2019 IEEE 58th Conference on Decision and Control (CDC)
Palais des Congrès et des Expositions Nice Acropolis
Nice, France, December 11-13, 2019

When is the Secure State-Reconstruction Problem Hard?

Yanwen Mao, Aritra Mitra, Shreyas Sundaram and Paulo Tabuada

***Abstract*— This paper addresses the problem of reconstructing the state of a linear time-invariant system from malicious sensor measurements. The first result establishes that this problem is, in general, NP-hard. We then identify classes of subproblems that can be solved in polynomial time. When there are at most s malicious sensors, the problem can be solved in polynomial time when each eigenvalue is observable by at least $2s+1$ sensors. When each eigenvalue has geometric multiplicity one, this condition is equivalent to the system being $2s-$sparse observable. In contrast, the situation becomes more nuanced when each eigenvalue is not observable by at least $2s+1$ sensors, as we describe in detail in the paper.**

I. INTRODUCTION

In recent years, security has become a central issue in the design of Cyber-Physical Systems (CPS). These large-scale systems are often distributed, requiring sensitive information to be communicated from sensors to controllers and controllers to actuators [1]. The distributed nature as well as the large amount of exchanged information render such CPS vulnerable to adversaries who may wish to learn or even alter its state [2]–[4].

Motivated by these considerations, researchers have invested significant effort into solving the problem of reconstructing the state in the presence of sensor attacks[1], called the Secure State-Reconstruction (SSR) problem in this paper. The first experimental demonstration of a stealthy attack on a control system was reported in [5], and it was followed by the first theoretical results developed for special classes of systems [6], [7]. Stealthy attacks were then formalized in [8], [9]. An important step in the conceptual understanding of these attacks was given in [10]–[12], where the existence of such attacks was characterized by the system theoretic notion of zero-dynamics.

In addition to detecting and identifying attacks, it is important to mitigate their effect by continuing to control the plant. Since this requires, in general, reconstructing or estimating the state, there was a large body of literature published on the SSR problem since the papers [13], [14]. The combinatorial nature (which subset of sensors is under attack?) of the SSR problem led researchers to ingenious ways of reducing the complexity of identifying the attacked sensors. Examples include: convex relaxations [14], [15], distributed detection filters [12] , specialized observers under sparsity constraints [16], satisfiability modulo theory techniques [17], and safety envelopes [18].

Implicit in the paper [19] is a polynomial time algorithm for the SSR problem, suggesting that the SSR problem may be tractable in certain cases. However, to the best of the authors knowledge, a detailed discussion of the complexity of the SSR problem has not appeared in the literature. The goal of this paper is precisely to clarify the computational hardness of the SSR problem. As we shall soon see, two alternate notions of observability, namely "sparse observability" introduced in [14], [16] (see also [10] for an equivalent notion in continuous time), and "eigenvalue observability" [20], [21], will play key roles in our characterization. Our contributions are the following:

1) We show that the SSR problem is NP-hard.
2) We offer a polynomial-time solution for the SSR problem under an eigenvalue observability assumption.
3) We show that checking sparse observability is coNP-complete.
4) We show that the notions of sparse observability and eigenvalue observability are equivalent when the geometric multiplicity of each eigenvalue of the system matrix $\mathbf{A}$ is 1.

These results can be understood as follows. When the eigenvalues of the system matrix $\mathbf{A}$ have unitary multiplicity, the SSR problem is tractable: checking sparse observability, a necessary and sufficient condition for the SSR problem to be solvable, can be done in polynomial time since this notion becomes equivalent to eigenvalue observability which can be tested in polynomial time. If sparse observability holds, then the SSR problem can also be solved in polynomial time. When at least one of the eigenvalues has geometric multiplicity greater than one, we can still check eigenvalue observability and, if successful, solve the SSR problem in polynomial time. However, in this case, eigenvalue observability is no longer necessary for the SSR problem to be solvable. Since even checking sparse observability is coNP-complete, we conjecture that the SSR problem may be intractable in this case.

II. PRELIMINARIES AND NOTATIONS

The cardinality of a finite set $\mathcal{I} = \{i_1, \ldots, i_p\}$ is denoted by $|\mathcal{I}| = p$. For matrices $\mathbf{Q}_{i_1}, \ldots, \mathbf{Q}_{i_p}$ over the same field and with the same number of columns, we define the

This work was funded in part by the Army Research Laboratory under Cooperative Agreement W911NF-17-2-0196, the UC-NL grant LFR-18-548554, and in part by NSF CAREER award 1653648.

Yanwen Mao and Paulo Tabuada are with the Department of Electrical and Computer Engineering, University of California, Los Angeles, California, 90095, USA {yanwen.mao,tabuada}@ucla.edu

Arita Mitra and Shreyas Sundaram are with the department of Electrical Engineering, Purdue University, West Lafayette, Indiana, 47907, USA {mitra14,sundara2}@purdue.edu

Shreyas Sundaram thanks Lintao Ye for helpful discussions pertaining to the Linear Degeneracy problem.

[1]We will equivalently use the expression "attacked sensors" and "malicious sensors" since the manner in which the attack is conducted will not play a role in the results of this paper.

matrix $\mathbf{Q}_{\mathcal{I}} = \begin{bmatrix}\mathbf{Q}_{i_1}^T & \cdots & \mathbf{Q}_{i_p}^T\end{bmatrix}^T$ by stacking the individual matrices vertically. For a matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$, we use $sp(\mathbf{A})$ to denote the spectrum of the matrix $\mathbf{A}$, and $g_{\mathbf{A}}(\lambda)$ to denote the geometric multiplicity of an eigenvalue $\lambda \in sp(\mathbf{A})$. We will use the notation $\mathbf{I}_n$ to represent the identity matrix of dimension n. Given a vector $\mathbf{b} \in \mathbb{R}^n$, we denote by $\|\mathbf{b}\|_0$ the number of non-zero entries in $\mathbf{b}$.

III. PROBLEM FORMULATION

A. System Model

Consider a discrete-time linear time-invariant system under sensor attacks of the following form:

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

where $\mathbf{x}[k] \in \mathbb{R}^n$, $\mathbf{y}_i[k] \in \mathbb{R}^{p_i}$ and $\mathbf{e}_i[k] \in \mathbb{R}^{p_i}$ represent the state of the system, the measurement acquired by sensor i, and the attack vector for sensor i at time-step $k \in \mathbb{N}$, respectively. Let $\mathcal{V}$ denote the set of sensors, and let $N = |\mathcal{V}|$. We use $\mathbf{C} = \begin{bmatrix}\mathbf{C}_1^T & \cdots & \mathbf{C}_N^T\end{bmatrix}^T$ to denote the collection of the sensor observation matrices, $\mathbf{y}[k] = \begin{bmatrix}\mathbf{y}_1^T[k] & \cdots & \mathbf{y}_N^T[k]\end{bmatrix}^T$ and $\mathbf{e}[k] = \begin{bmatrix}\mathbf{e}_1^T[k] & \cdots & \mathbf{e}_N^T[k]\end{bmatrix}^T$ to represent the collective measurement vector and the collective attack vector.

We define $\mathcal{O}_i = \begin{bmatrix}\mathbf{C}_i^T & (\mathbf{C}_i\mathbf{A})^T & \dots & (\mathbf{C}_i\mathbf{A}^{\tau_i-1})^T\end{bmatrix}^T$ to be the observability matrix of sensor i, with τ_i being the observability index of the pair $(\mathbf{A}, \mathbf{C}_i)$. We also define $\mathbf{Y}_i = \begin{bmatrix}\mathbf{y}_i^T[0] & \dots & \mathbf{y}_i^T[\tau_i-1]\end{bmatrix}^T$ and $\mathbf{E}_i = \begin{bmatrix}\mathbf{e}_i^T[0] & \dots & \mathbf{e}_i^T[\tau_i-1]\end{bmatrix}^T$ to be the collection of measurements and attacks of sensor i over time, respectively. An equivalent expression for the measurements is:

$$\mathbf{Y}_i = \mathcal{O}_i\mathbf{x}[0] + \mathbf{E}_i.$$

In the remainder of the paper, we drop the time indices to simplify notation.

B. The Secure State-Reconstruction Problem

Problem 1. *(Secure state-reconstruction)*
Input: *Matrices* $\mathcal{O}_i \in \mathbb{R}^{p_i\tau_i\times n},\ i = 1,\dots,N$, *and a set of vectors* $\mathbf{Y}_i \in \mathbb{R}^{p_i\tau_i},\ i = 1,\dots,N$.
Question: *Find a vector* $\mathbf{x} \in \mathbb{R}^n$ *and a set* $\mathcal{I}$ *of minimal cardinality such that* $\mathbf{Y}_j = \mathcal{O}_j\mathbf{x}$ *for all* $j \notin \mathcal{I}$.

In other words, the SSR problem requires the reconstruction of the state $\mathbf{x}$ and the simplest attack explanation in the form of the least number of attacked sensors.

C. Sparse Observability and Eigenvalue Observability

The notions of sparse observability and eigenvalue observability are instrumental to the results in this paper.

Definition 1 (Sparse observability index)**.** *The sparse observability index of the pair* $(\mathbf{A}, \mathbf{C})$ *in system* (1) *is the largest integer* k *such that* $\ker\mathcal{O}_{\mathcal{V}\setminus\mathcal{K}} = \{0\}$ *for any* $|\mathcal{K}| \leq k,\ \mathcal{K} \subseteq \mathcal{V}$. *When the sparse observability index is* r, *we say that system* (1) *is* $r-$*sparse observable.* □

It is proved in [14], [16] (see also [22] for a similar notion in continuous time) that if at most s sensors have been attacked in a system, the initial state $\mathbf{x}[0]$ can be uniquely reconstructed if and only if the system is at least $2s-$sparse observable. In view of this result, computing the sparse observability index of a system is of great interest since it characterizes the maximum number of arbitrary sensor attacks that can be tolerated without compromising the ability to uniquely reconstruct the state.

In addition to sparse observability, we will require the notion of "eigenvalue observability" [20], [21].

Definition 2 (Eigenvalue observability index)**.** *We say that an eigenvalue* $\lambda \in sp(\mathbf{A})$ *is observable w.r.t. sensor* i *if the following condition holds:*

$$rank\begin{bmatrix}\mathbf{A} - \lambda\mathbf{I}_n\\ \mathbf{C}_i\end{bmatrix} = n. \tag{2}$$

If the above condition is satisfied, we equivalently say that "sensor i *can observe eigenvalue* λ*". Let the set of all sensors that can observe an eigenvalue* λ *be denoted* $\mathcal{S}_\lambda$. *The eigenvalue observability index of system* (1) *is the largest integer* k *such that each eigenvalue of the matrix* $\mathbf{A}$ *is observable by at least* $k+1$ *distinct sensors. When the eigenvalue observability index is* k *we say that system* (1) *is* k*-eigenvalue observable.* □

We study the SSR problem under the following assumptions.

Assumption 1: For each attacked sensor $i \in \{1,\dots,N\}$, the adversary can only manipulate sensor i's measurements through the signal $\mathbf{e}_i[k]$ in (1).

Assumption 2: The adversary is omniscient, i.e., we assume the adversary has full knowledge of the system state, measurements, and plant model. Moreover, all the malicious sensors are allowed to work cooperatively.

IV. SSR IS HARD

Fawzi *et al.* established in [14] a connection between the SSR problem and compressed sensing by drawing inspiration from the ideas of Candes and Tao in [23]. We take this approach further by also using the ideas in [23] to establish that the SSR problem is NP-hard.

Problem 2. *(Compressed sensing)*
Input: *A full row rank matrix* $\mathbf{F} \in \mathbb{Q}^{m\times n}$, *a vector* $\mathbf{b} \in \mathbb{Q}^m$.
Question: *Find the sparsest solution of* $\mathbf{F}\mathbf{x} = \mathbf{b}$.

The compressed sensing problem yields the solution to the minimization problem:

$$\min_{\mathbf{e}} \|\mathbf{e}\|_0 \quad \text{s.t. } \mathbf{F}\mathbf{e} = \mathbf{b}. \tag{3}$$

Theorem 1. *The SSR problem is NP-hard.* □

Proof. Given an instance of the compressed sensing problem, we generate an instance of the SSR problem as follows. Let the system matrix be of the form $\mathbf{A} = \mathbf{I}_n$, and the collective observation matrix $\mathbf{C}$ satisfy $\mathrm{Im}\mathbf{C} = \ker\mathbf{F}$. Let the measurements of the sensors be scalar-valued, i.e., let $\mathbf{C}_i$ be the i-th row of $\mathbf{C}$. Note that based on the above $\mathbf{A}$ matrix, the observability index for each sensor $i \in \{1,\dots,N\}$ is given by $\tau_i = 1$, and thus $\mathcal{O}_i = \mathbf{C}_i$. Finally, let $\mathbf{Y}$ be any

solution to the equation $\mathbf{FY} = \mathbf{b}$. Since the linear equation $\mathbf{FY} = \mathbf{b}$ is underdetermined, finding a solution $\mathbf{Y}$ can be done in polynomial time [24]. For each $i \in \{1, \ldots, N\}$, set $\mathbf{Y}_i$ to be the i-th row of $\mathbf{Y}$. Thus, given an instance of the compressed sensing problem, the above instance of the SSR problem can be constructed in polynomial time.

The SSR problem for the constructed instance degenerates to:

$$\min_{\mathbf{x},\mathbf{e}} \|\mathbf{e}\|_0 \quad \text{s.t. } \mathbf{Cx} + \mathbf{e} = \mathbf{Y}.$$

We now show these two problems have the same solution. It is easy to see that any solution $(\mathbf{x}, \mathbf{e})$ of $\mathbf{Cx} + \mathbf{e} = \mathbf{Y}$ yields a solution to $\mathbf{Fe} = \mathbf{b}$, since by applying $\mathbf{F}$ we obtain:

$$\mathbf{F}(\mathbf{Cx} + \mathbf{e}) = \mathbf{FY} \Leftrightarrow \mathbf{Fe} = \mathbf{b}. \tag{4}$$

To prove the converse, we show that for every $\mathbf{e}$ such that $\mathbf{Fe} = \mathbf{b}$, there exists some $\mathbf{x}$ satisfying $\mathbf{Cx} + \mathbf{e} = \mathbf{Y}$. Recalling that $\mathbf{FY} = \mathbf{b}$, we obtain $\mathbf{F}(\mathbf{Y} - \mathbf{e}) = \mathbf{0}$, i.e., $\mathbf{Y} - \mathbf{e} \in \ker\mathbf{F}$. Since $\ker\mathbf{F} = \mathrm{Im}\mathbf{C}$, there exists an $\mathbf{x}$ such that $\mathbf{Cx} = \mathbf{Y} - \mathbf{e}$, as desired.

Note that the equations $\mathbf{Fe} = \mathbf{b}$ and $\mathbf{Cx} + \mathbf{e} = \mathbf{Y}$ have the same solutions for $\mathbf{e}$; we conclude they also have the same sparsest solution. In other words, if there exists an algorithm $\mathcal{A}$ that solves the SSR problem for the specific instance constructed by us, such an algorithm will also yield a solution to the given instance of the compressed sensing problem. It then follows that since the compressed sensing problem is NP-hard, the SSR problem is also NP-hard. □

V. CLASSES OF SSR PROBLEMS SOLVABLE IN POLYNOMIAL TIME

While in the previous section we established that the SSR problem is NP-hard, the goal of this section is to prove the positive result that specific instances of the problem do admit polynomial time solutions.

Theorem 2. *Consider the system (1), and suppose at most s sensors are compromised. Let the eigenvalue observability index of system (1) be at least $2s$. Then, the SSR problem can be solved in polynomial time.* □

Proof. We prove the result in two steps. In the first step, we show that each sensor can recover the portion of the initial condition corresponding to the eigenvalues it can observe. Given this result, in the second step we argue that the sensors can collaboratively recover the entire initial condition vector via majority voting. We now outline the details of these steps.

Step 1: Recovering locally observable portions of initial conditions: First, perform a similarity transformation $\mathbf{x}[k] = \mathbf{Tz}[k]$ that maps $\mathbf{A}$ to its Jordan canonical form $\mathbf{J}$, and transforms (1) to the following form:

$$\begin{aligned} \mathbf{z}[k+1] &= \mathbf{Jz}[k], \\ \mathbf{y}_i[k] &= \bar{\mathbf{C}}_i\mathbf{z}[k], \quad \forall i \in \{1, \cdots, N\}, \end{aligned} \tag{5}$$

where $\mathbf{A} = \mathbf{TJT}^{-1}$ and $\bar{\mathbf{C}}_i = \mathbf{C}_i\mathbf{T}$. Let the set of eigenvalues of $\mathbf{A}$ that are observable w.r.t. sensor i be denoted $\mathcal{U}_i$. Now perform a second similarity transformation $\mathbf{z}[k] = \mathbf{P}_i\bar{\mathbf{z}}_i[k]$ that permutes the state vector $\mathbf{z}[k]$ in (5), yielding:

$$\begin{aligned} \underbrace{\begin{bmatrix}\mathbf{z}_{\mathcal{U}_i}[k+1]\\ \mathbf{z}_{\bar{\mathcal{U}}_i}[k+1]\end{bmatrix}}_{\bar{\mathbf{z}}_i[k+1]} &= \underbrace{\begin{bmatrix}\bar{\mathbf{J}}_{\mathcal{U}_i} & \mathbf{0}\\ \mathbf{0} & \bar{\mathbf{J}}_{\bar{\mathcal{U}}_i}\end{bmatrix}}_{\bar{\mathbf{J}}_i}\underbrace{\begin{bmatrix}\mathbf{z}_{\mathcal{U}_i}[k]\\ \mathbf{z}_{\bar{\mathcal{U}}_i}[k]\end{bmatrix}}_{\bar{\mathbf{z}}_i[k]}, \\ \mathbf{y}_i[k] &= \underbrace{\begin{bmatrix}\bar{\mathbf{C}}_{\mathcal{U}_i} & \bar{\mathbf{C}}_{\bar{\mathcal{U}}_i}\end{bmatrix}}_{\bar{\mathbf{G}}_i}\bar{\mathbf{z}}_i[k]. \end{aligned} \tag{6}$$

In the above equations, $\bar{\mathbf{J}}_{\mathcal{U}_i}$ is the collection of the Jordan blocks corresponding to the eigenvalues observable w.r.t. sensor i, and $\bar{\mathbf{J}}_{\bar{\mathcal{U}}_i}$ comprises of the remaining Jordan blocks in $\mathbf{J}$. Similarly, $\bar{\mathbf{C}}_{\mathcal{U}_i}$ contains the columns of $\bar{\mathbf{C}}_i$ corresponding to the matrix $\bar{\mathbf{J}}_{\mathcal{U}_i}$, with an analogous definition for $\bar{\mathbf{C}}_{\bar{\mathcal{U}}_i}$. Based on the above discussion, notice that the states corresponding to the eigenvalue set $\mathcal{U}_i$ have been grouped into the vector $\mathbf{z}_{\mathcal{U}_i}[k] \in \mathbb{R}^{o_i}$, where o_i represents the dimension of the square-matrix $\bar{\mathbf{J}}_{\mathcal{U}_i}$. We now describe how $\mathbf{z}_{\mathcal{U}_i}[0]$ can be recovered via the measurements available at sensor i. To this end, let $\bar{\mathbf{T}}_i$ be a non-singular matrix that performs an observable canonical decomposition of the pair $(\bar{\mathbf{J}}_{\bar{\mathcal{U}}_i}, \bar{\mathbf{C}}_{\bar{\mathcal{U}}_i})$ in (6). Consider the following transformation matrix:

$$\mathbf{T}_i = \begin{bmatrix}\mathbf{I}_{o_i} & \mathbf{0}\\ \mathbf{0} & \bar{\mathbf{T}}_i\end{bmatrix}. \tag{7}$$

Define the coordinate transformation $\bar{\mathbf{z}}_i[k] = \mathbf{T}_i\mathbf{v}_i[k]$. Based on this transformation, and using (6), we obtain:

$$\begin{aligned} \underbrace{\begin{bmatrix}\mathbf{z}_{\mathcal{U}_i}[k+1]\\ \mathbf{w}_{\mathcal{U}_i}[k+1]\\ \mathbf{w}_{\bar{\mathcal{U}}_i}[k+1]\end{bmatrix}}_{\mathbf{v}_i[k+1]} &= \underbrace{\left[\begin{array}{c|cc}\bar{\mathbf{J}}_{\mathcal{U}_i} & \multicolumn{2}{c}{\mathbf{0}}\\ \hline \mathbf{0} & \mathbf{M}_{\mathcal{U}_i} & \mathbf{0}\\ & \star & \mathbf{M}_{\bar{\mathcal{U}}_i}\end{array}\right]}_{\mathbf{T}_i^{-1}\bar{\mathbf{J}}_i\mathbf{T}_i}\underbrace{\begin{bmatrix}\mathbf{z}_{\mathcal{U}_i}[k]\\ \mathbf{w}_{\mathcal{U}_i}[k]\\ \mathbf{w}_{\bar{\mathcal{U}}_i}[k]\end{bmatrix}}_{\mathbf{v}_i[k]}, \\ \mathbf{y}_i[k] &= \underbrace{\left[\begin{array}{c|cc}\bar{\mathbf{C}}_{\mathcal{U}_i} & \mathbf{H}_{\mathcal{U}_i} & \mathbf{0}\end{array}\right]}_{\bar{\mathbf{G}}_i\mathbf{T}_i}\mathbf{v}_i[k], \end{aligned} \tag{8}$$

where:

$$\begin{aligned} \bar{\mathbf{T}}_i^{-1}\bar{\mathbf{J}}_{\bar{\mathcal{U}}_i}\bar{\mathbf{T}}_i &= \begin{bmatrix}\mathbf{M}_{\mathcal{U}_i} & \mathbf{0}\\ \star & \mathbf{M}_{\bar{\mathcal{U}}_i}\end{bmatrix}, \\ \bar{\mathbf{C}}_{\bar{\mathcal{U}}_i}\bar{\mathbf{T}}_i &= \begin{bmatrix}\mathbf{H}_{\mathcal{U}_i} & \mathbf{0}\end{bmatrix}. \end{aligned} \tag{9}$$

Define:

$$\mathbf{A}_i \triangleq \mathrm{diag}(\bar{\mathbf{J}}_{\mathcal{U}_i}, \mathbf{M}_{\mathcal{U}_i}), \quad \mathbf{H}_i \triangleq \begin{bmatrix}\bar{\mathbf{C}}_{\mathcal{U}_i} & \mathbf{H}_{\mathcal{U}_i}\end{bmatrix}, \tag{10}$$

and $\mathbf{s}_i[k] \triangleq \begin{bmatrix}\mathbf{z}_{\mathcal{U}_i}{}^T[k] & \mathbf{w}_{\mathcal{U}_i}{}^T[k]\end{bmatrix}^T$. Based on the above definitions, and referring to (8), we obtain:

$$\underbrace{\begin{bmatrix}\mathbf{y}_i[0]\\ \mathbf{y}_i[1]\\ \vdots\\ \mathbf{y}_i[n_i-1]\end{bmatrix}}_{\mathbf{Y}_i[0:n_i-1]} = \underbrace{\begin{bmatrix}\mathbf{H}_i\\ \mathbf{H}_i\mathbf{A}_i\\ \vdots\\ \mathbf{H}_i\mathbf{A}_i^{(n_i-1)}\end{bmatrix}}_{\mathbf{F}_i}\mathbf{s}_i[0], \tag{11}$$

where n_i is the dimension of $\mathbf{A}_i$. We claim that $\mathbf{F}_i$ as defined in the above equation has full column rank. To see this, notice that each of the pairs $(\bar{\mathbf{J}}_{\mathcal{U}_i}, \bar{\mathbf{C}}_{\mathcal{U}_i})$ and $(\mathbf{M}_{\mathcal{U}_i}, \mathbf{H}_{\mathcal{U}_i})$ are observable based on our construction. Furthermore, $\bar{\mathbf{J}}_{\mathcal{U}_i}$ and $\mathbf{M}_{\mathcal{U}_i}$ do not share any eigenvalues. In view of these facts, the

pair $(\mathbf{A}_i, \mathbf{H}_i)$, as defined in (10), is observable. This readily justifies our claim of $\mathbf{F}_i$ being full column rank, since $\mathbf{F}_i$ is precisely the observability matrix of the pair $(\mathbf{A}_i, \mathbf{H}_i)$. Consequently, the measurements at sensor i can be used to uniquely recover $\mathbf{s}_i[0]$ based on (11). This in turn implies recovery of $\mathbf{z}_{\mathcal{U}_i}[0]$, since $\mathbf{z}_{\mathcal{U}_i}[0] = \begin{bmatrix}\mathbf{I}_{o_i} & \mathbf{0}\end{bmatrix} \mathbf{s}_i[0]$. With this development in place, we proceed to the next step.

Step 2: Recovery of the entire state vector based on majority voting: Consider any eigenvalue $\lambda_j \in sp(\mathbf{A})$. Let the portion of the vector $\mathbf{z}[0]$ that corresponds to this eigenvalue be denoted $\mathbf{z}_j[0]$, and let $z_j^{(i)}[0]$ indicate the i-th component of $\mathbf{z}_j[0]$. Our development in step 1 implies that $\mathbf{z}_j[0]$ (and hence $z_j^{(i)}[0]$) can be recovered using the measurements of each of the sensors in the set $\mathcal{S}_{\lambda_j}$.[2] Based on the hypothesis of the theorem, $|\mathcal{S}_{\lambda_j}| \geq (2s+1)$. Consequently, since at most s sensors have been compromised, we are guaranteed at least $s+1$ consistent copies of the state $z_j^{(i)}[0]$. In other words, each component of the vector $\mathbf{z}_j[0]$ can be recovered via majority voting. Since the assertion holds for each $\lambda \in sp(\mathbf{A})$, it follows that $\mathbf{z}[0]$ can be securely reconstructed. Since a non-singular transformation maps $\mathbf{z}[0]$ to $\mathbf{x}[0]$, the latter can also be securely reconstructed. □

VI. COMPLEXITY OF CHECKING SPARSE OBSERVABILITY

In the previous two sections we studied the complexity of the SSR problem, and in particular, identified instances of the problem that can be solved in polynomial time. Recall that under at most s sensor attacks on the system (1), $2s$-sparse observability turns out to be necessary and sufficient for the SSR problem to yield a unique solution, namely the true initial state vector $\mathbf{x}[0]$. Given this result, we now take a step back and ask: What is the complexity of deciding whether a given system is $2s$-sparse observable? As explained in Section III-C, the above question is highly relevant since it aims to identify the maximum number of sensor attacks that can be tolerated by a given system of the form (1). In what follows, we show that determining the sparse-observability index (see Definition 1) of a system is computationally hard; we will focus on the case of scalar-valued sensors throughout, as that suffices to show the computational complexity of the problem.

Problem 3. *(r-sparse observability)*
Input: *A matrix $\mathbf{A} \in \mathbb{Q}^{n\times n}$, a matrix $\mathbf{C} \in \mathbb{Q}^{p\times n}$ and a positive integer r.*
Question: *Is the pair $(\mathbf{A}, \mathbf{C})$ r-sparse observable?*

Note that if the answer to an instance of the r-sparse observability problem is "no", then there is a simple proof: one can provide a set of r rows of $\mathbf{C}$ that, if removed, result in a system that is no longer observable. However, it is not clear whether there is a similarly simple proof for "yes" instances. Thus, the r-sparse observability problem is in the class coNP.[3]

The complement of a decision problem is the problem obtained by switching the "yes" and "no" answers to all instances of that problem. If a problem is in the class coNP, then its complement is in the class NP, and vice versa.

We will show that the r-sparse observability problem is coNP-hard by showing that its complement is NP-hard. Specifically, we define the following complement problem to r-sparse observability.

Problem 4. *(r-sparse unobservability)*
Input: *A matrix $\mathbf{A} \in \mathbb{Q}^{n\times n}$, a matrix $\mathbf{C} \in \mathbb{Q}^{p\times n}$ and a positive integer r.*
Question: *Is there a set of r rows that can be removed from $\mathbf{C}$ in order to yield a matrix $\bar{\mathbf{C}}$ such that $(\mathbf{A}, \bar{\mathbf{C}})$ is unobservable?*

Note that the answer to an instance of r-sparse unobservability is "yes" if and only if the answer to the corresponding instance of r-sparse observability is "no" and vice versa. Further note that r-sparse unobservability is in the class NP.

We show that r-sparse unobservability is NP-complete by providing a reduction from the following *Linear Degeneracy* problem. This problem was shown to be NP-complete in [26].

Problem 5. *(Linear Degeneracy [26])*
Input: *A full column rank matrix $\mathbf{F} \in \mathbb{Q}^{p\times n}$.*
Question: *Does $\mathbf{F}$ contain a degenerate (i.e., noninvertible) $n \times n$ submatrix?*

In other words, the linear degeneracy problem asks whether it is possible to remove $p-n$ rows from matrix $\mathbf{F}$ so that the resulting (square) matrix is not full rank. We are now ready to prove the following result.

Theorem 3. *The r-sparse unobservability problem is NP-complete. Thus, the r-sparse observability problem is coNP-complete.* □

Proof. Given an instance of the linear degeneracy problem (with matrix $\mathbf{F} \in \mathbb{Q}^{p\times n}$), we construct an instance of the r-sparse unobservability problem as follows: set $\mathbf{A} = \mathbf{I}_n$, $\mathbf{C} = \mathbf{F}$, and $r = p - n$.

We now show that the answer to the constructed instance of r-sparse unobservability is "yes" if and only if the answer to the given instance of linear degeneracy is "yes".

First, suppose that the answer to the constructed instance of r-sparse unobservability is "yes." Then there exists a set of r rows of $\mathbf{C}$ that can be removed such that the remaining rows are not sufficient to yield observability. However, since $\mathbf{A} = \mathbf{I}$, the above implies that there is a set of r rows of $\mathbf{C}$ that can be removed such that the remaining rows are not full column rank. Since $\mathbf{C} = \mathbf{F}$ and $r = p - n$, this means that there is an $n \times n$ submatrix of $\mathbf{F}$ that loses rank, and thus the answer to the linear degeneracy problem is "yes."

Next, we show that if the answer to the given instance of linear degeneracy is "yes," then the answer to the constructed

[2]Recall that $\mathcal{S}_{\lambda_j}$ represents the set of sensors w.r.t. which λ_j is observable.

[3]See, e.g., [25] for additional details on the complexity classes NP and coNP.

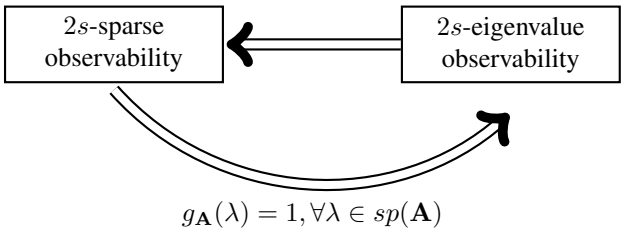

Fig. 1. Figure illustrating the hierarchy of relationships between different notions of observability.

instance of r-sparse unobservability is "yes." We will do this by showing the contrapositive: if the answer to the constructed instance of r-sparse unobservability is "no", then the answer to the given instance of linear degeneracy is "no." Suppose the answer to the constructed instance of r-sparse unobservability is "no." Then, by definition, the pair $(\mathbf{A}, \mathbf{C})$ is observable even after removing any arbitrary r rows from $\mathbf{C}$. However, since $\mathbf{A} = \mathbf{I}$, in order for the system to remain observable after removing r rows from $\mathbf{C}$, it must be the case that the remaining rows of $\mathbf{C}$ have full column rank. Thus, if the answer to the constructed instance of r-sparse unobserability is "no", then $\mathbf{C}$ has full column rank after removing any arbitrary $r = p - n$ rows. This means that every $n \times n$ submatrix of $\mathbf{C}$ is invertible. Since $\mathbf{C} = \mathbf{F}$, the answer to the given instance of linear degeneracy is "no" (i.e., there is no $n \times n$ submatrix of $\mathbf{F}$ that is degenerate).

Thus, we have shown that the answer to the constructed instance of r-sparse unobservability is "yes" if and only if the answer to the given instance of linear degeneracy is "yes." Since linear degeneracy is NP-complete, so is r-sparse unobservability.

Finally, since r-sparse observability is the complement of r-sparse unobservability (and since r-sparse observability is in coNP), we have that r-sparse observability is coNP-complete. □

VII. CONNECTIONS BETWEEN SPARSE OBSERVABILITY AND EIGENVALUE OBSERVABILITY

In Sections IV and VI, we showed that the SSR problem and the problem of determining the sparse observability index of a system are each computationally hard. At the same time, Section V gave us the positive result that certain instances of the SSR problem can be efficiently solved. This motivates the question: can the sparse observability index of a system be computed in polynomial time for certain specific instances? In this section, we show that this is indeed the case by identifying instances of the problem where the notions of sparse observability and eigenvalue observability coincide. Given that the eigenvalue observability index of a system can always be computed in polynomial time based on simple rank tests, an equivalence between the two notions of observability immediately yields instances of the problem where the sparse observability index of the system can also be computed in polynomial time. In this section we will prove each of the implications indicated in Figure 1. We begin with the following simple result.

Proposition 1. *Consider the linear system (1), and suppose its eigenvalue observability index is $2s$. Then, the pair $(\mathbf{A}, \mathbf{C})$ is at least $2s$-sparse observable.* □

Proof. Consider any subset of sensors $\mathcal{F} \subset \mathcal{V}$, such that $|\mathcal{F}| \leq 2s$. To establish that the pair $(\mathbf{A}, \mathbf{C})$ is at least $2s$-sparse observable, we need to show that the pair $(\mathbf{A}, \mathbf{C}_{\mathcal{V}\setminus\mathcal{F}})$ is observable. Based on the PBH test, this amounts to checking that each eigenvalue $\lambda \in sp(\mathbf{A})$ is observable w.r.t. the observation matrix $\mathbf{C}_{\mathcal{V}\setminus\mathcal{F}}$. A sufficient condition for this to happen is $|(\mathcal{V} \setminus \mathcal{F}) \cap \mathcal{S}_\lambda| \geq 1$, which is indeed true given that an eigenvalue observability index of $2s$ implies $|\mathcal{S}_\lambda| \geq (2f+1), \forall \lambda \in sp(\mathbf{A})$, and the fact that $|\mathcal{F}| \leq 2s$. □

Note that the reverse implication does not hold in general. In other words, $2s$-sparse observability of a system is in general less restrictive than the condition that the eigenvalue observability index of the system is $2s$. In what follows, we establish that the two aforementioned notions coincide when additional structure is imposed on the spectrum of $\mathbf{A}$.

Proposition 2. *Consider the linear system model (1), and suppose $\lambda \in sp(\mathbf{A})$ has geometric multiplicity 1. Consider any non-empty subset of sensors $\mathcal{S} = \{i_1, i_2, \ldots, i_{|\mathcal{S}|}\} \subseteq \mathcal{V}$. Then, the eigenvalue λ is observable w.r.t. the pair $(\mathbf{A}, \mathbf{C}_\mathcal{S})$ if and only if there exists a sensor $i_p \in \mathcal{S}$ such that λ is observable w.r.t. sensor i_p, i.e., λ is observable w.r.t. the pair $(\mathbf{A}, \mathbf{C}_{i_p})$.* □

Proof. Consider a similarity transformation that maps $\mathbf{A}$ to its Jordan canonical form $\mathbf{J}$. Let this transformation map $\mathbf{C}_\mathcal{S}$ to $\bar{\mathbf{C}}_\mathcal{S}$, and $\mathbf{C}_{i_j}$ to $\bar{\mathbf{C}}_{i_j}$, for each $i_j \in \mathcal{S}$. Since λ has geometric multiplicity 1, there exists a single Jordan block corresponding to λ in $\mathbf{J}$. Let this Jordan block be denoted $\mathbf{J}_\lambda$. Without loss of generality, suppose $\mathbf{J}$ is of the following form:

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_\lambda & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{J}} \end{bmatrix}, \tag{12}$$

where $\bar{\mathbf{J}}$ is the collection of the Jordan blocks corresponding to eigenvalues in $sp(\mathbf{A}) \setminus \{\lambda\}$. Based on the PBH test, λ is observable w.r.t. the pair $(\mathbf{J}, \bar{\mathbf{C}}_\mathcal{S})$ if and only if the following condition holds:

$$\text{rank} \begin{bmatrix} \mathbf{J} - \lambda \mathbf{I}_n \\ \bar{\mathbf{C}}_\mathcal{S} \end{bmatrix} = n. \tag{13}$$

Given the structure of $\mathbf{J}$ in (12), and the fact that λ has geometric multiplicity 1, it is easy to see that (13) holds if and only if there is at least one non-zero entry in the first column of $\bar{\mathbf{C}}_\mathcal{S}$. However, the preceding condition holds if and only if there exists some sensor $i_p \in \mathcal{S}$ with at least one non-zero entry in the first column of $\bar{\mathbf{C}}_{i_p}$; the latter is precisely the condition for observability of λ w.r.t. the sensor i_p, given that $g_{\mathbf{A}}(\lambda) = 1$. To complete the proof, it suffices to notice that a similarity transformation preserves the observability of an eigenvalue. □

We now use the above results to establish an equivalence between sparse observability and eigenvalue observability.

Proposition 3. *Consider the linear system model (1), and suppose every eigenvalue of* $\mathbf{A}$ *has geometric multiplicity* 1. *Then, the pair* $(\mathbf{A}, \mathbf{C})$ *is* $2s$*-sparse observable if and only if the eigenvalue observability of the system is* $2s$. □

Proof. For necessity, we proceed via contradiction. Suppose the pair $(\mathbf{A}, \mathbf{C})$ is $2s$-sparse observable, but there exists some $\lambda \in sp(\mathbf{A})$ that is observable w.r.t. at most $2s$ distinct sensors. Recall that the set of sensors w.r.t. which λ is observable is denoted $\mathcal{S}_\lambda$. Based on our hypothesis, $|\mathcal{S}_\lambda| \leq 2s$. Suppose $|\mathcal{S}_\lambda| = 2s$ (since an identical argument can be sketched when $|\mathcal{S}_\lambda| < 2s$). Based on our hypothesis, the pair $(\mathbf{A}, \mathbf{C}_{\mathcal{V}\setminus\mathcal{S}_\lambda})$ is observable. However, based on Prop. 2, this requires λ to be observable w.r.t. at least one sensor in $\mathcal{V} \setminus \mathcal{S}_\lambda$, leading to the desired contradiction. This completes the proof of necessity. For sufficiency, note from Prop. 1 that the pair $(\mathbf{A}, \mathbf{C})$ is at least $2s$-sparse observable whenever its eigenvalue observability index is $2s$; the fact that the observability index is no more than $2s$ follows from the additional assumption on the geometric multiplicity of eigenvalues, and arguments similar to those used for establishing necessity. □

It directly follows from the definition of eigenvalue observability that the eigenvalue observability index of a system can be computed in polynomial time. Hence, we have the following corollary.

Corollary 1. *When all the eigenvalues of the matrix* $\mathbf{A}$ *have geometric multiplicity 1, the sparse observability index of the system can be computed in polynomial time.* □

Corollary 2. *For a 2s-sparse observable system (1), when all the eigenvalues of the matrix* $\mathbf{A}$ *have geometric multiplicity 1, the SSR problem can be solved in polynomial time.* □

VIII. CONCLUSION

In this paper we showed that when the eigenvalues of the system matrix $\mathbf{A}$ have unitary multiplicity, the SSR problem is tractable since both checking the sparse observability (see Corollary 1) as well as solving the SSR problem (see Theorem 2) can be performed in polynomial time. When at least one of the eigenvalues has geometric multiplicity greater than one, we can still compute the eigenvalue observability index and, if it is at least $2s$, solve the SSR problem in polynomial time if at most s sensors are attacked. However, in this case eigenvalue observability is no longer necessary for the SSR problem to be solvable. Since even checking sparse observability is coNP-complete, we conjecture the SSR problem may be intractable in this case. The authors are currently investigating this conjecture.

REFERENCES

[1] D. P. Moller, *Guide to Computing Fundamentals in Cyber-Physical Systems*. Springer, 2016.

[2] A. A. Cárdenas, S. Amin, and S. Sastry, "Research challenges for the security of control systems," in *Proceedings of the 3rd Conference on Hot Topics in Security*, ser. HOTSEC'08, 2008, pp. 6:1–6:6.

[3] "Special issue on secure control of cyber physical systems," vol. 4, 2017.

[4] J. Giraldo, D. Urbina, A. Cardenas, J. Valente, M. Faisal, J. Ruths, N. O. Tippenhauer, H. Sandberg, and R. Candell, "A survey of physics-based attack detection in cyber-physical systems," *ACM Comput. Surv.*, vol. 51, no. 4, pp. 76:1–76:36, Jul. 2018.

[5] S. Amin, X. Litrico, S. S. Sastry, and A. M. Bayen, "Stealthy deception attacks on water scada systems," in *Proceedings of the 13th ACM International Conference on Hybrid Systems: Computation and Control*, ser. HSCC '10. ACM, 2010, pp. 161–170.

[6] H. Sandberg, A. Teixeira, and K. Johansson, "On security indices for state estimators in power networks," *Proc. 1st Workshop Secure Control Syst.*, 01 2010.

[7] A. Gupta, C. Langbort, and T. Baar, "Optimal control in the presence of an intelligent jammer with limited actions," in *49th IEEE Conference on Decision and Control (CDC)*, Dec 2010, pp. 1096–1101.

[8] R. S Smith, "A decoupled feedback structure for covertly appropriating networked control systems," *IFAC Proceedings Volumes (IFAC-PapersOnline)*, vol. 18, 08 2011.

[9] R. S. Smith, "Covert misappropriation of networked control systems: Presenting a feedback structure," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 82–92, Feb 2015.

[10] S. Sundaram and C. N. Hadjicostis, "Distributed function calculation via linear iterative strategies in the presence of malicious agents," *IEEE Trans. on Autom. Control*, vol. 56, no. 7, pp. 1495–1508, July 2011.

[11] F. Pasqualetti, F. Dörfler, and F. Bullo, "Cyber-physical security via geometric control: Distributed monitoring and malicious attacks," *Proc. of the 51st IEEE Conference on Decision and Control (CDC)*, pp. 3418–3425, 2012.

[12] ——, "Attack detection and identification in cyber-physical systems," *IEEE Trans. on Autom. Control*, vol. 58, no. 11, pp. 2715–2729, 2013.

[13] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure state-estimation for dynamical systems under active adversaries," in *Proc. of the 49th Annual Allerton Conference on Communication, Control, and Computing*, 2011, pp. 337–344.

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

[15] S. Z. Yong, M. Q. Foo, and E. Frazzoli, "Robust and resilient estimation for cyber-physical systems under adversarial attacks," in *2016 American Control Conference (ACC)*, July 2016, pp. 308–315.

[16] Y. Shoukry and P. Tabuada, "Event-triggered state observers for sparse sensor noise/attacks," *IEEE Trans. on Autom. Control*, vol. 61, no. 8, pp. 2079–2091, Aug 2016.

[17] Y. Shoukry, P. Nuzzo, A. L. Sangiovanni-Vincentelli, S. A. Seshia, G. J. Pappas, and P. Tabuada, "Smc: Satisfiability modulo convex programming," *Proceedings of the IEEE*, vol. 106, no. 9, pp. 1655–1679, Sep. 2018.

[18] A. Tiwari, B. Dutertre, D. Jovanović, T. de Candia, P. D. Lincoln, J. Rushby, D. Sadigh, and S. Seshia, "Safety envelope for security," in *Proceedings of the 3rd international conference on High confidence networked systems*. ACM, 2014, pp. 85–94.

[19] A. Mitra and S. Sundaram, "Byzantine-resilient distributed observers for LTI systems," *arXiv preprint arXiv:1802.09651*, 2018.

[20] C.-T. Chen, *Linear system theory and design*. Oxford University Press, Inc., 1998.

[21] A. Mitra and S. Sundaram, "Distributed observers for LTI systems," *IEEE Trans. on Autom. Control*, vol. 63, no. 11, pp. 3689–3704, 2018.

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

[23] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, Dec 2005.

[24] A. J. Laub, *Matrix Analysis for Scientists and Engineers*. SIAM, 2004.

[25] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms*. MIT press, 2009.

[26] L. Khachiyan, "On the complexity of approximating extremal determinants in matrices," *Journal of Complexity*, vol. 11, no. 1, pp. 138–153, 1995.