

Robust Resilient Signal Reconstruction under Adversarial Attacks

Olugbenga Moses Anubi¹, *Member, IEEE*, Lalit Mestha², *Fellow, IEEE*, Hema Achanta¹, *Member, IEEE*

¹GE Global Research, Niskayuna, NY 12309 USA

²KinetiCor Inc., Honolulu, HI 96816 USA

We consider the problem of signal reconstruction for a system under sparse unbounded signal corruption by an adversarial agent. The reconstruction problem follows the standard error coding problem that has been studied extensively in literature, with the added consideration of support estimation of the attack vector. The problem is formulated as a constrained optimization problem – merging exciting developments in the field of machine learning and estimation theory. Sufficient conditions for the *reconstructability* and the associated reconstruction error bounds were obtained for both exact and inexact support estimation of the attack vector. Special cases of data-driven model and linear dynamical systems were also considered.

Index Terms—Compressive Sensing, Signal Reconstruction, Cyber-physical Systems, Secure Estimation, Resilient.

I. INTRODUCTION

MAJORITY of Industrial systems and critical infrastructures are Cyber-physical Systems (CPS), in that they consist of an interplay between physical components (sensors, controllers and actuators) and digital components (computational algorithms, software systems, human-machine interfaces) via communication networks. This opens up a portal that makes them prime targets of cyber malicious activities. The resilient signal reconstruction is a filtering problem for removing undesirable effects created by malicious intent as in adversarial attacks or other similar unbounded phenomenon on some of the system’s monitoring nodes. For physical systems, if signal reconstruction is performed jointly on all the affected signals, then it can improve resiliency of critical infrastructure to cyber-activities and fault-induced anomalies to allow continued, safe operation [1].

There are numerous work in literature on the secure estimation for CPS [2]–[11]. However, we focus only on the ones which are optimization based - since that is the approach taken in this work. Moreover, due to sparsity assumption on the set of attacked nodes, majority of these works are based on the classical error correction problem [12]. Given a coding matrix $F \in \mathbb{R}^{n \times m}$ with far fewer rows than columns ($n \ll m$) and a vector of observed/measured quantities $\mathbf{y} \in \mathbb{R}^m$, the coding problem is to recover a sparse vector \mathbf{e} , $\|\mathbf{e}\|_{l_0} < m$ for which $\mathbf{y} = F\mathbf{e}$. This is cast as the compressive sensing problem:

$$\text{Minimize: } \|\mathbf{e}\|_{l_0} \quad \text{Subject to: } \mathbf{y} = F\mathbf{e}. \quad (1)$$

Hayden et. al [13] obtained a sufficient condition that if all subsets of $2q$ columns of F are full rank, then any error $\|\mathbf{e}\|_{l_0} \leq q$ can be reconstructed uniquely by the solution of the optimization problem in (1).

Although in some cases [6] the optimization problem in (1) is solved as is, in most cases, it does not lend itself to

a solution in polynomial time due to its nonconvexity. As a result, it is often replaced with its convex neighbor:

$$\text{Minimize: } \|\mathbf{e}\|_{l_1} \quad \text{Subject to: } \mathbf{y} = F\mathbf{e}. \quad (2)$$

The two programs, however, have been shown to be equivalent under the condition that the *Restricted Isometric Property (RIP)* holds [14]–[17]. Let $F^{\mathcal{T}} \triangleq ((F^{\top})_{\mathcal{T}})^{\top} \in \mathbb{R}^{n \times |\mathcal{T}|}$, $\mathcal{T} \triangleq \text{supp}(\mathbf{e}) \subset \{1, \dots, m\}$ be the submatrix obtained by extracting the columns of F corresponding to the indices in \mathcal{T} . Then the *S-restricted isometry constant* δ_S of F is defined as the smallest quantity such that

$$(1 - \delta_S) \|\mathbf{v}\|^2 \leq \|F^{\mathcal{T}_S} \mathbf{v}\|^2 \leq (1 + \delta_S) \|\mathbf{v}\|^2$$

for all subsets \mathcal{T}_S with $|\mathcal{T}_S| \leq S$ and vector $\mathbf{v} \in \mathbb{R}^{|\mathcal{T}_S|}$. This property essentially requires that every set of columns with cardinality less than S approximately behaves like an orthonormal system. Moreover, it was shown that if

$$\delta_S + \delta_{2S} + \delta_{3S} < 1,$$

then solving the optimization problem in (2) recovers any sparse signal \mathbf{e} for which $|\mathcal{T}| \leq S$.

Now, suppose there exists an *oracle* which estimates the support \mathcal{T} in advance, then the sparse vector \mathbf{e} can be estimated to an accuracy of $\frac{\varepsilon}{1 - \delta_{|\mathcal{T}|}}$ by the least square estimator:

$$\hat{\mathbf{e}} = \arg \min_{\mathbf{e}} \left\{ \|\mathbf{y}_{\mathcal{T}} - F_{\mathcal{T}} \mathbf{e}\|^2 \right\},$$

where ε is the model-measurement error. Of course, if the measurement is error-free, then the estimation is exact. The goal of this work is to investigate least-square-type estimator when such oracle is available subject to both oracle and measurement uncertainties. Such oracles are termed *localization oracles* for the purpose of this work. The motivation for this approach stemmed from the authors’ experience from working on the DOE funded program *Cyber Attack Detection and Accommodation for Energy Delivery systems*

where a team of Machine Learning experts have developed such algorithm using supporting data from other sources. The question then arises “What is the best simplest thing to do to reconstruct true signals given localization information with

*Research supported by US DOE’s (Department of Energy) Cyber Security for Energy Delivery Systems (CEDSS) R&D Program and GE Global Research
Corresponding author: O. Anubi (email: anubimoses@gmail.com).

The work presented in this paper was done while all authors were at the GE Global Research

uncertainty?”. This work is an attempt to provide an answer partially to that question.

The rest of the paper is organized as follows: In Section III, the measurement model considered is presented with the basic reconstruction problem. In Section IV, some achievable bounds are proved for the reconstruction problem where the exact support of the attack vector is assumed to be provided apriori by some support estimator. In Section V, the exact support knowledge assumption is replaced by an uncertainty model derived from the ROC statistic of the support estimator and new results are obtained based on the statistical information. Special cases of data-driven model and linear dynamical systems were considered in Section VI. Finally, some future directions are highlighted in Section VII

II. NOTATION

The following notions and conventions are employed throughout the 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. X^\top denotes the transpose of the quantity X . By $Q \succeq 0$, it is meant that Q is a positive semi-definite symmetric matrix, i.e $\mathbf{x}^\top Q \mathbf{x} \geq 0 \forall \mathbf{x} \neq 0$ and $Q \succ 0$ denotes positive definiteness which is defined with strict $>$ instead. Given $Q \succ 0$, the Q -weighted norm is defined as $\|\mathbf{x}\|_Q \triangleq \mathbf{x}^\top Q \mathbf{x}$. Normal-face lower-case letters ($x \in \mathbb{R}$) are used to represent real scalars, bold-face lower-case letter ($\mathbf{x} \in \mathbb{R}^n$) represents vectors, while normal-face upper case ($X \in \mathbb{R}^{n \times m}$) represents matrices. Let $\mathcal{T} \subseteq \{1, \dots, n\}$ then, for a matrix $X \in \mathbb{R}^{n \times m}$, $X_{\mathcal{T}} \in \mathbb{R}^{|\mathcal{T}| \times m}$ is the submatrix obtained by extracting the rows of X corresponding to the indices in \mathcal{T} . For a vector \mathbf{x} , \mathbf{x}_i denotes its i th element. The symbol \circ denotes element-wise multiplication of two vectors and is defined as $\mathbf{z} = \mathbf{x} \circ \mathbf{y}$, where $z_i = \mathbf{x}_i \mathbf{y}_i$. $\text{supp}(\mathbf{x})$ is the *support* of the vector \mathbf{x} given by the set $\text{supp}(\mathbf{x}) = \{i | \mathbf{x}_i \neq 0\}$. $\text{argsort} \downarrow (\mathbf{x})$ denotes a function that returns the sorted indices of vector \mathbf{x} in descending order (i.e $\text{argsort} \downarrow (\mathbf{x}) = \{i | \mathbf{x}_i \geq \mathbf{x}_{i+1}\}$). S^c denotes the complement of a set and the universal set on which it is defined will be clear from the context.

III. PRELIMINARIES

Consider the linear model:

$$\mathbf{y} = C\mathbf{x} + \mathbf{e} + \mathbf{v}, \quad (3)$$

where $\mathbf{y}, \mathbf{e}, \mathbf{v} \in \mathbb{R}^m$ are vectors of observation/measurements, attack/corruption due to an adversarial agent, and error term due to measurement noise/model uncertainty respectively. The matrix $C \in \mathbb{R}^{m \times n}$ a mapping from some internal state ($\subseteq \mathbb{R}^n$) to the output space ($\subseteq \mathbb{R}^m$). The following assumptions are made with respect to the model above:

Assumptions

- Redundancy: Measurements contain redundant information in that $m > n$
- Bounded Noise: There exists a known $\varepsilon > 0$ such that $\|\mathbf{v}\| \leq \varepsilon$
- Sparse Corruption: $|\text{supp}(\mathbf{e})| \ll m$
- Attack-Noise Orthogonality: WLOG, $\mathbf{e}^\top \mathbf{v} = 0$

Consequently, the reconstruction problem is given by:

$$\begin{aligned} & \text{Minimize: } \|\mathbf{e}\|_{l_0} + \|\mathbf{v}\|_{l_2} \\ & \text{Subject to:} \\ & \quad \mathbf{y} = C\mathbf{x} + \mathbf{e} + \mathbf{v} \\ & \quad \mathbf{e}^\top \mathbf{v} = 0 \end{aligned} \quad (4)$$

which is, in general, a very challenging problem to solve due to the index minimization objective and the degeneracy introduced by the complementarity constraint $\mathbf{e}^\top \mathbf{v} = 0$. However, if there exists a localization oracle that provides the support $\mathcal{T} = \text{supp}(\mathbf{e})$ apriori, then the reconstruction problem reduces to the unconstrained problem:

$$\text{Minimize: } \|\mathbf{y}_{\mathcal{T}^c} - C_{\mathcal{T}^c} \mathbf{x}\|_{l_2}. \quad (5)$$

Of course, there are obvious conditions under which the solution to the above optimization problem provides no guarantee of reconstructing the actual signal. In what follows, the reconstruction error bounds are studied in more details under different conditions.

IV. RECONSTRUCTION WITH EXACT SUPPORT KNOWLEDGE

In this section, we examine some bounds on the reconstruction error when the attack support is known exactly. Although, the exact knowledge assumption is not pragmatic, it does give us a lower bound and a benchmark for the cases where the support is not known exactly. The following theorem examines the performance of a least-square reconstruction from partial information.

Theorem 1 (Least Square Reconstruction). *Given the linear model*

$$\mathbf{y} = C\mathbf{x} + \boldsymbol{\nu}, \quad (6)$$

where $\mathbf{y} \in \mathbb{R}^m$ is a vector of measurements, $\mathbf{x} \in \mathbb{R}^n, n \leq m$ is a vector of internal states (or features), $C \in \mathbb{R}^{m \times n}$, and $\boldsymbol{\nu}$ is the model error with the associated error bound $\|\boldsymbol{\nu}\| \leq \varepsilon$ for a known constant $\varepsilon > 0$.

Consider any partial measurement $\mathbf{y}_{\mathcal{T}^c} \in \mathbb{R}^{m_1}, m_1 < n$ satisfying

$$\mathbf{y}_{\mathcal{T}^c} = C_{\mathcal{T}^c} \mathbf{x}^* + \boldsymbol{\nu}_{\mathcal{T}^c}, \quad (7)$$

where $\boldsymbol{\nu}_{\mathcal{T}^c}$ is the associated model error and the vector $\mathbf{x}^* \in \mathbb{R}^n$ is the unknown actual internal state associated with the complete measurement set as in (6).

The least-square estimator;

$$\hat{\mathbf{x}} = \arg \min \left\{ \frac{1}{2} \|\mathbf{y}_{\mathcal{T}^c} - C_{\mathcal{T}^c} \mathbf{x}\|^2 \right\}, \quad (8)$$

of \mathbf{x}^* , satisfies the error bound

$$\|\hat{\mathbf{x}} - \mathbf{x}^*\| \leq \frac{2}{\sigma} \varepsilon, \quad (9)$$

where σ is the smallest singular value of $C_{\mathcal{T}^c}$.

Proof. By the optimality of $\hat{\mathbf{x}}$, it follows that

$$\|\mathbf{y}_{\mathcal{T}^c} - C_{\mathcal{T}^c}\hat{\mathbf{x}}\|^2 \leq \|\mathbf{y}_{\mathcal{T}^c} - C_{\mathcal{T}^c}\mathbf{x}^*\|^2. \quad (10)$$

After using (7), the above inequality can be simplified and expanded as follows:

$$\begin{aligned} & \|C_{\mathcal{T}^c}\mathbf{x}^* - C_{\mathcal{T}^c}\hat{\mathbf{x}} + \boldsymbol{\nu}_{\mathcal{T}^c}\|^2 \leq \|\boldsymbol{\nu}_{\mathcal{T}^c}\|^2 \\ \Rightarrow & \|C_{\mathcal{T}^c}\tilde{\mathbf{x}}\|^2 \leq 2\boldsymbol{\nu}_{\mathcal{T}^c}^\top (C_{\mathcal{T}^c}\tilde{\mathbf{x}}), \end{aligned} \quad (11)$$

where $\tilde{\mathbf{x}} = \hat{\mathbf{x}} - \mathbf{x}^*$. After using Young's Inequality, for some $\delta > 1$ the above inequality yields¹,

$$\begin{aligned} & \|C_{\mathcal{T}^c}\tilde{\mathbf{x}}\|^2 \leq \delta \|\boldsymbol{\nu}_{\mathcal{T}^c}\|^2 + \frac{1}{\delta} \|C_{\mathcal{T}^c}\tilde{\mathbf{x}}\|^2, \\ \Rightarrow & \left(1 - \frac{1}{\delta}\right) \|C_{\mathcal{T}^c}\tilde{\mathbf{x}}\|^2 \leq \delta \|\boldsymbol{\nu}_{\mathcal{T}^c}\|^2 \\ \Rightarrow & \|C_{\mathcal{T}^c}\tilde{\mathbf{x}}\|^2 \leq 4\|\boldsymbol{\nu}_{\mathcal{T}^c}\|^2 \leq \frac{\delta^2}{\delta - 1} \|\boldsymbol{\nu}_{\mathcal{T}^c}\|^2. \end{aligned} \quad (12)$$

From which we conclude that

$$\|\tilde{\mathbf{x}}\| \leq \frac{2}{\sigma} \|\boldsymbol{\nu}\| \leq \frac{2}{\sigma} \varepsilon \quad (13)$$

□

Remark 1 (Rank-deficiency and RIP). *Necessarily* $|\mathcal{T}^c| \geq n$, otherwise the reconstruction error $\|\hat{\mathbf{x}} - \mathbf{x}^*\|$ is unbounded. Consequently, one can conclude that: $\|\hat{\mathbf{x}} - \mathbf{x}^*\| \leq \frac{2}{1-\delta_n} \varepsilon$, where δ_n is the n -restricted isometry constant of C^\top .

Although it was shown in [12] that random matrices satisfy the RIP condition with overwhelming probability, certifying such property is still an NP-hard problem [18]. In order to guarantee bounded reconstruction error for the cases where there are potential loss of row-rank after selection due to the localization oracle, we investigate the use of a special constraint in the reconstruction optimization problem.

Corollary 1 (Constrained Least Square Reconstruction). *Let $\mathcal{X} \subset \mathbb{R}^n$ be a set characterized by $\|\mathbf{x}_1 - \mathbf{x}_2\| \leq \delta$ for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$ and some $\delta > 0$. Consider the constrained least-square estimator:*

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathcal{X}} \left\{ \frac{1}{2} \|\mathbf{y}_{\mathcal{T}^c} - C_{\mathcal{T}^c}\mathbf{x}\|^2 \right\}. \quad (14)$$

If $\mathbf{x}^* \in \mathcal{X}$, then the reconstruction error can be upper bounded as:

$$\|\hat{\mathbf{x}} - \mathbf{x}^*\| \leq 2 \min \left\{ \frac{\delta}{2}, \frac{\varepsilon}{1-\delta_n} \right\}. \quad (15)$$

Proof. Using the optimality of $\hat{\mathbf{x}}$ and following similar argument as in the proof of Theorem 1, it is shown that $\|\mathbf{x} - \mathbf{x}^*\| \leq 2\frac{\varepsilon}{1-\delta_n}$. Next, using the feasibility of both \mathbf{x} and \mathbf{x}^* , it follows that $\|\mathbf{x} - \mathbf{x}^*\| \leq \delta$. Thus, $\|\hat{\mathbf{x}} - \mathbf{x}^*\| \leq \min \left\{ \delta, 2\frac{\varepsilon}{1-\delta_n} \right\}$. □

Remark 2. *Although Corollary 1 provides a guaranteed bound on the reconstruction error, it introduces another challenge of finding the set \mathcal{X} that contains the unknown vector \mathbf{x}^* . Fortunately, there is a host of supervised and*

unsupervised machine learning models and algorithms that can be used to find such set from historical data, together with some exogenous supporting measurement. In such cases, the bound is guaranteed with a probability depending on the ROC statistics of the underlying machine learning model. Interested readers are directed to the reference [1] where the authors used supervised learning with support vector machines and generated a local approximation for \mathcal{X} via a quadratic approximation of the boundary score function.

V. RECONSTRUCTION WITH INEXACT SUPPORT KNOWLEDGE

Even though, the constrained least square reconstruction described in the previous section provides guaranteed bound on the reconstruction error, it is impractical due to the exact knowledge assumption on the support estimation. Exact support knowledge alludes to a perfect localization oracle which is not possible, or extremely challenging at best, from a practical standpoint. Thus, it is imperative to understand the effect of the imperfection of the localization oracle on the reconstruction error bound. The goal of this section is to re-examine the constrained least square reconstruction with uncertain localization information.

For the unknown support $\mathcal{T} = \text{supp}(\mathbf{e})$, let the vector \mathbf{q} be an indicator of \mathcal{T}^c and defined element-wise as:

$$\mathbf{q}_i = \begin{cases} 0 & \text{if } i \in \mathcal{T} \\ 1 & \text{otherwise} \end{cases} \quad (16)$$

Suppose the localization oracle gives an estimated support $\hat{\mathcal{T}}$, with $\hat{\mathbf{q}}$ similarly defined, the following uncertainty model is used:

$$\mathbf{q}_i = \epsilon_i \hat{\mathbf{q}}_i + (1 - \epsilon_i)(1 - \hat{\mathbf{q}}_i) \quad (17)$$

where ϵ_i is a Bernoulli random variable with mean \mathbf{p}_i whose estimate is given by the *true positive* rate from the localization ROC statistic. Each $\epsilon_i \sim B(1, \mathbf{p}_i)$ captures the agreement between the estimated and actual support as

$$\epsilon_i = \begin{cases} 1 & \Rightarrow \hat{\mathbf{q}}_i = \mathbf{q}_i \\ 0 & \Rightarrow \hat{\mathbf{q}}_i = 1 - \mathbf{q}_i \end{cases}.$$

Consequently, the estimation error is characterized by

$$\tilde{\mathbf{q}}_i \triangleq \mathbf{q}_i - \hat{\mathbf{q}}_i = (2\hat{\mathbf{q}}_i - 1)(\epsilon_i - 1). \quad (18)$$

Clearly, $\sum_{i=1}^m \epsilon_i = m - |\text{supp}(\hat{\mathbf{q}})|$ and is poisson-binomially distributed. Let $\mathbf{r} \in \mathbb{R}^{m+1}$ be a vector whose elements correspond to the probability mass function of $\sum_{i=1}^m \epsilon_i$, i.e.,

$$\Pr \left(\sum_{i=1}^m \epsilon_i = k - 1 \right) = \mathbf{r}_k, \quad k = 1, \dots, m + 1,$$

then [19]

$$\mathbf{r} = \alpha \begin{bmatrix} -s_1 \\ 1 \end{bmatrix} * \begin{bmatrix} -s_2 \\ 1 \end{bmatrix} * \dots * \begin{bmatrix} -s_m \\ 1 \end{bmatrix}, \quad (19)$$

where

$$\alpha = \prod_{i=1}^m \mathbf{p}_i, \quad s_i = -\frac{1 - \mathbf{p}_i}{\mathbf{p}_i} \quad (20)$$

¹The same argument can be made for $0 < \delta < 1$ as well by using the Young's Inequality $2\boldsymbol{\nu}_{\mathcal{T}^c}^\top (C_{\mathcal{T}^c}\tilde{\mathbf{x}}) \leq \frac{1}{\delta} \|\boldsymbol{\nu}_{\mathcal{T}^c}\|^2 + \delta \|C_{\mathcal{T}^c}\tilde{\mathbf{x}}\|^2$ instead.

and the symbol $*$ denotes the convolution operator for vectors.

Next, we seek the maximum integer $l_\eta \in \{0, \dots, m\}$ for which the oracle will correctly localize at least l_η nodes with a probability of at least η . Explicitly, l_η is given by:

$$l_\eta = \max \left\{ k \mid \Pr \left(\sum_{i=1}^m \epsilon_i \geq k \right) \geq \eta \right\} \quad (21)$$

$$= \max \left\{ k \mid 1 - \sum_{i=1}^{k+1} \mathbf{r}_i \geq \eta \right\}$$

$$= \max \left\{ k \mid \sum_{i=1}^{k+1} \mathbf{r}_i \leq 1 - \eta \right\}, \quad (22)$$

where $\eta \in (0, 1]$ is a reliability level and is set apriori.

Theorem 2. Consider the linear measurement model given in (3). Suppose there exists a localization oracle which gives an estimate, $\hat{\mathcal{T}}$, of $\text{supp}(\mathbf{e})$ with uncertainty described by (17). Given a reliability level $\eta \in (0, 1]$ with corresponding integer l_η given by (22), let $\hat{\mathcal{T}}_\eta$ be a new support with the indicator $\hat{\mathbf{q}}_\eta$ (of $\hat{\mathcal{T}}_\eta^c$) obtained by randomly matching l_η elements of $\hat{\mathbf{q}}$ while setting the remaining elements to 0. If $l_\eta - |\text{supp}(\mathbf{e})| \geq n$ then, with a probability of at least η , the least-square estimator;

$$\hat{\mathbf{x}}_\eta = \arg \min_{\mathbf{x} \in \mathcal{X}} \left\{ \frac{1}{2} \left\| \mathbf{y}_{\hat{\mathcal{T}}_\eta^c} - C_{\hat{\mathcal{T}}_\eta^c} \mathbf{x} \right\|^2 \right\}, \quad (23)$$

satisfies the error bound

$$\|\hat{\mathbf{x}}_\eta - \mathbf{x}^*\| \leq 2 \min \left\{ \frac{\delta}{2}, \frac{\varepsilon}{1 - \delta_n} \right\}, \quad (24)$$

where δ_n is the n -restricted isometry constant of C^\top and $\mathbf{x}^* \in \mathbb{R}^n$ is the unknown true internal state.

Proof. To avoid repetition, we state upfront that all claims made in this proof holds with a probability of at least η .

From the definition of l_η in (21), it follows that $\hat{\mathcal{T}}_\eta^c \subseteq \{\text{supp}(\mathbf{e})\}^c$, which implies that $\left\| \mathbf{y}_{\hat{\mathcal{T}}_\eta^c} - C_{\hat{\mathcal{T}}_\eta^c} \mathbf{x}^* \right\| \leq \varepsilon$. Using the optimality of the estimator and following the same procedure as in the proof of Theorem 1 yields

$$\|\hat{\mathbf{x}}_\eta - \mathbf{x}^*\| \leq 2 \min \left\{ \frac{\delta}{2}, \frac{\varepsilon}{\sigma_\eta} \right\},$$

where σ_η is the smallest singular value of $C_{\hat{\mathcal{T}}_\eta^c}$. Next, since $|\hat{\mathcal{T}}_\eta| \leq |\hat{\mathcal{T}}|$ and at least l_η nodes are correctly localized by $\hat{\mathcal{T}}$, the following hold:

$$\begin{aligned} l_\eta - |\hat{\mathcal{T}}_\eta^c| &= l_\eta - (m - |\hat{\mathcal{T}}_\eta|) \\ &\leq l_\eta - (m - |\hat{\mathcal{T}}|) \\ &= l_\eta - |\hat{\mathcal{T}}^c| \\ &\leq \text{supp}(\mathbf{e}). \end{aligned}$$

Using $l_\eta - |\text{supp}(\mathbf{e})| \geq n$, it follows that:

$$l_\eta - |\hat{\mathcal{T}}_\eta^c| \leq l_\eta - n,$$

which implies that $|\hat{\mathcal{T}}_\eta^c| \geq n$. Thus σ_η can be lower bounded as $\sigma_\eta \geq (1 - \delta_n)$, from which the result follows. \square

Remark 3. Although the above theorem gives a recipe for constructing a robust support from $\hat{\mathcal{T}}$, other heuristics can be developed to obtain a more reliable robust support for the reconstruction optimization objective. For example, rather than randomly matching l_η elements of $\hat{\mathbf{q}}$, we retain the localization output for the first l_η most trusted nodes. Suppose $\mathbf{s} \in [0, 1]^m$ is a vector of confidence values² for the localization output for each node, then a robust support can be given by:

$$\hat{\mathcal{T}}_\eta^c = \left\{ \hat{\mathcal{T}} \cap \{\text{argsort} \downarrow (\mathbf{p} \circ \mathbf{s})\}_1^{l_\eta} \right\}^c \quad (25)$$

VI. CASE STUDIES

A. Data Driven Model

Suppose, instead of the linear model in (6), one only has available the synchronous data tuple $\{\boldsymbol{\sigma}_i, \mathbf{y}_i\}$, $i = 1, \dots, N_d$, where $\mathbf{y}_i \in \mathbb{R}^m$ are measurements collected from the process of interest and $\boldsymbol{\sigma}_i \in \mathbb{R}^{n_\sigma}$ are associated vectors of exogenous monitoring variables³. Let $f(\cdot; \boldsymbol{\theta}) : \mathbb{R}^{n_\sigma} \mapsto \mathbb{R}^m$ be a parametrized nonlinear mapping with parameter $\boldsymbol{\theta}$ selected to minimize the empirical loss function:

$$J(\boldsymbol{\theta}) = \frac{1}{N_d} \sum_{i=1}^{N_d} \|\mathbf{y}_i - f(\boldsymbol{\sigma}_i; \boldsymbol{\theta})\|^2. \quad (26)$$

If N_d is big enough, the residual $\{\mathbf{y}_i - f(\boldsymbol{\sigma}_i; \boldsymbol{\theta}^*)\}$ can be assumed to be unbiased. Thus, a linear model of the uncertainty can be obtained by a subspace reduction on the residual data. This can be obtained for instance by performing a Principal Component Analysis (PCA) dimensionality reduction on the error covariance matrix

$$E = \frac{1}{N_d} \sum_{i=1}^{N_d} (\mathbf{y}_i - f(\boldsymbol{\sigma}_i; \boldsymbol{\theta}^*)) (\mathbf{y}_i - f(\boldsymbol{\sigma}_i; \boldsymbol{\theta}^*))^\top.$$

Let $\Phi \in \mathbb{R}^{m \times n}$, $n < m$ be a matrix whose columns correspond to the first principal vectors of E . Then the data driven linear model of the process to be used for resilient reconstruction is given by

$$\mathbf{y} = f(\boldsymbol{\sigma}; \boldsymbol{\theta}^*) + \Phi \mathbf{x} + \boldsymbol{\nu} + \mathbf{e}, \quad (27)$$

with the model error bound $\|\boldsymbol{\nu}\| \leq \mathcal{O}(s_{n+1})$, where s_k is the k th singular value of E .

Thus, given $\mathcal{T} = \text{supp}(\mathbf{e})$ with $|\mathcal{T}| \geq n$ and $\mathcal{T}^c \triangleq \{1, \dots, m\} \setminus \mathcal{T}$, the reconstruction is obtained as follows:

$$\begin{aligned} \hat{\mathbf{x}} &= \arg \min_{\mathbf{x} \in \mathcal{X}} \|\mathbf{y}_{\mathcal{T}^c} - f(\boldsymbol{\sigma}; \boldsymbol{\theta}^*)_{\mathcal{T}^c} - \Phi_{\mathcal{T}^c} \mathbf{x}\|^2 \\ \hat{\mathbf{y}}_{\mathcal{T}^c} &= \mathbf{y}_{\mathcal{T}^c} \\ \hat{\mathbf{y}}_{\mathcal{T}} &= f(\boldsymbol{\sigma}; \boldsymbol{\theta}^*)_{\mathcal{T}} - \Phi_{\mathcal{T}^c} \hat{\mathbf{x}} \end{aligned} \quad (28)$$

with the reconstruction error given by Corollary 1 (or Theorem 2 if \mathcal{T} is replaced with $\hat{\mathcal{T}}_\eta$).

²In most Machine Learning classification decision, it is possible to return an associated confidence value by calculating a normalized distance from the boundary.

³These are supporting source of measurement which are assumed to be secure. Examples of such include; ambient conditions, external vibration measurements, thermal images, network login information, etc

B. Linear Dynamical System

Next, we turn our attention to discrete linear time-invariant (LTI) systems which have been studied quite extensively in literature with respect to resilient estimation. First, we recap some results and then build on top of them using materials from previous sections.

Consider the discrete LTI system

$$\mathbf{x}_{k+1} = A\mathbf{x}_k \quad (29)$$

$$\mathbf{y}_k = C\mathbf{x}_k + \mathbf{e}_k, \quad (30)$$

where $\mathbf{x}_k \in \mathbb{R}^n$ represents the state of the system at time $k \in \mathbb{N}$, $\mathbf{y}_k \in \mathbb{R}^m$ is the output of the monitoring nodes at time k and $\mathbf{e}_k \in \mathbb{R}^m$ denote the attack signals injected by malicious agents at the monitoring nodes. Let $\mathcal{T} \subset \{1, 2, \dots, m\}$ denote the set of attacked nodes, then for all k , $|\text{supp}(\mathbf{e}_k)| \subset \mathcal{T}$. The resilient estimation problem is then defined as reconstructing the initial state \mathbf{x}_0 from corrupt measurement $\{\mathbf{y}_k\}_{k=0}^T$, $T \in \mathbb{N}$. We look at two scenarios from literature: \mathcal{T} is time-invariant [2], [3] and \mathcal{T} is time-varying [5].

1) Secure estimation for Fixed Attacked Nodes [2]

Assuming that the set \mathcal{T} of attacked nodes is time-invariant:

Definition 1. q errors are correctable after T steps by the decoder $\mathcal{D} : (\mathbb{R}^m)^T \mapsto \mathbb{R}^n$ if for any $\mathbf{x}_0 \in \mathbb{R}^n$, any $\mathcal{T} \subset \{1, 2, \dots, m\}$ with $|\mathcal{T}| \leq q$, and any sequence of vectors $\mathbf{e}_0, \dots, \mathbf{e}_{T-1} \in \mathbb{R}^m$ such that $\text{supp}(\mathbf{e}_k) \subset \mathcal{T}$, we have $\mathcal{D}(\mathbf{y}_0, \dots, \mathbf{y}_{T-1}) = \mathbf{x}_0$, where $\mathbf{y}_k = CA^k\mathbf{x}_0 + \mathbf{e}_k$ for $k = 0, 1, \dots, T-1$.

Proposition 1. Let $T \in \mathbb{N} \setminus \{0\}$. The following are equivalent: (i) There is a decoder that can correct q errors after T steps; (ii) For all $\mathbf{z} \in \mathbb{R}^n \setminus \{0\}$, $|\text{supp}(C\mathbf{z}) \cup \text{supp}(CA\mathbf{z}) \cup \dots \cup \text{supp}(CA^{T-1}\mathbf{z})| > 2q$.

Consequently, the following optimal decoder is defined for when the set of attacked nodes is fixed:

$$\mathbf{x}_0 = \arg \min_{\mathbf{x}} \|\mathbf{Y}_T - \Phi_T(\mathbf{x})\|_{l_0} \quad (31)$$

where

$$\mathbf{Y}_T = [\mathbf{y}_0 \mid \mathbf{y}_1 \mid \dots \mid \mathbf{y}_{T-1}] \in \mathbb{R}^{m \times T}$$

and $\Phi_T : \mathbb{R}^n \mapsto \mathbb{R}^{m \times T}$ is a linear map given by:

$$\Phi_T(\mathbf{x}) = [C\mathbf{x} \mid CA\mathbf{x} \mid \dots \mid CA^{T-1}\mathbf{x}] \in \mathbb{R}^{m \times T}.$$

2) Secure estimation for Varying Attacked Nodes [5]

Assuming that the set \mathcal{T} of attacked nodes can change with time but bounded as in $|\mathcal{T}| \leq q$:

Definition 2. q errors are correctable after T steps by the decoder $\mathcal{D} : (\mathbb{R}^m)^T \mapsto \mathbb{R}^n$ if for any $\mathbf{x}_0 \in \mathbb{R}^n$ and any sequence of vectors $\mathbf{e}_0, \dots, \mathbf{e}_{T-1} \in \mathbb{R}^m$ such that $|\text{supp}(\mathbf{e}_k)| \leq q$, we have $\mathcal{D}(\mathbf{y}_0, \dots, \mathbf{y}_{T-1}) = \mathbf{x}_0$, where $\mathbf{y}_k = CA^k\mathbf{x}_0 + \mathbf{e}_k$ for $k = 0, 1, \dots, T-1$.

Proposition 2. Let $\mathcal{T} \in \mathbb{N} \setminus \{0\}$. The following are equivalent:

- (i) There is a decoder that can correct q errors after T steps;
(ii) For all $\mathbf{z} \in \mathbb{R}^n \setminus \{0\}$, $\sum_{k=0}^{T-1} |\text{supp}(CA^k\mathbf{z})| > 2q$.

Consequently, the following optimal decoder is defined for when the set of attacked nodes is not fixed:

$$\mathbf{x}_0 = \arg \min_{\mathbf{x}} \|\mathbf{y}_{(T)} - \Phi_{(T)}\mathbf{x}\|_{l_0} \quad (32)$$

where

$$\mathbf{y}_{(T)} = \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_{T-1} \end{bmatrix} \in \mathbb{R}^{mT},$$

$$\Phi_{(T)} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{T-1} \end{bmatrix} \in \mathbb{R}^{mT \times n}.$$

Throughout the rest of the paper, we assume that the set \mathcal{T} of attacked nodes can change with time but bounded as in $|\mathcal{T}| \leq q$. Interested readers are referred to the reference [5] for detailed discussion of the advantage of this assumption over more traditional assumption of fixed set of attacked nodes.

3) Featurization of Observability Matrix

Given an integer $n_g \leq n$, consider the singular value decomposition of the observability matrix in (32) given by:

$$\begin{aligned} \Phi_{(T)} &= [U_1 \quad U_2] \left[\begin{array}{c|c} \Sigma_1 & \\ \hline & \Sigma_2 \end{array} \right] \begin{bmatrix} V_1^\top \\ V_2^\top \end{bmatrix} \\ &= U_1 \Sigma_1 V_1^\top + U_2 \Sigma_2 V_2^\top, \end{aligned} \quad (33)$$

where $U_1 \in \mathbb{R}^{mT \times n_g}$, $U_2 \in \mathbb{R}^{mT \times (n-n_g)}$, $\Sigma_1 \in \mathbb{R}^{n_g \times n_g}$ and $U_2 \in \mathbb{R}^{(n-n_g) \times (n-n_g)}$. We define the PCA feature corresponding to the actual set of measurement $\mathbf{y}_{(T)}^*$ as the linear transformation

$$\mathbf{g} \triangleq \Sigma_1 V_1^\top \mathbf{x}_0 = U_1^\top \mathbf{y}_{(T)}^*, \quad (34)$$

which results in the measurement model

$$\mathbf{y}_{(T)} = U_1 \mathbf{g} + \mathbf{e} + \mathbf{v}, \quad (35)$$

with $\|\mathbf{v}\| \leq \bar{\sigma}_{n_g+1} \|\mathbf{x}_0\|$ and $\mathbf{g} \in \mathcal{G}$, where the set \mathcal{G} is characterized by $\|\mathbf{g}_1 - \mathbf{g}_2\| \leq \delta$ for all $\mathbf{g}_1, \mathbf{g}_2 \in \mathcal{G}$ and $\bar{\sigma}_i$ is the i th biggest singular value of $\Phi_{(T)}$.

As will be seen later, using the feature measurement model in (35) for resilient estimation creates a tradeoff between reconstructability⁴ and reconstruction error.

It is also possible to use a more general nonlinear feature transform of the form $\mathbf{g} = \psi(\mathbf{y}_{(T)})$, $\psi(\cdot) : \mathbb{R}^{mT} \mapsto \mathbb{R}^{n_g}$. This introduces additional challenge of computing the inverse function $\psi^{-1}(\cdot) : \mathbb{R}^{n_g} \mapsto \mathbb{R}^{mT}$ which is not straightforward in general. However, since the scope of the present paper is limited to linear systems, the linear feature transform above turns out to be sufficient for the current work and the consideration of nonlinear feature transform for nonlinear systems is reserved for future work.

⁴defined as the total number of measurements that can be accurately reconstructed by a resilient estimator

4) Robust Resilient Reconstruction

The robust resilient estimator using the feature measurement model in (35) is given by:

$$\hat{\mathbf{g}}_k = \arg \min_{\|\mathbf{g} - \hat{\mathbf{g}}_{k-1}\| \leq \delta} \left\| \mathbf{y}(k) \hat{\tau}_\eta^c(k) - U_1 \hat{\tau}_\eta^c(k) \mathbf{g} \right\|^2 \quad (36)$$

$$\hat{\mathbf{x}}_0 = V_1 \Sigma_1^{-1} \hat{\mathbf{g}}_k$$

where

$$\mathbf{y}(k) = \begin{bmatrix} \mathbf{y}_{k-T} \\ \mathbf{y}_{k-T+1} \\ \vdots \\ \mathbf{y}_{k-1} \end{bmatrix} \in \mathbb{R}^{mT},$$

and $\hat{\tau}_\eta(k)$ is the robust support estimate, with a reliability level η , at time k .

Theorem 3. *Suppose a localization oracle, with a true positive rate \mathbf{p}_i for each node i , gives an estimated support $\hat{\mathcal{T}}$. Let $\hat{\mathcal{T}}_\eta(k)$ be given as described in Theorem 2 (or (25)). If $l_\eta - q \geq n_g$ then, with a probability of at least η , q errors are correctable by the resilient estimator given in (36). Furthermore, the estimation error is bounded as:*

$$\|\hat{\mathbf{x}}_0 - \mathbf{x}_0\| \leq \frac{\delta}{\bar{\sigma}_{n_g}} \min \left\{ 1, \frac{\bar{\sigma}_{n_g+1}}{1 - \delta_{n_g}} \right\}. \quad (37)$$

Proof. The result follows by following the same procedure in the proof of Theorem 2 using the measurement model in (35) and noting that

$$\begin{aligned} \|\mathbf{v}\| &\leq \bar{\sigma}_{n_g+1} \|\mathbf{x}_0\| = \bar{\sigma}_{n_g+1} \|V_1 \Sigma_1^{-1} \mathbf{g}\| \\ &\leq \frac{\bar{\sigma}_{n_g+1}}{\bar{\sigma}_{n_g}} \delta. \end{aligned}$$

□

Remark 4. *The size n_g of the feature vector \mathbf{g} creates a tradeoff between reconstructability and reconstruction error bound in that the maximum number of correctable error $q_{max} = \max\{q \in \mathbb{Z} | q \leq l_\eta - n_g\}$ decreases with n_g while the error bound improves ($1 - \delta_{n_g}$ increases) with n_g .*

VII. CONCLUSION

The problem of reconstructing compromised signals for a cyber-physical system under adversarial attack is considered. The approach merges results from machine learning an estimation theory to produce and optimization-based resilient estimator. It was shown that, under certain conditions, the signals can be reconstructed even with uncertain estimation of the support of the attack vector. The basic result was applied to two special cases; data-driven model and linear time-invariant system.

Finally, we highlight some open problems for future work. There is need to validate the theoretical claims via numerical simulation and experimental setup. Extension of the results to nonlinear systems via nonlinear feature transform is an open problem and reserved for future work. The constraint in the reconstruction optimization problem is a simple quadratic constraint. In future, we plan to consider more complicated

boundary function. The overall problem is set up for an open-loop scenario. It remains an open question what kind of effects would be expected if the resilient estimator is cascaded with a controller in a closed-loop setting. Also, only measurement error is considered with the attack uncertainty in the measurement model. It will be beneficial to see some results pertaining to model uncertainties using popular uncertainty model like integral quadratic constraints.

ACKNOWLEDGMENT

This study is supported under contract number DEOE0000833 awarded in 2016 by the United States Department of Energy (DOE)s Cyber-security for Energy Delivery (CEDs) R&D Program. Authors acknowledge the DOE\CEDs R&D Staff and the support of Justin John and Daniel Holzhauser from GE Global Research.

REFERENCES

- [1] L. K. Mestha, O. M. Anubi, and M. Abbaszadeh, "Cyber-attack detection and accommodation algorithm for energy delivery systems," in *Control Technology and Applications (CCTA), 2017 IEEE Conference on*. IEEE, 2017, pp. 1326–1331.
- [2] 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.
- [3] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, 2013.
- [4] J. Kim, C. Lee, H. Shim, Y. Eun, and J. H. Seo, "Detection of sensor attack and resilient state estimation for uniformly observable nonlinear systems," in *Decision and Control (CDC), 2016 IEEE 55th Conference on*. IEEE, 2016, pp. 1297–1302.
- [5] Q. Hu, D. Fooladivanda, Y. H. Chang, and C. J. Tomlin, "Secure state estimation for nonlinear power systems under cyber attacks," *arXiv preprint arXiv:1603.06894*, 2016.
- [6] M. Pajic, J. Weimer, N. Bezzo, O. Sokolsky, G. J. Pappas, and I. Lee, "Design and implementation of attack-resilient cyberphysical systems: With a focus on attack-resilient state estimators," *IEEE Control Systems*, vol. 37, no. 2, pp. 66–81, 2017.
- [7] G. Fiore, Y. H. Chang, Q. Hu, M. D. D. Benedetto, and C. J. Tomlin, "Secure state estimation for cyber physical systems with sparse malicious packet drops," in *2017 American Control Conference (ACC)*, May 2017, pp. 1898–1903.
- [8] S. Mishra, Y. Shoukry, N. Karamchandani, S. Diggavi, and P. Tabuada, "Secure state estimation: Optimal guarantees against sensor attacks in the presence of noise," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 2929–2933.
- [9] A. Martinelli, *Nonlinear Unknown Input Observability: Analytical expression of the observable codistribution in the case of a single unknown input*, pp. 9–15. [Online]. Available: <https://epubs.siam.org/doi/abs/10.1137/1.9781611974072.2>
- [10] S. Sefati, N. J. Cowan, and R. Vidal, "Linear systems with sparse inputs: Observability and input recovery," *2015 American Control Conference (ACC)*, pp. 5251–5257, 2015.
- [11] X. Liu, Y. Mo, and E. Garone, "Secure dynamic state estimation by decomposing kalman filter," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 7351 – 7356, 2017, 20th IFAC World Congress. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2405896317320682>
- [12] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE transactions on information theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [13] D. Hayden, Y. H. Chang, J. Goncalves, and C. J. Tomlin, "Sparse network identifiability via compressed sensing," *Automatica*, vol. 68, pp. 9–17, 2016.
- [14] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.
- [15] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2558–2567, 2002.

- [16] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE transactions on Information theory*, vol. 49, no. 12, pp. 3320–3325, 2003.
- [17] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information theory*, vol. 50, no. 10, pp. 2231–2242, 2004.
- [18] A. S. Bandeira, E. Dobriban, D. G. Mixon, and W. F. Sawin, "Certifying the restricted isometry property is hard," *IEEE transactions on information theory*, vol. 59, no. 6, pp. 3448–3450, 2013.
- [19] M. Fernández and S. Williams, "Closed-form expression for the poisson-binomial probability density function," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 2, pp. 803–817, 2010.