Attack-Resilient Weighted ℓ_1 Observer with Prior Pruning

Yu Zheng[1] and Olugbenga Moses Anubi[2]

***Abstract*— Security related questions for Cyber Physical Systems (CPS) have attracted much research attention in searching for novel methods for attack-resilient control and/or estimation. Specifically, false data injection attacks (FDIAs) have been shown to be capable of bypassing bad data detection (BDD), while arbitrarily compromising the integrity of state estimators and robust controller even with very sparse measurements corruption. Moreover, based on the inherent sparsity of pragmatic attack signals, ℓ_1-*Minimization* scheme has been used extensively to improve design attack-resilient estimators. For this, the theoretical maximum for the percentage of compromised nodes that can be accommodated has been shown to be 50%. In order to guarantee correct state recoveries for larger percentage of attacked nodes, 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 tolerated attacked percentages and the precision of the *prior* information. In this paper, we present a *pruning* method to improve the precision of the *prior* information, given a stochastic uncertainty characteristics of the underlying Machine Learning model. Then a *Weighted* ℓ_1-*Minimization* is proposed based on the *pruned prior*. The theoretical and simulation results show that the pruning method significantly improves the observer performance for much larger attack percentages, even when moderately accurate Machine Learning model used.**

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 positive real numbers. Normal-face lower-case letters ($e.g.\ x \in \mathbb{R}$) are used to represent real scalars, bold-face lower-case letter ($e.g\ \mathbf{x} \in \mathbb{R}^n$) represents vectors, while normal-face upper case ($e.g.\ X \in \mathbb{R}^{n\times m}$) represents matrices. Let $\mathcal{T} \subseteq \{1,\ldots,n\}$, then for a matrix $X \in \mathbb{R}^{m\times n}$, $X_{\mathcal{T}} \in \mathbb{R}^{m\times|\mathcal{T}|}$ is the sub-matrix obtained by extracting the columns of X corresponding to the indices in $\mathcal{T}$. $\mathcal{T}^c$ denotes the complement of a set $\mathcal{T}$ and the universal set on which it is defined will be clear from the context. $\mathcal{S}_k^m$ denotes the set of all vectors $\mathbf{v} \in \mathbb{R}^m$ such that $|\mathsf{supp}(\mathbf{v})| \leq k$ (i.e the set of k-sparse vectors). $X_{.}^{\top}$ denotes a combination of transposition of the sub-matrix of X ($e.g.\ X = [X_1\ \ X_2]$, $X.^{\top} = \begin{bmatrix} X_1^{\top} \\ X_2^{\top} \end{bmatrix}$). The symbol $*$ denotes the convolution operator for vectors. $\mathsf{supp}(\mathbf{x})$ denotes the support of the vector $\mathbf{x}$ given by the set $\mathcal{T} = \mathsf{supp}(\mathbf{x}) = \{i|\mathbf{x}_i \neq 0\}$. $\mathsf{argsort}\downarrow(\mathbf{x})$ denotes a function that returns the sorted indices of vector $\mathbf{x}$ in descending order.

The *best k-sparse approximation error*, measured by ℓ_p norm, is given, for $\mathbf{e} \in \mathcal{S}_k^m$,

$$\sigma_k(\mathbf{e})_{\ell_p} \triangleq \min_{\mathbf{z}\in\mathcal{S}_k^n} \|\mathbf{e}-\mathbf{z}\|_{\ell_p} = \|\mathbf{e}_{\mathcal{T}^c}\|_{\ell_p}. \qquad (1)$$

where $\mathcal{T}$ is the support of $\mathbf{e}_i$ with first k largest magnitude.

The symbol $\circ$ denotes element-wise multiplication of two vectors and is defined as $\mathbf{z} = \mathbf{x} \circ \mathbf{y}$, where $\mathbf{z}_i = \mathbf{x}_i \cdot \mathbf{y}_i$.

I. INTRODUCTION

Cyber-physical System has application potential in various areas [1]. The authors in [2] pointed out that the ideal CPS must operate dependably, safely, securely, efficiently and in real-time.

Security questions in CPSs are more challenging than traditional IT security because of the combination of temporal dynamics brought by the physical environment and the heterogeneous nature of the operation of CPSs [3]. Failure of CPS is more complicated than random failures or well-defined uncertainty for which many results exist on reliability and robustness, since they may be caused by stealth malicious attacks. One of such powerful deception attacks, named *false data injection attack* (FDIA), has shown ability to bypass *bad data detector* (BDD), while compromising the integrity of observer and robust controller with sparse measurements corruption [4, 5]. Consequently, much research attention have been directed to develop appropriate protection schemes.

Active detection approaches have been considered [6], where the defenders adjust the detection rules online in order to identify the attack scenarios. There are some *machine learning* algorithms being considered to localize the attacks, such as *Gaussian process regression* [7], *support vector machine*[8], *markov graphs*[9], *generative adversarial networks*[10] and more. These *machine learning* localization algorithms generate estimated support of attacked (or safe) nodes. This can then be used as a *prior* information for a resilient estimation program.

Due to sparsity assumption of the attack vector $|\mathsf{supp}(\mathbf{e})| \leq k$, the resilient estimation problem has been cast as a classical error correction problem [5, 11]. Consider a linear observation model $\mathbf{y} = H\mathbf{x} + \mathbf{e}$, where $H \in \mathbb{R}^{N\times n}$ denotes an observation matrix, then the resilient estimation is formulated as 0-norm minimization problem [12]. But it imposes a restriction of maximum attack percentage of 50%

[1] Yu Zheng is Ph.D candidate in Department of Electrical and Computer Engineering, Florida State University, Tallahassee, FL 32310, USA yz19b@fsu.edu

[2] Olugbenga Moses Anubi is IEEE member, Assistant professor of the Department of Electrical and Computer Engineering, Center for Advanced Power systems, Florida State University, Tallahassee, FL 32310, USA oanubi@fsu.edu

for correct recovery of $\mathbf{x}$. Moreover, since 0-norm minimization decoder is an NP-hard problem, an alternative 1-norm minimization decoder has been considered in literature [7, 12], which can be solved by linear programming [13]. The condition to bridge the two decoders is *Restricted Isometry Property* (RIP) [14] that defines the sparse recoverability of observation matrix H.

In order to guarantee correct state recoveries for larger percentage of attacks, *prior* information has been considered for resilient estimation scheme in literature: *Measurement Prior* [7, 15], *Support Prior* [5] and *State Prior* [16]. In [16], the author considered the *Prior* information of estimated states in three forms: *sparsity information* of $\mathbf{x}_0$, $(\alpha, \bar{n}_0)$ *sparsity information* where α replaces 0 in sparsity definition, and *side information* that is knowledge of the initial state from the physical attribution of the system and cannot be manipulated by malicious third parties. In [7, 15], the author constructed a data-driven auxiliary model between system measurement and auxiliary state by trained *Gaussian Process Regression* (GPR), the attacked measurements are visible to defender if they cannot be explained based on the *measurement model prior* with high likelihood. In this paper, we consider *support prior* which gives an estimated set of attack location, and is generated by any of the afore mentioned localization algorithms. However, there are two drawbacks, namely: uncertainty and training price. Thus, we propose a *Pruning* method to improve the precision of the *support prior* without training process. Then a weighted 1-norm minimization scheme [17] is given based on the resulting *pruned support prior*. The pruning idea originates from [5]. Moreover, due to the perfect localization precision of *pruned support prior*, resilient *Unscented Kalman Filter* (UKF) against FDIA was given in [18] by performing UKF based on the pruned safe set.

The remainder of this paper is organized as follows. In *Section* **??**, we describe the concurrent models that will be used for the development in subsequent sections, including physical model of CPS, threat model and prior model. In *Section* **??**, we develop the pruning methods, and construct a weighted ℓ_1 observer with *pruned support prior*. A numerical simulation and an application simulation on IEEE-14 bus system show the proposed observer indeed enhance the system resilience in *Section IV*. Finally, conclusion remarks follows in *Section V*.

II. PROBLEM STATEMENT

A physical attacked dynamic model of CPS can be given:

$$\begin{aligned} \mathbf{x}_{i+1} &= A\mathbf{x}_i \\ \mathbf{y}_i &= C\mathbf{x}_i + \mathbf{e}_i \end{aligned} \tag{2}$$

where, $\mathbf{x}_k \in \mathbb{R}^n, \mathbf{y}_k \in \mathbb{R}^m$, with $m > n$, denote state vector and measurement vector at time i respectively, $\mathbf{e}_k \in \mathcal{S}_k^m$ denotes the sparse attack vector, A, C are system dynamic parameters. A control input may be included in the model above. However, since the control input is generally irrelevant to state estimation problems, we suppress in the model considered here. By iterating the system model (2) T time steps, the T horizon observation model is given

$$\mathbf{y}_T = H\mathbf{x}_{i-T+1} + \mathbf{e}_T \tag{3}$$

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 $i-T+1$, $\mathbf{e}_T = [\mathbf{e}_i^\top \;\; \mathbf{e}_{i-1}^\top \cdots \mathbf{e}_{i-T+1}^\top]^\top \in \mathcal{S}_k^{Tm}$ is the sequence of attack vectors in the same moving window with $\mathbf{e}_i \in \mathcal{S}_{k/T}^m, \forall i \in [i-T+1 \;\; i]$ and

$$H = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{T-1} \end{bmatrix} = U \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix} V^\top$$

where, $U = [U_1 \;\; U_2]$, with $U_1 \in \mathbb{R}^{Tm \times n}, U_2 \in \mathbb{R}^{Tm \times Tm-n}$, is a matrix of left singular vectors of H, $V \in \mathbb{R}^{n \times n}$ is the corresponding matrix of right singular vectors, $\Sigma_1 \in \mathbb{R}^{n \times n}$ is a diagonal matrix of the singular values of H, which are non-zero since H is full rank.

Definition 1 (Decoder): Given a sequence of observation $\mathbf{y}_T \in \mathbb{R}^{Tm}$, the decoder for the measurement model in (3) is a mapping of the form $\mathcal{D} : \mathbb{R}^{Tm} \mapsto \mathbb{R}^n$ given by

$$\hat{\mathbf{x}} = \mathcal{D}(\mathbf{y}_T) = V\Sigma_1^{-1} \arg\min_{\mathbf{z}} \|\mathbf{y}_T - U_1\mathbf{z}\|_{\ell_1} \tag{4}$$

Definition 2 (Detector): Given a positive threshold parameter ε, a detector for the decoder $\mathcal{D}$ and the measurement model in (3) is a binary classifier of the form

$$\mathcal{D}_\varepsilon(\mathbf{y}_T) = \begin{cases} 1 & \text{if } \|\mathbf{y}_T - H\mathcal{D}(\mathbf{y}_T)\|_{\ell_1} > \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Remark 1: The positive class for $\mathcal{D}_\varepsilon(\cdot)$ is the set of attacked (unsafe) measurements, while the negative class is the set of safe measurements.

False data injection attack is modeled as an addictive bias through $\mathbf{e}$ on the sparse measurement nodes. We will assume the attacker has knowledge of system (A, C).

Definition 3 (Successful FDIA [4]): Consider the CPS in (2) and the corresponding measurement model (3), the attack sequence $\mathbf{e}_T \in \mathcal{S}_k^{Tm}$ is said to be (ε, α)-successful against the decoder-detector pair $\{\mathcal{D}, \mathcal{D}_\varepsilon\}$ if

$$\|\mathbf{x}^\star - \mathcal{D}(\mathbf{y}_T)\| \ge \alpha, \;\; \mathcal{D}_\varepsilon(\mathbf{y}_T) = 0, \tag{6}$$

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

Remark 2: (ε, α)-successful FDIA can result in a state bias of size at least α while by-passing the detector $\mathcal{D}_\varepsilon(\cdot)$.

The following theorem gives a machanism for constructing such (ε, α)-successful FDIA if the attack support is predetermined.

Theorem 2.1: Given the support sequence $\mathcal{T} = \{\mathcal{T}_i \;\; \mathcal{T}_{i-1} \cdots \mathcal{T}_{i-T+1}\}$ with $|\mathcal{T}_i| \le k$. Let $\mathbf{z}_e$ be an optimal solution of the optimization program

$$\begin{aligned} \text{Maxmize :} \quad & \|\Sigma_1^{-1}\mathbf{z}\|_{\ell_2}, \\ \text{Subject to :} \quad & \|U_{1,\mathcal{T}^c}\mathbf{z}\|_{\ell_2} \le \varepsilon. \end{aligned} \tag{7}$$

If $\bar{\sigma}_{\mathcal{T}^c}\sqrt{Tm-|\mathcal{T}|}<1$, then the FDIA

$$\mathbf{e}_{\mathcal{T}} = U_{1,\mathcal{T}}\cdot\mathbf{z}_e,\ \ \mathbf{e}_{\mathcal{T}^c}=\mathbf{0} \tag{8}$$

is (ε,α)-successful against the decoder-detector pair $\{\mathcal{D},\mathcal{D}_\varepsilon\}$, with

$$\alpha = \frac{\varepsilon}{\underline{\sigma}\overline{\sigma}_{\mathcal{T}^c}}(1-\overline{\sigma}_{\mathcal{T}^c}\sqrt{Tm-|\mathcal{T}|}),$$

where $\overline{\sigma}_{\mathcal{T}^c}$ is the largest singular value of $U_{1,\mathcal{T}^c}$ and $\underline{\sigma}$ is the smallest nonzero singular value of H.

Proof: Since

$$\left\{\mathbf{z}\,\middle|\ \|\mathbf{z}\|_{\ell_2}\le\frac{\varepsilon}{\overline{\sigma}_{\mathcal{T}^c}}\right\}\subset\left\{\mathbf{z}\,\middle|\ \|U_{1,\mathcal{T}^c}\mathbf{z}\|_{\ell_2}\le\varepsilon\right\},$$

where $\overline{\sigma}_{\mathcal{T}^c}$ is the largest singular value of $U_{1,\mathcal{T}^c}$, then

$$\|\Sigma_1^{-1}\mathbf{z}_e\|_{\ell_2}\ge\max_{\|\mathbf{z}\|_{\ell_2}\le\frac{\varepsilon}{\overline{\sigma}_{\mathcal{T}^c}}}\|\Sigma_1^{-1}\mathbf{z}\|_{\ell_2}\ge\frac{\varepsilon}{\underline{\sigma}\overline{\sigma}_{\mathcal{T}^c}}$$

Also

$$\hat{\mathbf{x}}=\mathcal{D}\left(\mathbf{y}_T^\star+P\begin{bmatrix}U_{1,\mathcal{T}}\\0\end{bmatrix}\mathbf{z}_e\right)$$

for an appropriate permutation matrix P satisfying $U_1=P\begin{bmatrix}U_{1,\mathcal{T}}\\U_{1,\mathcal{T}^c}\end{bmatrix}$. Hence

$$\begin{aligned}\hat{\mathbf{x}}&=V\Sigma_1^{-1}\arg\min_{\mathbf{z}}\left\|U_1(\mathbf{z}^\star+\mathbf{z}_e+\mathbf{z})-P\begin{bmatrix}0\\U_{1,\mathcal{T}^c}\end{bmatrix}\mathbf{z}_e\right\|_{\ell_1}\\&=V\Sigma_1^{-1}(\mathbf{z}^\star+\mathbf{z}_e-\mathbf{z}_e^\perp)\end{aligned}$$

where $\mathbf{z}^\star=\Sigma_1V^\top\mathbf{x}^\star$ corresponds to the true state and $\mathbf{z}_e^\perp$ is the projection given by

$$\mathbf{z}_e^\perp=\arg\min_{\mathbf{z}}\left\|U_1\mathbf{z}-P\begin{bmatrix}0\\U_{1,\mathcal{T}^c}\end{bmatrix}\mathbf{z}_e\right\|_{\ell_1}$$

Thus,

$$\begin{aligned}\|\hat{\mathbf{x}}\|_{\ell_2}&=\|\Sigma_1^{-1}(\mathbf{z}_e-\mathbf{z}_e^\perp)\|_{\ell_2}\\&\ge\|\Sigma_1^{-1}\mathbf{z}_e\|_{\ell_2}-\|\Sigma_1^{-1}\mathbf{z}_e^\perp\|_{\ell_2}\ge\frac{\varepsilon}{\underline{\sigma}\overline{\sigma}_{\mathcal{T}^c}}-\frac{1}{\underline{\sigma}}\|\mathbf{z}_e^\perp\|_{\ell_2}\end{aligned}$$

Note[1]

$$\begin{aligned}\|\mathbf{z}_e^\perp\|_{\ell_2}\le\|\mathbf{z}_e^\perp\|_{\ell_1}&\le\|U_{1,\mathcal{T}^c}\mathbf{z}_e\|_{\ell_1}\\&\le\sqrt{Tm-|\mathcal{T}|}\cdot\|U_{1,\mathcal{T}^c}\mathbf{z}_e\|_{\ell_2}\\&\le(\sqrt{Tm-|\mathcal{T}|})\varepsilon\end{aligned}$$

Therefore,

$$\begin{aligned}\|\hat{\mathbf{x}}-\mathbf{x}^\star\|_{\ell_2}&\ge\frac{\varepsilon}{\underline{\sigma}\overline{\sigma}_{\mathcal{T}^c}}-\frac{\varepsilon}{\underline{\sigma}}\sqrt{Tm-|\mathcal{T}|}\\&\ge\frac{\varepsilon}{\underline{\sigma}}\left(\frac{1}{\overline{\sigma}_{\mathcal{T}^c}}-\sqrt{Tm-|\mathcal{T}|}\right)\triangleq\alpha\end{aligned}$$

Moreover,

$$\begin{aligned}\|\mathbf{y}_T-H\hat{\mathbf{x}}\|_{\ell_1}&=\left\|U_1(\mathbf{z}^\star+\mathbf{z}_e)-P\begin{bmatrix}0\\U_{1,\mathcal{T}^c}\end{bmatrix}\mathbf{z}_e-H\hat{\mathbf{x}}\right\|_{\ell_1}\\&=\left\|U_1(\mathbf{z}^\star+\mathbf{z}_e)-P\begin{bmatrix}0\\U_{1,\mathcal{T}^c}\end{bmatrix}\mathbf{z}_e-U_1(\mathbf{z}^\star+\mathbf{z}_e-\mathbf{z}_e^\perp)\right\|_{\ell_1}\\&=\left\|U_1\mathbf{z}_e^\perp-P\begin{bmatrix}0\\U_{1,\mathcal{T}^c}\end{bmatrix}\mathbf{z}_e\right\|_{\ell_1}\le\sqrt{Tm-|\mathcal{T}|}\cdot\overline{\sigma}_{\mathcal{T}^c}\varepsilon\end{aligned}$$

So, if $\bar{\sigma}_{\mathcal{T}^c}\sqrt{Tm-|\mathcal{T}|}<1$, then $\|\mathbf{y}_T-H\hat{\mathbf{x}}\|_{\ell_1}\le\varepsilon$ ■

III. MAIN RESULT

As discussed in *Section I*, we start from modeling the uncertainty of machine learning localization result (*Support Prior*), then a pruning method is given to generate a new estimated support (*Prune Support Prior*), with precision improvement. Finally, a weighted ℓ_1 observer is given based on *Prune Support Prior*.

Assume the actual support of safe nodes is represented by $\mathcal{T}^c$, let the vector $\mathbf{q}\in\{0,1\}^{Tm}$ be an indicator of the support $\mathcal{T}$, such that

$$\mathbf{q}_i=\begin{cases}1 & \text{if } i\in\mathcal{T}^c\\0 & \text{otherwise}\end{cases} \tag{9}$$

The outputted *Support Prior* $\hat{\mathcal{T}}^c$ of any underlying machine learning localization algorithm is an estimate of $\mathcal{T}^c$, and the estimated indicator $\hat{\mathbf{q}}\in\{0,1\}^{Tm}$ is defined similarly to (9). Then an uncertainty model is defined as

$$\mathbf{q}_i=\varepsilon_i\hat{\mathbf{q}}_i+(1-\varepsilon_i)(1-\hat{\mathbf{q}}_i) \tag{10}$$

where $\varepsilon_i\sim\mathcal{B}(1,\mathbf{p}_i)$, with known $\mathbf{p}_i\in\mathbb{R}_+$, and $E[\mathbf{p}_i]=T_r$, where T_r is *true rate* obtained from the *Receiver Operator Characteristic* (ROC) of the underlying machine learning algorithm.

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

$$\mathsf{PPV}=\frac{\|\mathbf{q}\circ\hat{\mathbf{q}}\|_{\ell_0}}{\|\hat{\mathbf{q}}\|_{\ell_0}} \tag{11}$$

Lemma 3.1: ([19]) Given mutually independent Bernoulli random variables $\varepsilon_i\sim\mathcal{B}(1,\mathbf{p}_i)$, $i=1,\cdots,Tm$, then

$$\mathsf{Pr}\left\{\sum_{i=1}^{Tm}\varepsilon_i=k-1\right\}=\mathbf{r}(k),k=1,\cdots,Tm+1 \tag{12}$$

where,

$$\mathbf{r}=\beta\cdot\begin{bmatrix}-\mathbf{s}_1\\1\end{bmatrix}*\begin{bmatrix}-\mathbf{s}_2\\1\end{bmatrix}*\cdots*\begin{bmatrix}-\mathbf{s}_{Tm}\\1\end{bmatrix}$$

with $\beta=\prod_{i=1}^{Tm}\mathbf{p}_i$, $\mathbf{s}_i=-\dfrac{1-\mathbf{p}_i}{\mathbf{p}_i}$

Next, based on the knowledge of uncertainty of *Prior*, the pruning process will perform in two steps. Firstly, given a reliability level $\eta\in(0,1)$, the maximum integer $l_\eta(\le Tm)$

[1] Let $f(\mathbf{z})=\left\|U_1\mathbf{z}-P\begin{bmatrix}0\\U_{1,\mathcal{T}^c}\end{bmatrix}\mathbf{z}_e\right\|_{\ell_1}$ which is a convex function.The unique minimizer $\mathbf{z}_e^\perp$ satisfies $f(\mathbf{z}_e^\perp)\le f(0)$.

of safe nodes being correctly localized with a probability of at least η is obtained:

$$\begin{aligned} l_\eta &= \mathsf{max}\left\{k \mid \mathsf{Pr}\left\{\sum_{i\in\hat{\mathcal{T}}^c}\varepsilon_i \geq k\right\} \geq \eta\right\} \\ &= \mathsf{max}\left\{k \mid \sum_{i=1}^{k+1}\mathbf{r}_{\hat{\mathcal{T}}^c}(i) \leq 1-\eta\right\} \end{aligned} \tag{13}$$

where $\mathbf{r}_{\hat{\mathcal{T}}^c}$ is given by (12), using the index set $\hat{\mathcal{T}}^c$.

Then a *Prune Support Prior* is obtained through a robust extraction:

$$\hat{\mathcal{T}}^c_\eta = \left\{\mathsf{argsort}\downarrow(\mathbf{p}\circ\hat{\mathbf{q}})\right\}_1^{l_\eta} \tag{14}$$

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

Definition 5 (Weighted ℓ_1 Observer with Pruned Prior): Consider the CPS in (2) and the corresponding measurement model (3), given *Prune Support Prior* $\hat{\mathcal{T}}^c_\eta$ by (14), a weighted ℓ_1 observer is defined

$$\mathsf{Minimize}_{\mathbf{z}\in\mathbb{R}^n}\ \|\mathbf{y}_T - H\mathbf{z}\|_{\ell_1,\mathrm{w}}, \ \text{ with } \mathrm{w}_i = \begin{cases} 1, & i\in\hat{\mathcal{T}}^c_\eta \\ \omega, & i\in\hat{\mathcal{T}}_\eta \end{cases} \tag{15}$$

where, $0\leq\omega\leq 1$, $\|\mathbf{z}\|_{1,\mathrm{w}} \triangleq \sum_i \mathrm{w}_i|\mathbf{z}_i|$ is the weighted ℓ_1 norm.

Theorem 3.2: Consider the system model in (2) together with the corresponding measurement model (3). Assume there exists a *Support Prior* $\hat{\mathcal{T}}^c$ generated by an underlying machine learning localization algorithm with uncertainty model in (10) and

$$\sum_{i=1}^{Tm}\mathbf{p}_i > Tm\cdot p_A. \tag{16}$$

Given a positive parameter η satisfying

$$\eta \leq 1 - e\left(1 - \frac{(e-1)\sum_{i\in\hat{\mathcal{T}}^c}\mathbf{p}_i}{e|\mathcal{T}^c|}\right)^{|\mathcal{T}^c|}, \tag{17}$$

where e is Euler's number, a *Pruned Support Prior* $\hat{\mathcal{T}}^c_\eta$ given by (13) and (14). If there exists an integer $a \geq \max\{\rho-1,1\}$, where $\rho = \frac{|\hat{\mathcal{T}}_\eta|}{k}$, such that any coding matrix F ($FH=0$) satisfies RIP condition

$$(1-\delta_S)\|\mathbf{h}\|^2_{\ell_2} \leq \|F_{\mathcal{T}}\mathbf{h}\|^2_{\ell_2} \leq (1+\delta_S)\|\mathbf{h}\|^2_{\ell_2} \tag{18}$$

for all sparse vector $\mathbf{h}\in\mathcal{S}^{Tm}_S$, and RIP constant satisfies

$$\delta_{ak} - C\delta_{(a+1)k} \leq C-1 \tag{19}$$

then, with a probability of at least η, the estimation error of the Weighted ℓ_1 Observer in (15) can be upper bounded as

$$\|\hat{\mathbf{x}}-\mathbf{x}\|_{\ell_2} \leq \frac{C_1}{\underline{\sigma}_H\sqrt{k}}\left(\omega\sigma_k(\mathbf{e})_{\ell_1} + (1-\omega)\|\mathbf{e}_{\hat{\mathcal{T}}^c_\eta}\|_{\ell_1}\right), \tag{20}$$

where $\underline{\sigma}_H$ is the smallest singular value of H, and

$$C_1 = \frac{2a^{-\frac{1}{2}}(\sqrt{1-\delta_{(a+1)k}} + \sqrt{1+\delta_{ak}})}{\sqrt{1-\delta_{(a+1)k}} - \frac{1}{C}\sqrt{1+\delta_{ak}}},$$

$$C = \frac{a}{(\omega+(1-\omega)\sqrt{\rho-1})^2}.$$

Proof: Expanding (10) yields

$$\mathbf{q}_i = 2\varepsilon_i\hat{\mathbf{q}}_i + 1 - \hat{\mathbf{q}}_i - \varepsilon_i = 2\mathbf{q}_i\hat{\mathbf{q}}_i + 1 - \hat{\mathbf{q}}_i - \varepsilon_i$$

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 expectation of both sides yield

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

Using the condition in (16) yields

$$E\left[\|\mathbf{q}\circ\hat{\mathbf{q}}\|_{\ell_0} - \frac{1}{2}\|\hat{\mathbf{q}}\|_{\ell_0}\right] > 0$$

which means $\mathsf{PPV} > 1/2$, then

$$\mathsf{Pr}\{\mathbf{q}\circ\hat{\mathbf{q}} = 1 \mid \hat{\mathbf{q}} = 1\} > \mathsf{Pr}\{\mathbf{q}\circ\hat{\mathbf{q}} = 0 \mid \hat{\mathbf{q}} = 1\}$$

and with the operation in (14), it follows

$$\hat{\mathcal{T}}^c_\eta \subseteq \hat{\mathcal{T}}^c \cap \mathcal{T}^c, \ \text{ if } \ 0 < l_\eta \leq \|\mathbf{q}\circ\hat{\mathbf{q}}\|$$

From (13), it is true that

$$\mathsf{Pr}\{0 \leq l_\eta \leq \|\mathbf{q}\circ\hat{\mathbf{q}}\|\} \geq \eta, \tag{21}$$

As well known, Poisson-binomial distribution can be approximated by Binomial distribution, such that

$$\sum_{i\in\hat{\mathcal{T}}^c}\varepsilon_i \sim \mathcal{B}\left(|\mathcal{T}^c|, \frac{\sum_{i\in\hat{\mathcal{T}}^c}\mathbf{p}_i}{|\mathcal{T}^c|}\right)$$

then, from condition (17), $\eta \leq 1 - e\mathcal{M}_{\sum_{i\in\hat{\mathcal{T}}^c}\varepsilon_i}(-1)$, where $\mathcal{M}_{\sum_{i\in\hat{\mathcal{T}}^c}\varepsilon_i}(t)$ is the moment generating function of $\sum_{i\in\hat{\mathcal{T}}^c}\varepsilon_i$, thus

$$E\left[\exp\left(1 - \sum_{i\in\hat{\mathcal{T}}^c}\varepsilon_i\right)\right] \leq 1-\eta$$

Since $\exp(z)\geq 1$ for any $z\geq 0$, we obtain $E[\exp(z)] \geq \mathsf{Pr}(z\geq 0)$, then it yields

$$\mathsf{Pr}\left\{1 - \sum_{i\in\hat{\mathcal{T}}^c}\varepsilon_i \geq 0\right\} \leq 1-\eta \ \Leftrightarrow \ \mathsf{Pr}\left\{\sum_{i\in\hat{\mathcal{T}}^c}\varepsilon_i \geq 1\right\} \geq \eta$$

which means $l_\eta \neq 0$, combining with (21), we obtain $\mathsf{Pr}\{0 < l_\eta \leq \|\mathbf{q}\circ\hat{\mathbf{q}}\|\} \geq \eta$, thus, $\mathsf{Pr}\{\hat{\mathcal{T}}^c_\eta \subseteq \hat{\mathcal{T}}^c\cap\mathcal{T}^c \subseteq \mathcal{T}^c\} \geq \eta$, then

$$\mathsf{Pr}\left\{\mathsf{PPV}_\eta = 1\right\} \geq \eta \tag{22}$$

To avoid repetition, we state upfront that all claims made in this rest of proof holds with a probability of at least η. Under this probability, we have $\hat{\mathcal{T}}^c_\eta \subset \mathcal{T}^c$, $\mathcal{T} \subset \hat{\mathcal{T}}_\eta$.

Let $\hat{\mathbf{x}}$ be a minimizer to (15) with $H\hat{\mathbf{x}} = H\mathbf{x} + \mathbf{h}$, then the corresponding attack vector $\mathbf{e}^\star = \mathbf{e} + \mathbf{h}$, thus $\|\mathbf{e}^\star\|_{\ell_1,\omega} \leq \|\mathbf{e}\|_{\ell_1,\omega}$. Following the definition of weight w in (15),

$$\omega\|\mathbf{e}_{\hat{\mathcal{T}}_\eta} + \mathbf{h}_{\hat{\mathcal{T}}_\eta}\|_{\ell_1} + \|\mathbf{e}_{\hat{\mathcal{T}}^c_\eta} + \mathbf{h}_{\hat{\mathcal{T}}^c_\eta}\|_{\ell_1} \leq \omega\|\mathbf{e}_{\hat{\mathcal{T}}_\eta}\|_{\ell_1} + \|\mathbf{e}_{\hat{\mathcal{T}}^c_\eta}\|_{\ell_1}$$

Since ℓ_1 norm is decomposable for disjoint sets, such for $\mathcal{T}$ and $\mathcal{T}^c$, and $\hat{\mathcal{T}}_\eta^c \cap \mathcal{T} = \emptyset$, $\hat{\mathcal{T}}_\eta^c \cap \mathcal{T}^c = \hat{\mathcal{T}}_\eta^c$, $\mathcal{T} \cap \hat{\mathcal{T}}_\eta = \mathcal{T}$, it follows

$$\begin{aligned}\omega\|\mathbf{e}_{\mathcal{T}} + \mathbf{h}_{\mathcal{T}}\|_{\ell_1} + \omega\|\mathbf{e}_{\hat{\mathcal{T}}_\eta \cap \mathcal{T}^c} + \mathbf{h}_{\hat{\mathcal{T}}_\eta \cap \mathcal{T}^c}\|_{\ell_1} + \|\mathbf{e}_{\hat{\mathcal{T}}_\eta^c} + \mathbf{h}_{\hat{\mathcal{T}}_\eta^c}\|_{\ell_1} \\ \leq \omega\|\mathbf{e}_{\mathcal{T}}\|_{\ell_1} + \omega\|\mathbf{e}_{\hat{\mathcal{T}}_\eta \cap \mathcal{T}^c}\|_{\ell_1} + \|\mathbf{e}_{\hat{\mathcal{T}}_\eta^c}\|_{\ell_1}\end{aligned}$$

By triangle inequality, it follows

$$\omega\|\mathbf{h}_{\hat{\mathcal{T}}_\eta \cap \mathcal{T}^c}\|_{\ell_1} + \|\mathbf{h}_{\hat{\mathcal{T}}_\eta^c}\|_{\ell_1} \leq \omega\|\mathbf{h}_{\mathcal{T}}\|_{\ell_1} + 2\left(\|\mathbf{e}_{\hat{\mathcal{T}}_\eta^c}\|_{\ell_1} + \omega\|\mathbf{e}_{\hat{\mathcal{T}}_\eta \cap \mathcal{T}^c}\|_{\ell_1}\right)$$

Adding and subtracting $\omega\|\mathbf{h}_{\hat{\mathcal{T}}_\eta^c \cap \mathcal{T}^c}\|_{\ell_1} (= \omega\|\mathbf{h}_{\hat{\mathcal{T}}_\eta^c}\|_{\ell_1})$ on LHS, $2\omega\|\mathbf{e}_{\hat{\mathcal{T}}_\eta^c \cap \mathcal{T}^c}\|_{\ell_1} (= 2\omega\|\mathbf{e}_{\hat{\mathcal{T}}_\eta^c}\|_{\ell_1})$ on RHS, it yields

$$\begin{aligned}\omega\|\mathbf{h}_{\mathcal{T}^c}\|_{\ell_1} + (1-\omega)\|\mathbf{h}_{\hat{\mathcal{T}}_\eta^c}\|_{\ell_1} \leq \omega\|\mathbf{h}_{\mathcal{T}}\|_{\ell_1} + \\ 2\left(\omega\|\mathbf{e}_{\mathcal{T}^c}\|_{\ell_1} + (1-\omega)\|\mathbf{e}_{\hat{\mathcal{T}}_\eta^c}\|_{\ell_1}\right)\end{aligned}$$

And since $\|\mathbf{h}_{\mathcal{T}^c}\|_{\ell_1} = \omega\|\mathbf{h}_{\mathcal{T}^c}\|_{\ell_1} + (1-\omega)\|\mathbf{h}_{\hat{\mathcal{T}}_\eta^c}\|_{\ell_1} + (1-\omega)\|\mathbf{h}_{\hat{\mathcal{T}}_\eta \cap \mathcal{T}^c}\|_{\ell_1}$, it yields

$$\begin{aligned}\|\mathbf{h}_{\mathcal{T}^c}\|_{\ell_1} \leq \omega\|\mathbf{h}_{\mathcal{T}}\|_{\ell_1} + (1-\omega)\|\mathbf{h}_{\hat{\mathcal{T}}_\eta \cap \mathcal{T}^c}\|_{\ell_1} + \\ 2\left(\omega\|\mathbf{e}_{\mathcal{T}^c}\|_{\ell_1} + (1-\omega)\|\mathbf{e}_{\hat{\mathcal{T}}_\eta^c}\|_{\ell_1}\right)\end{aligned} \tag{23}$$

Next, sort the coefficients of $\mathbf{h}_{\mathcal{T}^c}$ in descending order, and let $\mathcal{T}_j, j \in \{1, 2, \cdots\}$ denote jth support in $\mathbf{h}_{\mathcal{T}^c}$ with size $ak \in \mathbb{Z}$, where $a > 1$. Since $\|\mathbf{h}_{\mathcal{T}_{j-1}}\|_{\ell_1} \geq ak \cdot \min_{i \in \mathcal{T}_{j-1}} (\mathbf{h}_{\mathcal{T}_{j-1}})_i \geq ak\|\mathbf{h}_{\mathcal{T}_j}\|_{\ell_\infty}$, let $\mathcal{T}_0 = \mathcal{T} \cup \mathcal{T}_1$, we have

$$\|\mathbf{h}_{\mathcal{T}_0^c}\|_{\ell_2} \leq \sum_{j \geq 2} \|\mathbf{h}_{\mathcal{T}_j}\|_{\ell_\infty} \leq ak^{-1/2} \sum_{j \geq 1} \|\mathbf{h}_{\mathcal{T}_j}\|_{\ell_1} = ak^{-1/2}\|\mathbf{h}_{\mathcal{T}^c}\|_{\ell_1} \tag{24}$$

Note that $|\hat{\mathcal{T}}_\eta \cap \mathcal{T}^c| = (\rho - 1)k$ since $\mathsf{PPV} = 1$, and $|\mathcal{T}| = k$, then $\|\mathbf{h}_{\hat{\mathcal{T}}_\eta \cap \mathcal{T}^c}\|_{\ell_1} \leq \sqrt{(\rho-1)k}\|\mathbf{h}_{\hat{\mathcal{T}}_\eta \cap \mathcal{T}^c}\|_{\ell_2}$, $\|\mathbf{h}_{\mathcal{T}}\|_{\ell_1} \leq \sqrt{k}\|\mathbf{h}_{\mathcal{T}}\|_{\ell_2} \leq \sqrt{k}\|\mathbf{h}_{\mathcal{T}_0}\|_{\ell_2}$. Since $a \geq \rho - 1$, we obtain $|\hat{\mathcal{T}}_\eta \cap \mathcal{T}^c| = (\rho-1)k \leq ak = |\mathcal{T}_1|$, then $\|\mathbf{h}_{\hat{\mathcal{T}}_\eta \cap \mathcal{T}^c}\|_{\ell_2} = \|\mathbf{h}_{\mathcal{T} \cup (\hat{\mathcal{T}}_\eta \cap \mathcal{T}^c)}\|_{\ell_2} \leq \|\mathbf{h}_{\mathcal{T}_0}\|_{\ell_2}$. Then combine with (23) and (24),

$$\begin{aligned}\|\mathbf{h}_{\mathcal{T}_0^c}\|_{\ell_2} \leq & \frac{\omega + (1-\omega)\sqrt{\rho - 1}}{\sqrt{a}}\|\mathbf{h}_{\mathcal{T}_0}\|_{\ell_2} \\ & + \frac{2}{\sqrt{ak}}\left(\omega\|\mathbf{e}_{\mathcal{T}^c}\|_{\ell_1} + (1-\omega)\|\mathbf{e}_{\hat{\mathcal{T}}_\eta^c}\|_{\ell_1}\right)\end{aligned} \tag{25}$$

Since $\|F\mathbf{h}\|_{\ell_2} = \|(F\mathbf{e}^\star - F\mathbf{y}_T) - (F\mathbf{e} - F\mathbf{y}_T)\|_{\ell_2} = 0$, it follows, based on triangle inequality and RIP condition,

$$\begin{aligned}\sqrt{1 - \delta_{(a+1)k}}\|\mathbf{h}_{\mathcal{T}_0}\|_{\ell_2} \leq \|F\mathbf{h}_{\mathcal{T}_0}\|_{\ell_2} \leq \|F\mathbf{h}_{\mathcal{T}_0^c}\|_{\ell_2} \leq \\ \sum_{j \geq 2} \|F\mathbf{h}_{\mathcal{T}_j}\|_{\ell_2} \leq \sqrt{1 + \delta_{ak}} \sum_{j \geq 2} \|\mathbf{h}_{\mathcal{T}_j}\|_{\ell_2} \leq \frac{\sqrt{1 + \delta_{ak}}}{\sqrt{ak}}\|\mathbf{h}_{\mathcal{T}^c}\|_{\ell_1}\end{aligned}$$

Combining with (23), it yields

$$\begin{aligned}\sqrt{1 - \delta_{(a+1)k}}\|\mathbf{h}_{\mathcal{T}_0}\|_{\ell_2} \leq \omega\frac{\sqrt{1 + \delta_{ak}}}{\sqrt{ak}}\|\mathbf{h}_{\mathcal{T}}\|_{\ell_1} + \\ (1-\omega)\frac{\sqrt{1 + \delta_{ak}}}{\sqrt{ak}}\|\mathbf{h}_{\hat{\mathcal{T}}_\eta \cap \mathcal{T}^c}\|_{\ell_1} \\ +2\frac{\sqrt{1 + \delta_{ak}}}{\sqrt{ak}}\left(\omega\|\mathbf{e}_{\mathcal{T}^c}\|_{\ell_1} + (1-\omega)\|\mathbf{e}_{\hat{\mathcal{T}}_\eta^c}\|_{\ell_1}\right)\end{aligned}$$

Notice, we have $\|\mathbf{h}_{\mathcal{T}}\|_{\ell_1} \leq \sqrt{k}\|\mathbf{h}_{\mathcal{T}_0}\|_{\ell_2}$ and $\|\mathbf{h}_{\hat{\mathcal{T}}_\eta \cap \mathcal{T}^c}\|_{\ell_1} \leq \sqrt{(\rho-1)k}\|\mathbf{h}_{\mathcal{T}_0}\|_{\ell_2}$. Combining with the upper inequality,

$$\|\mathbf{h}_{\mathcal{T}_0}\|_{\ell_2} \leq \frac{2\frac{\sqrt{1+\delta_{ak}}}{\sqrt{ak}}\left(\omega\|\mathbf{e}_{\mathcal{T}^c}\|_{\ell_1} + (1-\omega)\|\mathbf{e}_{\hat{\mathcal{T}}_\eta^c}\|_{\ell_1}\right)}{\sqrt{1 - \delta_{(a+1)k}} - \frac{\omega+(1-\omega)\sqrt{\rho-1}}{\sqrt{a}}\sqrt{1 + \delta_{ak}}}$$

Since $\|\mathbf{h}\|_{\ell_2} \leq \|\mathbf{h}_{\mathcal{T}_0}\|_{\ell_2} + \|\mathbf{h}_{\mathcal{T}_0^c}\|_{\ell_2}$, combine the above inequality with (25), it yields

$$\|\mathbf{h}\|_{\ell_2} \leq \frac{2\frac{\sqrt{1-\delta_{(a+1)k}}+\sqrt{1+\delta+ak}}{\sqrt{ak}}\left(\omega\|\mathbf{e}_{\mathcal{T}^c}\|_{\ell_1} + (1-\omega)\|\mathbf{e}_{\hat{\mathcal{T}}_\eta^c}\|_{\ell_1}\right)}{\sqrt{1 - \delta_{(a+1)k}} - \frac{\omega+(1-\omega)\sqrt{\rho-1}}{\sqrt{a}}\sqrt{1 + \delta_{ak}}}$$

where $\|\mathbf{e}_{\mathcal{T}^c}\|_{\ell_1} = \sigma_k(\mathbf{e})_{\ell_1}$ is best k sparse approximation error of $\mathbf{e}$ defined in (1), and the estimation error $\|\hat{\mathbf{x}} - \mathbf{x}\|_{\ell_2} \leq \underline{\sigma}_H^{-1}\|\mathbf{h}\|_{\ell_2}$, and $\underline{\sigma}_H$ is the smallest singular value of H. Moreover, to maintain the denominator to be positive, a condition is given

$$\delta_{ak} + C\delta_{(a+1)k} < C - 1$$

where $C = \frac{a}{(\omega+(1-\omega)\sqrt{\rho-1})^2}$. ■

Remark 3: The condition in (16) guarantees the underlying machine learning algorithm works better than random flip of a fair coin, which is a reasonable assumption.

Remark 4: A precision improvement conclusion (22) is given in proof, and notice that the upper bound restriction of η is an overwhelming probability.

Remark 5: To avoid the sacrifice of redundancy of measurements, we set weight ω close to zero, but not zero, even if we are confident in the precision of $\hat{\mathcal{T}}_\eta^c$.

IV. SIMULATION

In this section, a numerical simulation[2] and an application simulation on IEEE-14 bus system are presented.

A. Numerical Simulation

In this simulation, we compare the resiliency of three estimation scheme: ℓ_1 observer without prior, weighted ℓ_1 observer with prior generated an underlying attack localization algorithm with a true rate of $T_r = 0.6$, and weighted ℓ_1 observer with the pruned prior. The system dimension is set as $m = 20, n = 10$, and then a full observable system is generated with random pair (A, C) of independent Gaussian entries [13]. For different attack percentage P_A, the FDIA is designed by (8) on random support $\mathcal{T}$. By defining ”success” as that the estimation error is less than 0.1% of the real state, the success percentage is calculated from 1000 trials. In Figure 1, a performance comparison is presented for varying attack percentages. As proved in literature [12], ℓ_1 *observer without prior* cannot work when attack percentage is larger than $1/2$. The prior information can improve the resiliency of the estimator, but the improvement is limited because the precision of the prior information is not enough. By including pruning method, the resiliency is improved a lot.

[2] we make the numerical simulation be open source at `https://github.com/ZYblend/Resilient-Pruning-Observer`

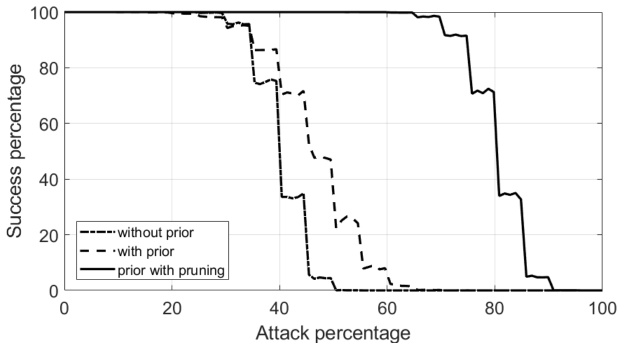

Fig. 1. A comparison of estimation performance under different attack percentage $P_A \in [0,1]$ between ℓ_1 *Observer without Prior*, *Weighted* ℓ_1 *Observer with Prior*, and *Weighted* ℓ_1 *Observer with Prior Pruning*.

B. Application Example

In this subsection, an additional simulation is carried out for a more realistic IEEE Bus 14 system with a similar setup in [15]. Here, 30% of the measurement nodes are attacked. A comparison of the resulting phase angles estimation for an ℓ_1 *observer without prior* and the proposed *weighted* ℓ_1 *observer with prior pruning* is shown in Figure 2. Moreover, two comparison metrics for the estimation errors are given in Table I. The first one is the *root-mean-square* (RMS) value of error, the second one is the *maximum absolute value* of the errors. Since the attack percentage is small, ℓ_1 *observer without prior* works well. However, there are still notable spikes that may induce closed loop instability. As shown in the Figure, the proposed *weighted* ℓ_1 *observer with prior pruning* completely eliminates the spikes!

TABLE I
ERROR METRIC VALUES

	RMS Metric			Max. Ans. Metric		
	LO	**L1O**	**WL1P**	**LO**	**L1O**	**WL1P**
δ_1	2.5359	0.0002	0.00004	5.7480	0.0034	0.0005
δ_2	2.3917	0.0002	0.0001	5.7000	0.0016	0.0016
δ_3	2.6353	0.0012	0.0001	7.5232	0.0215	0.0013
δ_4	2.3685	0.0006	0.0005	5.7236	0.0063	0.0052
δ_5	2.6638	0.0007	0.0001	7.8757	0.0085	0.0016

LO: Luenberger observer, **L1O**: ℓ_1 observer without prior
WL1P: Weighted ℓ_1 observer with prior pruning

V. CONCLUSIONS

This paper proposed a weighted ℓ_1 observer with prior pruning scheme against FDIAs. The pruning method gives a method to improve localization precision of any underlying localization algorithm without training effort. Moreover, the weighted ℓ_1 observer with prior pruning is capable of coping with high-percentage of attacks among measurement nodes, which relaxes the transitional restriction on the maximum attack percentage for resilient ℓ_1 observer, thereby improve the resiliency of systems.

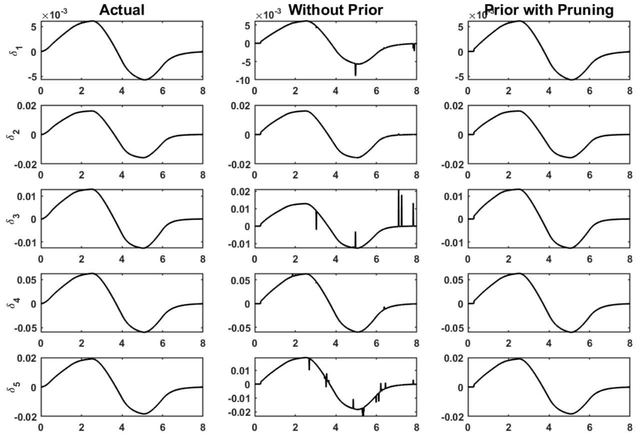

Fig. 2. Angle estimation for IEEE-14 bus system under FDIA, ”without Prior” means ℓ_1 *Observer without Prior*, ”Prior with Pruning” means *Weighted* ℓ_1 *Observer with Prior Pruning*.

REFERENCES

[1] E. A. Lee and S. A. Seshia, *Introduction to embedded systems: A cyber-physical systems approach*. Mit Press, 2016.

[2] R. Rajkumar, I. Lee, L. Sha, and J. Stankovic, “Cyber-physical systems: The next computing revolution,” in *Design automation conference*, IEEE, 2010, pp. 731–736.

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

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

[5] O. M. Anubi, L. Mestha, and H. Achanta, “Robust resilient signal reconstruction under adversarial attacks,” *arXiv preprint arXiv:1807.08004*, 2018.

[6] S. Weerakkody, O. Ozel, Y. Mo, B. Sinopoli, *et al.*, “Resilient control in cyber-physical systems: Countering uncertainty, constraints, and adversarial behavior,” *Foundations and Trends® in Systems and Control*, vol. 7, no. 1-2, pp. 1–252, 2019.

[7] O. M. Anubi and C. Konstantinou, “Enhanced resilient state estimation using data-driven auxiliary models,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 1, pp. 639–647, 2019.

[8] M. Ozay, I. Esnaola, F. T. Y. Vural, S. R. Kulkarni, and H. V. Poor, “Machine learning methods for attack detection in the smart grid,” *IEEE transactions on neural networks and learning systems*, vol. 27, no. 8, pp. 1773–1786, 2015.

[9] H. Sedghi and E. Jonckheere, “Statistical structure learning to ensure data integrity in smart grid,” *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1924–1933, 2015.

[10] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley, “Least squares generative adversarial networks,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2794–2802.

[11] O. M. Anubi, C. Konstantinou, and R. Roberts, “Resilient optimal estimation using measurement prior,” *arXiv preprint arXiv:1907.13102*, 2019.

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

[13] E. J. Candes and T. Tao, “Decoding by linear programming,” *IEEE transactions on information theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[14] E. Candes, J. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” 2005.

[15] O. M. Anubi, C. Konstantinou, C. A. Wong, and S. Vedula, “Multi-model resilient observer under false data injection attacks,” in *2020 IEEE Conference on Control Technology and Applications (CCTA)*, IEEE, 2020, pp. 1–8.

[16] T. Shinohara, T. Namerikawa, and Z. Qu, “Resilient reinforcement in secure state estimation against sensor attacks with a priori information,” *IEEE Transactions on Automatic Control*, vol. 64, no. 12, pp. 5024–5038, 2019.

[17] M. P. Friedlander, H. Mansour, R. Saab, and Ö. Yilmaz, “Recovering compressively sampled signals using partial support information,” *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1122–1134, 2011.

[18] Y. Zheng and O. M. Anubi, “Attack-resilient observer pruning for path-tracking control of wheeled mobile robot,” in *2020 ASME Dynamic Systems and Control (DSC) Conference*, ASME, 2020, pp. 1–9.

[19] ——, “Resilient pruning observer design for cyber-physical systems under false data injection attacks,” 2020.