{\mathscr{T}}$. Specifically, the quantified relationship between the precision of support prior PPV and the disagreement size κ is given by:

$$\kappa = \rho - 1 + \frac{2(1-\mathsf{PPV})(Tm - \rho|\mathscr{T}|)}{|\mathscr{T}|},$$

where, ρ is given in (35). It is seen that the precision of support prior has a negative correlation to the disagreement size κ. Thus, it has a positive correlation to the attack-resiliency of the underlying observer. Another way to see this is to observe that the condition in Theorem 6 can be stated as $|\mathscr{T}_i| \leq \frac{Tk}{\kappa}$ and $H \triangleright \mathsf{CSP}(Tk)$, from which it is clear that $\kappa < 1$ implies that more attacks can be accommodated by the observer with prior. This is the main motivation for the pruning algorithm. Next, the following corollary gives a better attack-resiliency of weighted ℓ_1 observer with the pruned support $\hat{\mathscr{T}}_\eta$.

Corollary 1 (Resilient Recovery with Pruned Prior $\hat{\mathscr{T}}_\eta$). *Given a support prior $\hat{\mathscr{T}} = \{\hat{\mathscr{T}}_i, \hat{\mathscr{T}}_{i-1}, \cdots, \hat{\mathscr{T}}_{i-T+1}\}$ generated by the localization algorithm in Algorithm 1. Let $\hat{\mathscr{T}}_\eta$ be the pruned support prior obtained from $\hat{\mathscr{T}}$ according to Algorithm 2 with a parameter $\eta \in (0,1)$. Let the precision of $\hat{\mathscr{T}}_\eta$ be denoted by PPV_η. If $H \triangleright \mathsf{CSP}(\kappa_1 Tk)$, where*

$$\kappa_1 = \frac{|\mathscr{T}^c| + l_\eta(1 - 2\mathsf{PPV}_\eta)}{|\mathscr{T}|},$$

then the estimation error due to (32) *with $\hat{\mathscr{T}}_\eta$ can be upper bounded as*

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq \frac{2(1+\beta)}{\underline{\sigma}(1-\beta)}\varepsilon, \tag{43}$$

for some $\beta \in (0,1)$, and $\underline{\sigma}$ is the smallest singular value of H.

Furthermore, with probability at least η, the smallest disagreement size is obtained as

$$\kappa_1 = \frac{Tm - l_\eta}{|\mathscr{T}|} - 1. \tag{44}$$

Proof. (43) can be obtained by following the proof of Theorem 6 but using PPV_η instead. To obtain (44), observe that with probability at least η, $\mathsf{PPV}_\eta = 1$.

5 Simulation Results

In this section, three application examples are given in power grid, wheeled mobile robot, and water distributed system respectively. These application examples are used to demonstrate how (1) to implement the developed observer in previous sections and (2) the resulting pruning-based observers improves compares, performance-wise, with some well-known results in the literature.

5.1 Resilient Power Grid

Here, we implement the proposed pruning observer on a IEEE 14-bus system. The simulation scenario is shown in Fig. 5. The bus system has $n_b = 14$ buses and $n_g = 5$ generators. It is assumed that each bus in the network is equipped with IIoT sensor devices which provide the corresponding active power injection and flow measurements.

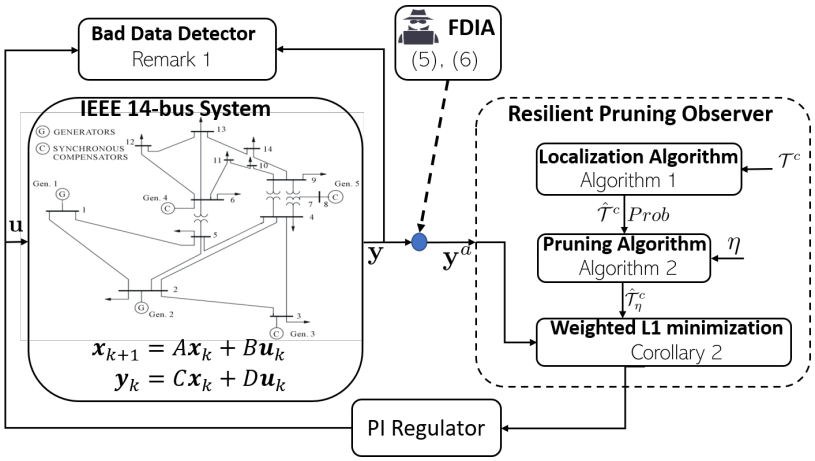

Fig. 5 Block diagram depiction of resilient power grid

A small signal model is constructed by linearizing the generator swing and power flow equations around the operating point. The following linearizing assumptions are made:

1. Voltage is tightly controlled at their nominal value;
2. Angular difference between each bus is small;
3. Conductance is negligible therefore the system is lossless.

By ordering the buses such that the generator nodes appear first, the admittance-weighted Laplacian matrix can be expressed as $L = \begin{bmatrix} L_{gg} & L_{lg} \\ L_{gl} & Lll \end{bmatrix} \in \mathbb{R}^{N\times N}$, where $N = n_g + n_b$. Thus, the dynamical linearized swing equations and algebraic DC power flow equations are given by:

$$\begin{bmatrix} I & 0 & 0 \\ 0 & M & 0 \\ 0 & 0 & 0 \end{bmatrix} \dot{\mathbf{x}} = - \begin{bmatrix} 0 & -I & 0 \\ L_{gg} & D_g & L_{lg} \\ L_{gl} & 0 & L_{ll} \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 & 0 \\ I & 0 \\ 0 & I \end{bmatrix} \mathbf{u}, \tag{45}$$

where, $\mathbf{x} = [\delta^\top \ \ \omega^\top \ \ \theta^\top]^\top \in \mathbb{R}^{2n_g+n_b}$ is the state vector containing generator rotor angle $\delta \in \mathbb{R}^{n_b}$, generator frequency $\omega \in \mathbb{R}^{n_g}$, and voltage bus angles $\theta \in \mathbb{R}^{n_b}$.

$\mathbf{u} = [P_g^\top P_d^\top]^\top \in \mathbb{R}^{n_g+n_b}$ is the input vector consisting of mechanical input power from each generator $P_g \in \mathbb{R}^{n_g}$ and active power demand at each bus $P_d \in \mathbb{R}^{n_b}$, M is a diagonal matrix of inertial constants for each generator, and D_g is a diagonal matrix of damping coefficients. A PI regulator is included to regulate the generator frequency in order to control the P_g. The system in (45) is then simplified as follows:

$$\begin{aligned}
\begin{bmatrix}\dot{\delta}\\ \dot{\omega}\end{bmatrix} &= \begin{bmatrix}0 & I\\ -M^{-1}(L_{gg}-L_{gl}L_{ll}^{-1}L_{lg}) & -M^{-1}D_g\end{bmatrix}\begin{bmatrix}\delta\\ \omega\end{bmatrix} + \begin{bmatrix}0 & 0\\ M^{-1} & -M^{-1}L_{gl}L_{ll}^{-1}\end{bmatrix}\mathbf{u},\\
\begin{bmatrix}\omega\\ P_{net}\end{bmatrix} &= \begin{bmatrix}0 & I\\ -P_{node}L_{ll}^{-1}Llg & 0\end{bmatrix}\begin{bmatrix}\delta\\ \omega\end{bmatrix} + \begin{bmatrix}0 & 0\\ P_{node}L_{ll}^{-1} & 0\end{bmatrix}\mathbf{u},\\
\theta &= -Lll^{-1}(L_{lg}\delta - P_d),
\end{aligned} \tag{46}$$

where P_{node} is a function of the system incidence and susceptance matrices obtained by linearizing the active power injections at the buses [54], and P_{net} is the net power injected at each bus. As shown in Fig. 5, the FDIA designed using (5) and (6) is injected into system through the sensor channels. The bad data detection residual is then calculated after the FDIA is injected, as shown in Fig. 6. The figure indicates the designed FDIA can bypass bad data detector.

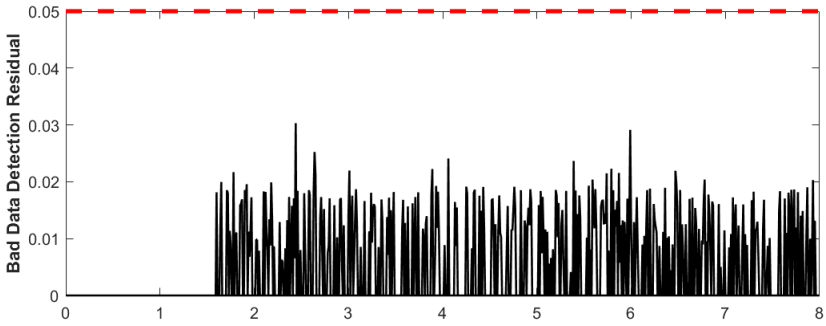

Fig. 6 Bad data detection result (the residual threshold is set as 0.05, 60% of measurement nodes are attacked)

The prior model is a set of trained Gaussian process regression models mapping from the real load data of New York (NY) state provided by the NY Independent System Operator (NYISO) to IEEE 14-bus model measurements. Five-minute load data of NYISO for 3 months (between January and March) in 2017 and 2018 are used. As shown in Fig. 7, the IEEE 14-bus model discussed above is mapped onto the NYISO transmission grid as follows: $A \to 2$, $B \to 3$, $C \to 4$, $D \to 5$, $E \to 6$, $F \to 9$, $G \to 10$, $H \to 11$, $I \to 12$, $J \to 13$, $K \to 14$. Then, the market variables downloaded from the respective nodes of NYISO transmission grid are collected into the auxiliary vector variable $\mathbf{z} = [z_{lbmp} \;\; z_{mcl} \;\; z_{mcc}]$, where z_{lbmp} is the *locational bus marginal prices* (\$/MWh), z_{mcl} is the *marginal cost loses* (\$/MWh), and z_{mcc} is the *marginal cost congestion* (\$/MWh). Using the load data downloaded at NY load

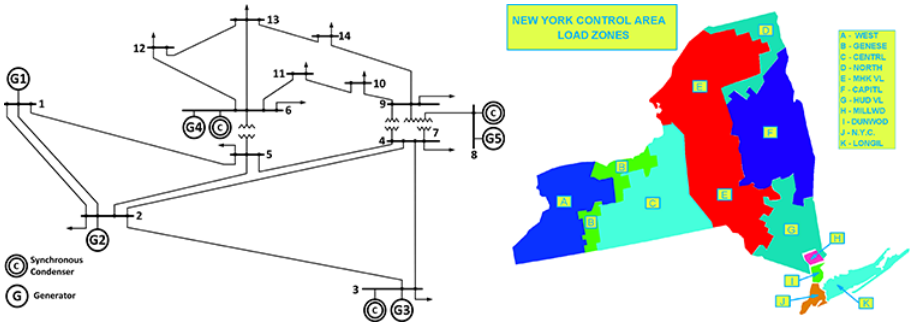

Fig. 7 Mapping from NY load zones to IEEE 14-bus system (Left is IEEE 14-bus system test case [52], right is NYISO control area load zone map [53])

zones for the same time period and interval as output, GPR models were trained to map the auxiliary vector $\mathbf{z}$ to each corresponding bus measurements $\mathbf{y}_j$ containing active power and reactive power of load buses. As shown in (13), the trained GPR models are executed to give the mean $\mu(\mathbf{z})$ and the covariance $\Sigma(\mathbf{z})$ of prior model for each of measurements. The prediction performance of those GPR models, measured by the mean relative absolute errors (MRAE), are shown in Fig. 8. Finally, the

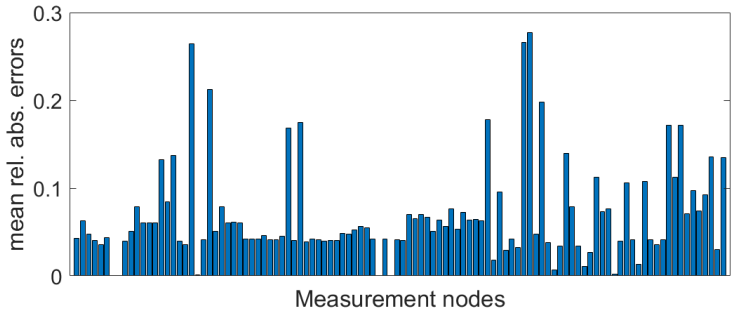

Fig. 8 GPRs' prediction error metrics for all measurement nodes (The mean relative absolute error is used to evaluate the prediction performance)

localization algorithm in Algorithm 1 is implemented on the system model in (46). The precision of the generated support prior calculated for each time step is shown in Fig. 9. The mean of precision is 0.655, which indicates the localization algorithm at least works better than random flip of fair coin.

Furthermore, the developed resilient observer with support prior pruning is compared with some well-known resilient observers in literature. Luenberger observer (LO) is also included to serve as a reference and to show the effectiveness of the designed FDIA. The unconstrained ℓ_1 observer (UL1O) (28), event-triggered Luenberger observer (ETLO) [18], and multi-model observer (MMO) [25] are all resilient observers included in the comparison. MMO is a ℓ_1 observer with multiple

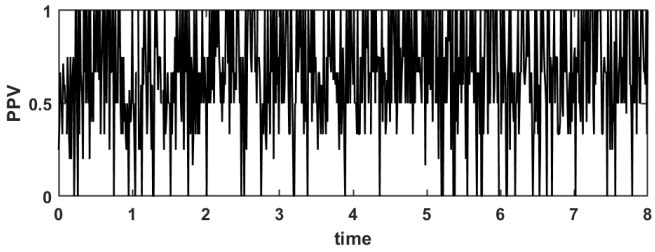

Fig. 9 The precision of support prior generated by the localization algorithm in Algorithm 1 for the power grid (The mean of precision is 0.655)

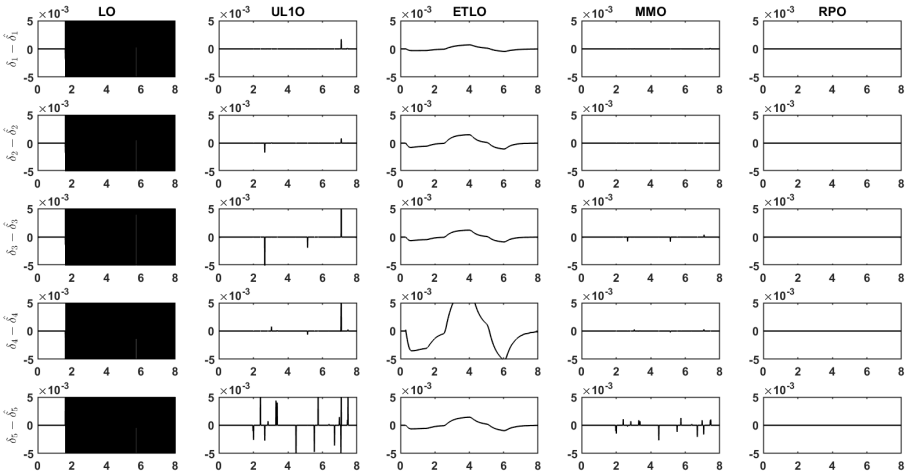

Fig. 10 A comparison result of estimation error of bus angles by 5 observers on IEEE 14-bus system (46). (LO: Luenberger observer, UL1O: unconstrained ℓ_1 observer, ETLO: event-triggered Luenberger observer, MMO: multi-model observer, RPO: resilient pruning observer)

constraints including system updating law and the measurement prior in (13). The core optimization problem solved for the MMO is:

$$\begin{aligned} \text{Minimize} \quad & \sum_{i=k-T+1}^{k} \|\mathbf{y}_i - C\mathbf{x}_i\|_1 \\ \text{Subject to} \quad & \mathbf{x}_{i+1} - A\mathbf{x}_i - B\mathbf{u}_i = 0 \quad j = i-T+1, \cdots, i-1 \\ & \|C\mathbf{x}_k - \mu(\mathbf{z}_k)\|^2_{\Sigma^{-1}(\mathbf{z})} \leq \chi^2_m(\tau) \end{aligned} \tag{47}$$

where, $\chi^2_m(\tau)$ is the quantile function for probability τ of the chi-squared distribution with m degrees of freedom, and τ is the a pre-defined confidence threshold.

ETLO uses event-triggered projected gradient descent technique to achieve fast and reliable solution to the batch optimization problem

$$\begin{aligned} &\text{Minimize: } \|\mathbf{Y}_t - [H \;\; I]\mathbf{z}_t\|_2^2 \\ &\text{Subject to: } \mathbf{z}_t \in \mathbb{R}^n \times \Sigma_{Tk}, \end{aligned} \tag{48}$$

where the decision variable $\mathbf{z}_t$ is an augmented states containing desired initial states and all injected measurement error in T time horizon, $Y_t = [\mathbf{y}_1(t-T+1)^\top \;\; \mathbf{y}_1(t-T+2)^\top \;\; \cdots \;\; \mathbf{y}_1(t)^\top \;\; \cdots\cdots \;\; \mathbf{y}_m(t-T+1)^\top \;\; \mathbf{y}_m(t-T+2)^\top \;\; \cdots \;\; \mathbf{y}_m(t)^\top]^\top \in \mathbf{R}^{Tm}$ is the collection of measurements in T time horizon. A recursive solution to (48) is then implemented as a Luenberger-like update

$$\hat{\mathbf{z}}_t^{(m+1)} = \hat{\mathbf{z}}_t^{(m)} + 2[H \;\; I]^\top (Y_t - [H \;\; I]\hat{\mathbf{z}}_t^{(m)}), \tag{49}$$

alternated with a projection

$$\hat{\mathbf{z}}_\Pi = \Pi(\hat{\mathbf{z}}), \tag{50}$$

where $\Pi : \mathbb{R}^n \times \mathbb{R}^{Tm} \mapsto \mathbb{R}^n \times \Sigma_{Tk}$ is the associated projection operator.

Fig. 10 shows the comparison of the bus angles estimation errors for the different observers. It is seen that the RPO has the least error of all 5 observers. The Luenberger observer is completely unstable as a result of the FDIA, which was designed by compromising 19 sensors measurements. For the MMO, the value of $\tau = 0.1$ was used for the confidence value. For the ETLO, the value of $v = -0.01$ was used for the decreasing level of V.

5.2 Resilient Water Distribution System

In this subsection, we introduce another application example on a water tank coupling control system, shown in Fig. 11. The tank coupling system in [55] was extended to a 11-tanks system which contains 10 operating water tank and a storage tank. The goal is to regulate all operating tanks' water levels around desired values. The magnetic valves v at the entrance pipelines of operating tanks are controlled to adjust the tank water levels. The magnetic valve at the entrance of the storage tank is fixed at a constant opening value. It is assumed that there are water level measurement sensors and pressure senors in the pipelines.The pressure sensors can measure the difference of water level between adjoin tanks on each line. Thus, there are 19 measurements total. The water level adjustment process can be approximated by the LTI model:

$$\begin{aligned} \dot{\mathbf{h}} &= A\mathbf{h} + B\mathbf{v} \\ \mathbf{y} &= C\mathbf{h} \end{aligned} \tag{51}$$

where $\mathbf{h}, \mathbf{v} \in \mathbb{R}^{10}$, $\mathbf{y} \in \mathbf{R}^{19}$. The system dynamics is given by

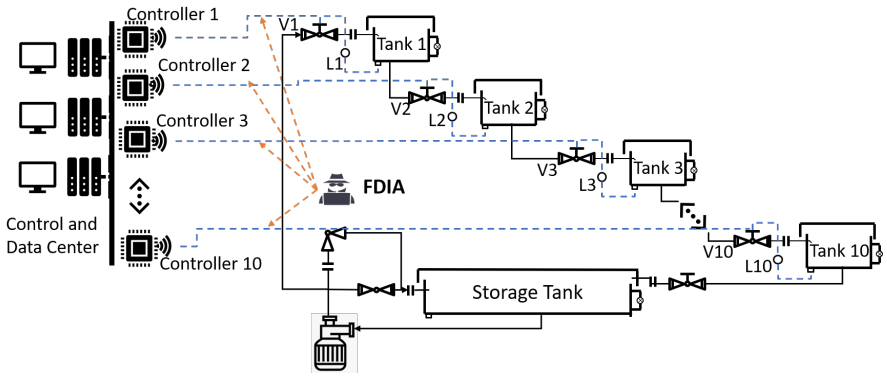

Fig. 11 A water distribution tank coupling control system under false data injection attacks (black solid lines are water pipelines, blue dotted lines are wireless data transmission lines for sensors data and control commands, orange dotted lines are attack injection paths)

$$A = \begin{bmatrix} -0.5815 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.1870 & -0.5906 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.1870 & -0.5127 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1870 & -0.5913 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.1870 & -0.5632 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.1870 & -0.5098 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.1870 & -0.5278 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.1870 & -0.5547 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1870 & -0.5958 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1870 & -0.5965 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.8315 & -0.8450 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.9941 & -0.8450 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.9914 & -0.8450 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.8971 & -0.8450 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.9610 & -0.8450 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.8284 & -0.8450 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.8844 & -0.8450 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9831 & -0.8450 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9584 & -0.8450 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9919 \end{bmatrix}$$

$$C = \begin{bmatrix} & & & & & I_{10} & & & & \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

The model in (51) was discretized using Euler discretization scheme with sampling time $0.01s$. A discrete LQR controller is designed using $Q = 10^3 \times \text{diag}\{2,1,1,2,1,1,2,1,1,2\}$ and $R = 0.2 \times I_{10}$ to obtain the feedback control gain K to regulate the water levels at $\mathbf{h}_d = 0.01 * \mathbf{1}_{10}$, The control law is given by

$$\mathbf{v} = -K(\mathbf{h} - \mathbf{h}_d) - B^{-1}A\mathbf{h}_d + B^{-1}\mathbf{h}_d.$$

The attack percentage is set as $P_A = 0.6$, and by using the designed FDIA (6), it can bypass the bad data detection threshold. Due to lack of actual auxiliary data for this case, a sample support prior is created by generating uniformly distributed random numbers in the interval $[0,\ \ 1]$ for each measurement node. These numbers represent the localization confidence values $\mathbf{p}_i$'s used in Algorithm 2. The generated prior information represents a localization algorithm whose performance is comparable to random flip of a fair coin. The reason for this is to show how the observers perform using a relatively poor localization algorithm. The precision of the generated support prior is shown in Fig. 12, the mean of precision is 0.5588. For a more realistic situation, possible candidate auxiliary variables include atmospheric data like temperature, humidity, atmospheric pressure, or any other values that can affect the flow of water in a long pipe. Market data and time of day are also great candidates for auxiliary variables.

Then the resilient estimation schemes descried in last subsection are also implemented for this system. The comparison of the resulting estimation errors is presented in Table 1, in which relative mean square error and maximum absolute error are given. Again, as seen in the table, the RPO outperforms the other observers in terms of the given error metrics.

5.3 Resilient Wheeled Mobile Robot

For this example, a nonlinear observer scheme based on prior information is given for the resilient motion control of wheeled mobile robot. Non-holonomic wheeled mobile robot is considered with IIoT sensors, its dynamical and kinematic model

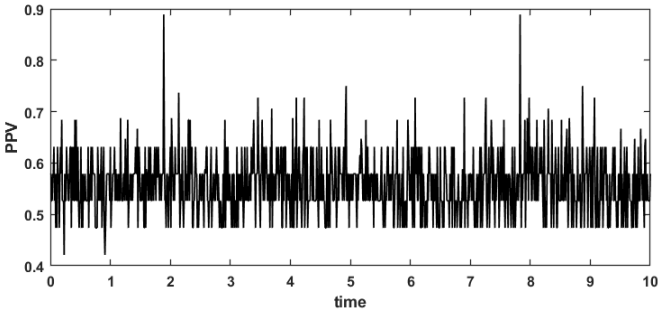

Fig. 12 The precision of support prior generated by the localization algorithm in Algorithm 1 for water tank coupling system (The mean of precision is 0.5588)

	RMS Metric				Max. Ans. Metric			
	LO	**UL1O**	**MMO**	**RPO**	**LO**	**UL1O**	**MMO**	**RPO**
$\mathbf{e}_1$	1.4434	2.5657e-6	2.0794e-6	3.5704e-10	21.8421	4.6746e-5	4.6655e-5	5.7127e-9
$\mathbf{e}_2$	1.5088	5.8117e-6	2.1444e-8	4.3079e-10	23.0772	1.5826e-4	5.1996e-7	5.7700e-9
$\mathbf{e}_3$	0.8018	4.1172e-8	2.1381e-10	4.3901e-10	13.9374	1.1886e-6	3.4310e-9	9.6873e-9
$\mathbf{e}_4$	0.7350	4.5476e-6	4.5476e-6	3.2479e-10	14.4943	1.4388e-4	1.4388e-4	4.4373e-9
$\mathbf{e}_5$	0.5645	2.2444e-5	1.7216e-5	3.5302e-10	9.8116	4.7845e-4	3.7122e-4	4.8049e-9
$\mathbf{e}_6$	1.0332	3.3578e-5	1.7473e-5	4.1156e-10	15.5191	5.5419e-4	3.7748e-4	8.1021e-9
$\mathbf{e}_7$	1.1802	2.2776e-5	1.6834e-5	3.8149e-10	17.1720	4.1583e-4	3.7724e-4	5.5387e-9
$\mathbf{e}_8$	1.2172	3.8198e-5	2.2591e-6	1.1470e-6	20.5512	0.0010	6.1343e-5	3.6289e-5
$\mathbf{e}_9$	0.9802	2.2720e-5	2.2118e-5	2.0543e-6	18.1152	3.4706e-4	3.4706e-4	6.2776e-5
$\mathbf{e}_{10}$	2.6151	1.0291e-4	2.0641e-6	2.3344e-7	28.5826	0.0030	6.1424e-5	7.3509e-6

Table 1 Error metric values for four resilient observers on water tank coupling system (RMS Metric: relative mean square error, Max. Ans. Metric: maximum absolute error)

can be described as [56]

$$\begin{aligned} \dot{\mathbf{q}} &= M^{-1}(-D\mathbf{q}+B\tau)+\mathbf{w} \triangleq g(\mathbf{x},\mathbf{u})+\mathbf{w} \\ \begin{bmatrix} \dot{\theta} \\ \cdots \\ \dot{\mathbf{z}} \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ \cdots & \\ C(\theta) & \end{bmatrix} \mathbf{q} \triangleq \bar{C}(\theta)\mathbf{q}, \end{aligned} \tag{52}$$

where, $\mathbf{q} = [v \;\; \omega]^\top$ is the generalized body velocities vector, $\mathbf{u} \triangleq \tau = [\tau_R \;\; \tau_L]^\top$ is a vector of the wheels torques, and $\mathbf{z} = [x \;\; y]^\top$ is the task-space position vector, $\mathbf{x} = [\theta \;\; v \;\; \omega]^\top$ is defined as a state vector, $\mathbf{w} \sim \mathcal{N}(0,R)$ is the process noise in dynamics. The kinematic and dynamical parameters are given by:

$$\begin{aligned} M &= \begin{bmatrix} m & 0 \\ 0 & md^2+J \end{bmatrix}, \;\; D = \begin{bmatrix} 0 & -md\omega \\ md\omega & 0 \end{bmatrix} \\ B &= \frac{1}{r}\begin{bmatrix} 1 & 1 \\ L & -L \end{bmatrix}, \;\; C(\theta) = \begin{bmatrix} \cos(\theta) & -d\sin(\theta) \\ \sin(\theta) & d\cos(\theta) \end{bmatrix}. \end{aligned}$$

The corresponding measurements system is given by

$$\mathbf{y} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1/4r & L/4r \\ 1/4r & -L/4r \\ \cos(\theta) & -d\sin(\theta) \\ \sin(\theta) & d\cos(\theta) \end{bmatrix} \cdot \mathbf{q} + \mathbf{v} \triangleq f(\mathbf{x}) + \mathbf{v} + \mathbf{e}, \tag{53}$$

where $\mathbf{v}$ denotes measurement noises, $\mathbf{e}$ denotes the attack vector.

Given a desired 2D "Figure-8" path described by the continuous function:

$$\mathbf{z}_d = \begin{bmatrix} x_d(t) \\ y_d(t) \end{bmatrix} = \begin{bmatrix} \frac{a\cos(t)}{1+\sin^2(t)} \\ \frac{a\sin(t)\cos(t)}{1+\sin^2(t)} \end{bmatrix},$$
$$\theta_d(t) = \arctan\left(\frac{y_d(t)}{x_d(t)}\right),$$

a stable path-tracking controller was given in [16] as

$$\tau = B^{-1}(M\mathbf{u} + D\mathbf{q}), \tag{54}$$

where,

$$\mathbf{u} = -k_q(\mathbf{q} - \mathbf{q}_d) + \dot{\mathbf{q}}_d - \bar{C}(\theta)^\top \tilde{\mathbf{e}}$$

with

$$\mathbf{q}_d = C^{-1}(\theta)(\dot{\mathbf{z}}_d - k_e\mathbf{e}_\mathbf{z}),$$
$$\dot{\mathbf{q}}_d = -k_e(\dot{C}^{-1}(\theta)\mathbf{e}_\mathbf{z} + \mathbf{q}) + C^{-1}(\theta)[\ddot{\mathbf{z}}_d + (k_e + C(\theta)\dot{C}^{-1}(\theta))\dot{\mathbf{z}}_d].$$

and k_q, k_e are positive scalar control gains.

Next task is to design a nonlinear observer to recover the real state $\mathbf{x}$ under the compromised measurements $\mathbf{y}$, shown in Fig. 13. According to Theorem 1, the precision of $\hat{\mathscr{T}}_\eta^c$ can achieve 100% with a probability lower bound. Thus, Unscented Kalman Filter (UKF) can be used on the safe subset of measurements denoted by $\hat{\mathscr{T}}_\eta^c$. The control system with resilient Kalman filter is shown schematically in Fig. 13. Fig. 14 and Fig. 15 show comparisons of the tracking performances between UKF, UKF with the prior , and UKF with prior and pruning. It is well-known in literature that KF cannot recover exact states in the presence of FDIA. Fig. 14 and Fig. 15 confirms this fact. Specifically, it is seen that the path-tracking task and state estimation totally fail with only UKF. By adding a prior information obtained by the localization algorithm whose mean of precision is around 0.6, the motion control performance is improved but has big oscillatory due to the imperfect precision. However, with the developed pruning algorithm, the robot was able to track the reference path very closely and smoothly.

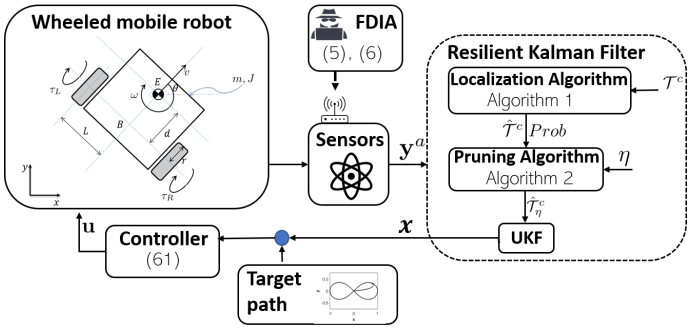

Fig. 13 Block diagram depiction the resilient motion control of wheeled mobile robot

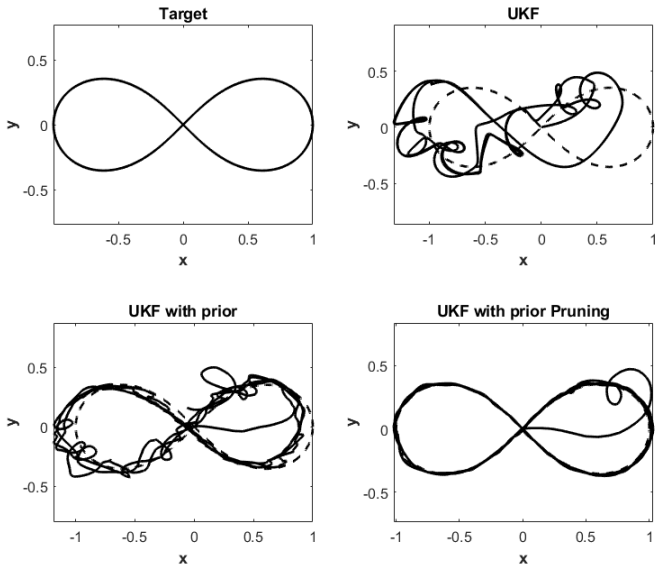

Fig. 14 Path tracking performance (UKF: unscented Kalman filter, UKF with prior: unscented Kalman filter with the prior generated by localization algorithm in Algorithm 1, UKF with prior pruning: unscented Kalman filter with pruned prior generated by Algorithm 2)

6 Conclusion

In this chapter, we talk through the resilient observer design with prior pruning. Firstly, we prove that a good support prior (better than random flip of fair coin) improve the attack-resiliency boundary (50%) of resilient observers admitted by literature. Secondly, the pruning algorithm is a method to improve the precision of support prior without training effort, which make prior information more useful in control scenario. Finally, the proposed pruning observer scheme showed its superior in worse attacking situation compared to other resilient observers in literature.

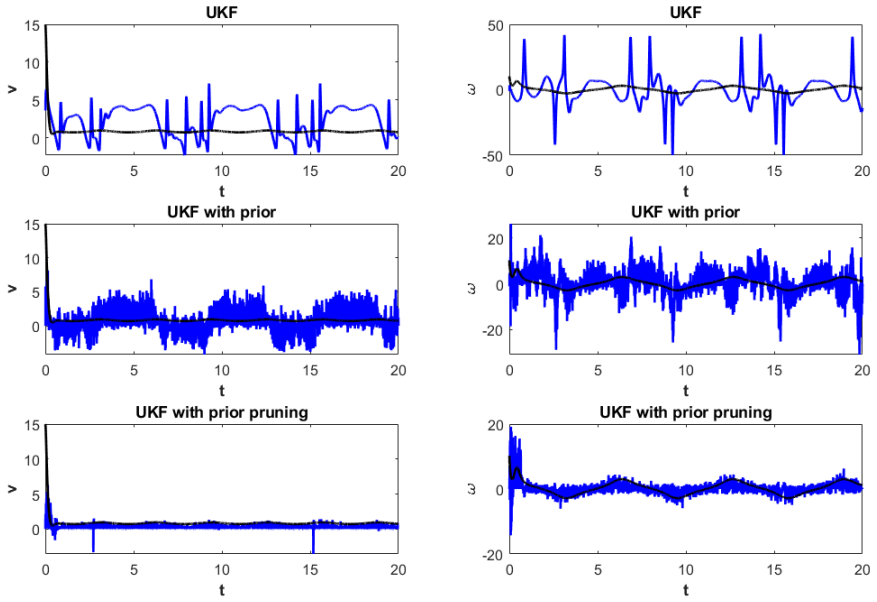

Fig. 15 Estimations of robot's forward velocity v and angular velocity ω by three observers (Black line is the nominal state estimation, blue line is the estimation by those three observers in presence of FDIA)

Moreover, another minor contributions of this chapter include a formal definition of successful FDIA and optimization-based FDIA design.

References

1. Khaitan, S. K., McCalley, J. D. (2014). Design techniques and applications of cyber physical systems: A survey. IEEE Systems Journal, 9(2), 350-365.
2. Zetter, K. (2015). A cyber attack has caused confirmed physical damage for the second time ever.
3. Lee, R. M., Assante, M. J., Conway, T. (2014). German steel mill cyber attack. Industrial Control Systems, 30, 62.
4. Slay, J., Miller, M. (2007, March). Lessons learned from the maroochy water breach. In International conference on critical infrastructure protection (pp. 73-82). Springer, Boston, MA.
5. Chen, T. M., Abu-Nimeh, S. (2011). Lessons from stuxnet. Computer, 44(4), 91-93.
6. Burg, A., Chattopadhyay, A., Lam, K. Y. (2017). Wireless communication and security issues for cyber–physical systems and the Internet-of-Things. Proceedings of the IEEE, 106(1), 38-60.
7. Bishop, M. (2003). What is computer security?. IEEE Security and Privacy, 1(1), 67-69.
8. Pelechrinis, K., Iliofotou, M., Krishnamurthy, S. V. (2010). Denial of service attacks in wireless networks: The case of jammers. IEEE Communications surveys and tutorials, 13(2), 245-257.

9. Liu, Y., Ning, P., Reiter, M. K. (2011). False data injection attacks against state estimation in electric power grids. ACM Transactions on Information and System Security (TISSEC), 14(1), 1-33.
10. Guo, Z., Shi, D., Johansson, K. H., Shi, L. (2016). Optimal linear cyber-attack on remote state estimation. IEEE Transactions on Control of Network Systems, 4(1), 4-13.
11. Mo, Y., Sinopoli, B. (2010, April). False data injection attacks in control systems. In Preprints of the 1st workshop on Secure Control Systems (pp. 1-6).
12. Sui, T., Mo, Y., Marelli, D., Sun, X. M., Fu, M. (2020). The Vulnerability of Cyber-Physical System under Stealthy Attacks. IEEE Transactions on Automatic Control.
13. Pasqualetti, F., Dörfler, F., Bullo, F. (2013). Attack detection and identification in cyber-physical systems. IEEE transactions on automatic control, 58(11), 2715-2729.
14. Fang, C., Qi, Y., Cheng, P., Zheng, W. X. (2020). Optimal periodic watermarking schedule for replay attack detection in cyber–physical systems. Automatica, 112, 108698.
15. de Sá, A. O., da Costa Carmo, L. F. R., Machado, R. C. (2017). Covert attacks in cyber-physical control systems. IEEE Transactions on Industrial Informatics, 13(4), 1641-1651.
16. Y. Zheng and O. M. Anubi. (2020) "Attack-resilient observer pruning for path-tracking control of wheeled mobile robot," in2020 ASME Dynamic Systems and Control(DSC) Conference, ASME, pp. 1–9.
17. Fawzi, H., Tabuada, P., Diggavi, S. (2014). Secure estimation and control for cyber-physical systems under adversarial attacks. IEEE Transactions on Automatic control, 59(6), 1454-1467.
18. Shoukry, Y., Tabuada, P. (2015). Event-triggered state observers for sparse sensor noise/attacks. IEEE Transactions on Automatic Control, 61(8), 2079-2091.
19. Chong, M. S., Wakaiki, M., Hespanha, J. P. (2015, July). Observability of linear systems under adversarial attacks. In 2015 American Control Conference (ACC) (pp. 2439-2444). IEEE.
20. Pajic, M., Tabuada, P., Lee, I., Pappas, G. J. (2015, December). Attack-resilient state estimation in the presence of noise. In 2015 54th IEEE Conference on Decision and Control (CDC) (pp. 5827-5832). IEEE.
21. Lee, C., Shim, H., Eun, Y. (2015, July). Secure and robust state estimation under sensor attacks, measurement noises, and process disturbances: Observer-based combinatorial approach. In 2015 European Control Conference (ECC) (pp. 1872-1877). IEEE.
22. Nakahira, Y., Mo, Y. (2018). Attack-Resilient $\mathscr{H}_2$, $\mathscr{H}_\infty$, and ℓ_1 State Estimator. IEEE Transactions on Automatic Control, 63(12), 4353-4360.
23. Shoukry, Y., Nuzzo, P., Puggelli, A., Sangiovanni-Vincentelli, A. L., Seshia, S. A., Tabuada, P. (2017). Secure state estimation for cyber-physical systems under sensor attacks: A satisfiability modulo theory approach. IEEE Transactions on Automatic Control, 62(10), 4917-4932.
24. Shinohara, T., Namerikawa, T., Qu, Z. (2019). Resilient reinforcement in secure state estimation against sensor attacks with a priori information. IEEE Transactions on Automatic Control, 64(12), 5024-5038.
25. Anubi, O. M., Konstantinou, C., Wong, C. A., Vedula, S. (2020, August). Multi-Model Resilient Observer under False Data Injection Attacks. In 2020 IEEE Conference on Control Technology and Applications (CCTA) (pp. 1-8). IEEE.
26. Anubi, O. M., Konstantinou, C. (2019). Enhanced resilient state estimation using data-driven auxiliary models. IEEE Transactions on Industrial Informatics, 16(1), 639-647.
27. Anubi, O. M., Mestha, L., Achanta, H. (2018). Robust resilient signal reconstruction under adversarial attacks. arXiv preprint arXiv:1807.08004.
28. Liu, H., Mo, Y., Yan, J., Xie, L., Johansson, K. H. (2020). An online approach to physical watermark design. IEEE Transactions on Automatic Control, 65(9), 3895-3902.
29. Weerakkody, S., Sinopoli, B. (2015, December). Detecting integrity attacks on control systems using a moving target approach. In 2015 54th IEEE Conference on Decision and Control (CDC) (pp. 5820-5826). IEEE.
30. Esmalifalak, M., Liu, L., Nguyen, N., Zheng, R., Han, Z. (2014). Detecting stealthy false data injection using machine learning in smart grid. IEEE Systems Journal, 11(3), 1644-1652.

31. He, Y., Mendis, G. J., Wei, J. (2017). Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism. IEEE Transactions on Smart Grid, 8(5), 2505-2516.
32. Ozay, M., Esnaola, I., Vural, F. T. Y., Kulkarni, S. R., Poor, H. V. (2015). Machine learning methods for attack detection in the smart grid. IEEE transactions on neural networks and learning systems, 27(8), 1773-1786.
33. Abbaszadeh, M., Mestha, L. K., Bushey, C., Holzhauer, D. F. (2019). U.S. Patent No. 10,417,415. Washington, DC: U.S. Patent and Trademark Office.
34. Deldjoo, Y., Noia, T. D., Merra, F. A. (2021). A survey on adversarial recommender systems: from attack/defense strategies to generative adversarial networks. ACM Computing Surveys (CSUR), 54(2), 1-38.
35. Huang, Y., Tang, J., Cheng, Y., Li, H., Campbell, K. A., Han, Z. (2014). Real-time detection of false data injection in smart grid networks: An adaptive CUSUM method and analysis. IEEE Systems Journal, 10(2), 532-543.
36. Zheng, Y., Anubi, O. M. (2021). Attack-Resilient Weighted ℓ_1 Observer with Prior Pruning. In 2021 American Control Conference (ACC).
37. Anubi, O. M., Konstantinou, C., Roberts, R. (2019). Resilient optimal estimation using measurement prior. arXiv preprint arXiv:1907.13102.
38. Lee, E. A. (2010, June). CPS foundations. In Design Automation Conference (pp. 737-742). IEEE.
39. Derler, P., Lee, E. A., Vincentelli, A. S. (2011). Modeling cyber–physical systems. Proceedings of the IEEE, 100(1), 13-28.
40. Mo, Y., Sinopoli, B. (2015). On the performance degradation of cyber-physical systems under stealthy integrity attacks. IEEE Transactions on Automatic Control, 61(9), 2618-2624.
41. Rasmussen, C. E. (2003, February). Gaussian processes in machine learning. In Summer school on machine learning (pp. 63-71). Springer, Berlin, Heidelberg.
42. Urtasun, R., Darrell, T. (2008, June). Sparse probabilistic regression for activity-independent human pose inference. In 2008 IEEE Conference on Computer Vision and Pattern Recognition (pp. 1-8). IEEE.
43. Liu, M., Chowdhary, G., Da Silva, B. C., Liu, S. Y., How, J. P. (2018). Gaussian processes for learning and control: A tutorial with examples. IEEE Control Systems Magazine, 38(5), 53-86.
44. Fawcett, T. (2006). An introduction to ROC analysis. Pattern recognition letters, 27(8), 861-874.
45. Donoho, D. L. (2006). Compressed sensing. IEEE Transactions on information theory, 52(4), 1289-1306.
46. Candès, E. J., Romberg, J., Tao, T. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on information theory, 52(2), 489-509.
47. Candes, E. J., Tao, T. (2005). Decoding by linear programming. IEEE transactions on information theory, 51(12), 4203-4215.
48. Candes, E., Tao, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. Annals of statistics, 35(6), 2313-2351.
49. Fornasier, M., Rauhut, H. (2015). Compressive Sensing. Handbook of mathematical methods in imaging, 1, 187-229.
50. Friedlander, M. P., Mansour, H., Saab, R., Yilmaz, Ö. (2011). Recovering compressively sampled signals using partial support information. IEEE Transactions on Information Theory, 58(2), 1122-1134.
51. N. Y. I. S. Operator, "Load Data," [Online]. `https://www.nyiso.com/load-data`
52. Power System Test Case Archive, 14 bus power flow test case. [Online]. `http://labs.ece.uw.edu/pstca/pf14/pg_tca14bus.htm`
53. New York control area load zone map. [Online]. `https://www.nyiso.com/documents/20142/1397960/nyca_zonemaps.pdf`

54. Scholtz, E. (2004). Observer-based monitors and distributed wave controllers for electromechanical disturbances in power systems (Doctoral dissertation, Massachusetts Institute of Technology).
55. Yang, J., Zhou, C., Tian, Y. C., An, C. (2020). A Zoning-Based Secure Control Approach Against Actuator Attacks in Industrial Cyber-Physical Systems. IEEE Transactions on Industrial Electronics, 68(3), 2637-2647.
56. Dhaouadi, R., Hatab, A. A. (2013). Dynamic modelling of differential-drive mobile robots using lagrange and newton-euler methodologies: A unified framework. Advances in Robotics & Automation, 2(2), 1-7.