Complexity Reduction for Resilient State Estimation of Uniformly Observable Nonlinear Systems

Junsoo Kim, *Member, IEEE*, Jin Gyu Lee, *Member, IEEE*, Henrik Sandberg, *Fellow, IEEE*, and Karl H. Johansson, *Fellow, IEEE*

***Abstract*—A resilient state estimation scheme for uniformly observable nonlinear systems, based on a method for local identification of sensor attacks, is presented. The estimation problem is combinatorial in nature, and so many methods require substantial computational and storage resources as the number of sensors increases. To reduce the complexity, the proposed method performs the attack identification with local subsets of the measurements, not with the set of all measurements. A condition for nonlinear attack identification is introduced as a relaxed version of existing redundant observability condition. It is shown that an attack identification can be performed even when the entire state cannot be recovered from the measurements. As a result, although a portion of measurements are compromised, they can be locally identified and excluded from the state estimation, and thus, the true state can be recovered. Simulation results demonstrate the effectiveness of the proposed scheme.**

***Index Terms*—Nonlinear detection, redundancy, resilient state estimation, security, sensor attack identification.**

I. Introduction

Networked control based on advances in communication and computational power has enabled significant development in a number of industrial fields. Meanwhile, while more network connectivity leads to more abilities and versatility, the network layer is inherently more exposed to unauthorized access from third parties, such as malicious cyberattacks [1]. With the increasing threat of possible cyberattacks, a number of studies on security issues in networked systems have been presented [2], [3], [4].

A feature of both attack strategies and defense mechanisms for large-scale networked systems is that they may be based on substantial computation resources. When the attack is generated with knowledge of the system model and control signals as in [1], detection and identification of such an attack is generally NP-hard and combinatorial in nature [5].

Recent studies on the resilient state estimation problem [6] is a representative example where computational complexity has been of major interest, considering the combinatorial nature. Assuming that an adversary arbitrarily corrupts a limited number of sensor measurements in the control system, the problem is to figure out the compromised measurements, exclude their effect, and perform the state estimation correctly. For example, to identify and exclude q compromised sensors out of p sensors $(q < p)$, some early results in [5] and [7] construct not less than $\binom{p}{q}$ observers for the computation, and methods in [8] and [9] consider $\binom{p}{q}$ cases to search out the uncompromised sensors. As the algorithms generally cannot solve the problem in polynomial time and the required computation or storage resource drastically increases depending on the number of sensors, intensive efforts have been made to reduce the complexity of resilient state estimation.

Particularly for linear systems, many sufficient conditions that allow complexity reduction have been found. For example, a condition for converting the problem into a convex optimization problem was introduced in [6], and an assumption for making use of gradient decent algorithms was found in [10]. Conditions for dividing the problem into smaller problems, by a divide and conquer approach, was investigated in [11], [12]. Linear systems for which the problem is not NP-hard were discussed in [13], [14], [15], [16], and [17]. Distributed solutions to make the implementation scalable were further developed in [14], [15], and [16].

Despite that most control systems are nonlinear in practice, reducing the complexity of nonlinear resilient state estimation has rarely been considered. Exceptions include the results in [18] and [19], which extend the approaches of [7] and [8] to a class of nonlinear systems, respectively, but the order of the computation/storage complexity is at least $\binom{p}{q}$. Another attempt is made in [20], in which the satisfiability modulo theory is used to reduce the combinatorial complexity, but it assumes not only that the state but also the input of the system is reconstructed from the outputs, so that, for example, it is not applicable for systems having nontrivial zero dynamics.

This article presents resilient state estimation for uniformly observable nonlinear systems, based on a novel local attack identification method for reducing the computational complexity required. We first introduce a redundancy notion for nonlinear functions, under which attack identification is possible. Conventionally, it has been known that redundant observability is a necessary condition for resilient state estimation, as shown for linear systems in [6], and uniformly observable nonlinear systems [18]. Considering that resilient state estimation is to first identify the uncompromised sensors and then reconstruct the correct state estimate, we propose that observability or left-invertibility itself is not necessary when performing attack identification. Based on the introduced notion of redundant functions, we propose an attack inspection method, which guarantees that the presence of attack is detected whenever its size is large enough to be distinguished from measurement noise. The proposed method checks whether a set of measurement data is included in a predefined set or not, by measuring the distance from the set, and ensures that the data under inspection are not corrupted if and only if the measured distance is less than a precomputed threshold.

Received 6 June 2024; accepted 31 August 2024. Date of publication 11 September 2024; date of current version 30 January 2025. This work was supported in part by Seoul National University of Science and Technology, and in part by the Swedish Research Council. Recommended by Associate Editor A. Zemouche. *(Corresponding author: Jin Gyu Lee.)*

Junsoo Kim is with the Department of Electrical and Information Engineering, Seoul National University of Science and Technology, Seoul 01811, South Korea.

Jin Gyu Lee is with the ASRI, Department of Electrical and Computer Engineering, Seoul National University, Seoul 08826, South Korea.

Henrik Sandberg and Karl H. Johansson are with the Division of Decision and Control System, KTH Royal Institute of Technology, SE-10044 Stockholm, Sweden.

Digital Object Identifier 10.1109/TAC.2024.3459413

We further provide a condition under which the nonlinear resilient state estimation problem can be divided into smaller local problems which can be solved using only local information and measurements. The complexity can be significantly reduced when each local problem involves a proper subset of measurements. Projection mappings are used for removing a portion of measurement data which are not redundant.

The rest of this article is organized as follows. Section II formulates the problem, and Section III presents the main results. Section IV illustrates a numerical example. Finally, Section V concludes this article.

Notation: The set of natural and real numbers are denoted by $\mathbb{N}$ and $\mathbb{R}$, respectively. Define $[p] := \{1, 2, \dots, p\}$ for $p \in \mathbb{N}$. For a set of column vectors $z_i = \Phi_i(x)$, $i \in [p]$, and functions, for $I = \{i_1, i_2, \dots, i_k\} \subseteq [p]$, we simply define

$$\{z_i\}_{i\in I} := \begin{bmatrix} z_{i_1} \\ z_{i_2} \\ \vdots \\ z_{i_k} \end{bmatrix} \quad \text{and} \quad \{\Phi_i\}_{i\in I}(x) := \begin{bmatrix} \Phi_{i_1}(x) \\ \Phi_{i_2}(x) \\ \vdots \\ \Phi_{i_k}(x) \end{bmatrix},$$

respectively. The infinity norm of $z \in \mathbb{R}^n$ is denoted by $\|z\|$. For a compact set $\mathcal{X}$ and a vector z in $\mathbb{R}^n$, denote the distance from z to the set $\mathcal{X}$ by $d(z, \mathcal{X}) := \min_{x\in\mathcal{X}} \|z - x\|$. A function ϕ defined on a set X is said to be Lipschitz (on X), if there exists $L \geq 0$ such that $\|\phi(x_1) - \phi(x_2)\| \leq L\|x_1 - x_2\|$ holds for all $x_1 \in X$ and $x_2 \in X$. Let $L(\phi)$ denote the infimum of such constants L. A differentiable function ϕ on a set X is called an immersion if its Jacobian matrix at $x \in X$, denoted by $D\phi(x)$, is injective for all $x \in X$.

II. System Model and Problem

Consider a continuous-time input-affine system given by

$$\dot{x}(t) = f(x(t)) + g(x(t))u(t) \tag{1a}$$

$$y(t) = h(x(t)) + a(t) + v(t) \tag{1b}$$

where $x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}$ the input, $y(t) \in \mathbb{R}^p$ the output, $a(t) \in \mathbb{R}^p$ the attack signal, and $v(t) \in \mathbb{R}^p$ the measurement noise. We assume that $x(t)$, $u(t)$, and $v(t)$ are bounded. Especially for the state $x(t)$, we let $\mathcal{X} \subset \mathbb{R}^n$ be a compact set, which is known, such that $x(t) \in \mathcal{X}$ for all $t \geq 0$.

Attack Model: The model generating the sensor attack $a(t)$ is specified. First, the number of compromised sensors is limited; there exists $q < p$ such that up to q components of the attack $a(t) = \{a_i(t)\}_{i=1}^p$ (out of its p components) can be nonzero. Specifically, the cardinality of the set

$$I_0 := \{i \in [p] : a_i(t) = 0 \quad \forall t \geq 0\} \tag{2}$$

is such that $|I_0| \geq p - q$. There is no assumption for the nonzero components of the attack; for $i \in [p] \setminus I_0$, the attack signal $a_i(t)$ can be an arbitrary unbounded signal, which may be designed with knowledge of the system (1).

Problem 1 (Resilient state estimation): Construct a state estimator whose error does not depend on the attack. □

Although the set of compromised sensors are unknown, they should be found and excluded from the estimation. The next section provides a sufficient condition under which the resilient estimation is enabled and the required complexity for the method can be reduced.

III. Main Results

We begin by constructing a "partial" state observer for each individual measurement $y_i(t)$ of $y(t) \in \mathbb{R}^p$, $i = 1, 2, \dots, p$, employing the methods in [18] and [21]. It will facilitate contrasting the measurements and figuring out the uncorrupted ones.

A. Design of Partial High Gain Observers [18], [21]

As uniform observability [22] allows for nonlinear observer design in most cases, we assume that a certain portion of the state $x(t)$ is uniformly observable from each individual output $y_i(t)$. We particularly note that the state $x(t)$ is not being assumed to be (uniformly) observable from every output $y_i(t)$.

Assumption 1: For each $i \in [p]$, the system (1a) with the individual output $y_i(t)$ is diffeomorphic to the form

$$\dot{z}_i(t) = \{\dot{z}_{i,j}(t)\}_{j=1}^{n_i} = \begin{bmatrix} z_{i,2}(t) \\ \vdots \\ z_{i,n_i}(t) \\ \alpha_i(z_i(t)) \end{bmatrix} + \begin{bmatrix} \beta_{i,1}(z_{i,1}(t)) \\ \beta_{i,2}(z_{i,1}(t), z_{i,2}(t)) \\ \vdots \\ \beta_{i,n_i}(z_{i,1}(t), \dots, z_{i,n_i}(t)) \end{bmatrix} u(t) \tag{3a}$$

$$\dot{z}_i'(t) = F_i'(z_i(t), z_i'(t)) + G_i'(z_i(t), z_i'(t))u(t) \tag{3b}$$

$$y_i(t) = z_{i,1}(t) + a_i(t) + v_i(t) \tag{3c}$$

where $z_i(t) \in \mathbb{R}^{n_i}$ and $z_i'(t) \in \mathbb{R}^{n-n_i}$, with some $n_i \in \mathbb{N}$. □

We note that the function that maps the state $x(t) \in \mathbb{R}^n$ to the substate $z_i(t) \in \mathbb{R}^{n_i}$ is given by the ith component $h_i(\cdot)$ of the output function $h(\cdot) = \{h_i(\cdot)\}_{i=1}^p$ of (1b), and its Lie-derivatives along the vector field f, as

$$z_i(t) = \begin{bmatrix} z_{i,1}(t) \\ z_{i,2}(t) \\ \vdots \\ z_{i,n_i}(t) \end{bmatrix} = \begin{bmatrix} h_i(x(t)) \\ L_f h_i(x(t)) \\ \vdots \\ L_f^{n_i-1} h_i(x(t)) \end{bmatrix} =: \Phi_i(x(t)) \tag{4}$$

which can be verified by comparing (1) and (3) with $u(t) \equiv 0$.

Despite that the state $x(t)$ is generally not able to be reconstructed from an individual output $y_i(t)$, the substate $z_i(t) = \Phi_i(x(t))$ can be recovered as a piece of $x(t)$, unless the output $y_i(t)$ is corrupted. Indeed, for each $i \in [p]$, let a high gain observer for the system (3a) be designed, as

$$\dot{\hat{z}}_i(t) = \{\dot{\hat{z}}_{i,j}(t)\}_{j=1}^{n_i} = \begin{bmatrix} \hat{z}_{i,2}(t) \\ \vdots \\ \hat{z}_{i,n_i}(t) \\ \alpha_i(\hat{z}_i(t)) \end{bmatrix} + \begin{bmatrix} \beta_{i,1}(\hat{z}_{i,1}(t)) \\ \beta_{i,2}(\hat{z}_{i,1}(t), \hat{z}_{i,2}(t)) \\ \vdots \\ \beta_{i,n_i}(\hat{z}_{i,1}(t), \dots, \hat{z}_{i,n_i}(t)) \end{bmatrix} u(t) - P_i^{-1} C_i^\top (\hat{z}_{i,1}(t) - y_i(t)) \tag{5}$$

where $\hat{z}_i(t) \in \mathbb{R}^{n_i}$ is the estimate, $C_i := [1, 0, \dots, 0] \in \mathbb{R}^{n_i}$, and $P_i \in \mathbb{R}^{n_i \times n_i}$ is the unique positive-definite solution of

$$0 = -\theta_i P_i - A_i^\top P_i - P_i A_i + C_i^\top C_i$$

in which the parameter θ_i is to be determined, and

$$A_i := \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{n_i \times n_i}.$$

Since $x(t) \in \mathcal{X}$, and the substate $z_i(t) = \Phi_i(x(t))$ belongs to the image set $\Phi_i(\mathcal{X})$ of the function Φ_i, we let the initial state $\hat{z}_i(0)$ of (5) be chosen such that $\hat{z}_i(0) \in \Phi_i(\mathcal{X})$.

Now, the following proposition states that the observer (5) with the ith individual output $y_i(t)$ recovers the information of $z_i(t) = \Phi_i(x(t))$, in case the output $y_i(t)$ is not corrupted.

Proposition 1 ([18], [22]): There exist $\theta_i^* \geq 1$ and functions $\eta_i(\theta_i)$ and $\epsilon_i(\theta_i)$ such that, if $\theta_i \geq \theta_i^*$, the observer (5) ensures

$$\|\hat{z}_i(t) - z_i(t)\| \leq \max\{\eta_i(\theta_i) \cdot e^{-\frac{\theta_i}{8}t}, \epsilon_i(\theta_i)\} =: \delta_i(t) \tag{6}$$

in case $a_i(t) \equiv 0$. □

Proof: The proof can be found in [18]. ■

Remark 1: The benefit of constructing an observer for each individual output is that the majority of the estimates can be preserved, as many as the number of uncorrupted measurements. The estimates $\{\hat{z}_i(t)\}_{i\in I_0}$ from the uncorrupted measurements $\{y_i(t)\}_{i\in I_0}$, where the set I_0 is such that $|I_0| \geq p - q$, as defined in (2), ensures that (6) holds. Compared with the methods in [5] or [7], in which not less than $\binom{p}{q} = \frac{p!}{q!(p-q)!}$ observers are designed against q corrupted measurements out of p measurements, we construct comparatively less number of observers, as many as the number of measurements. □

In case there were no attacked elements in the partial estimates $\{\hat{z}_i(t)\}_{i\in[p]}$, obtaining the estimate for the state $x(t)$ would be possible, assuming that the system (1) is observable from the whole measurements $\{y_i(t)\}_{i=1}^p$, that is, the map

$$\Phi(\cdot) := \{\Phi_i(\cdot)\}_{i=1}^p$$

is left-invertible.

B. Condition for Attack Identification

One may expect that the resilient state estimation can be performed if the set of uncompromised measurements can be identified and the system (1) is observable or detectable from the identified measurements. Aside from the second step of state estimation and observability, this subsection investigates a condition under which the following problem is solvable, which can be regarded as a subproblem of Problem 1.

Problem 2 (Attack identification): Given information of $\{\hat{z}_i(t)\}_{i=1}^p$ and $\{\Phi_i\}_{i=1}^p$, find an index set $I \subseteq [p]$, $|I| = p - q$, such that $\|\hat{z}_i(t) - \Phi_i(x(t))\| \leq \delta_{\hat{z}}(t)$ holds for all $i \in I$, where the bound $\delta_{\hat{z}}(t) \geq 0$ should not depend on the attacks. □

We propose that the attack identification is possible in case the collections $\{\hat{z}_i(t)\}_{i=1}^p$ and $\{\Phi_i\}_{i=1}^p$ are "redundant." For a set $I \subseteq [p]$, we define the collections with respect to I, as

$$\hat{z}_I(t) := \{\hat{z}_i(t)\}_{i\in I} \quad \Phi_I(\cdot) := \{\Phi_i(\cdot)\}_{i\in I}.$$

Then, a novel notion of redundancy is introduced as follows.

Definition 1: A collection of maps $\Phi(\cdot) = \{\Phi_i(\cdot)\}_{i=1}^p$ on $\mathcal{X}$ is called k-redundant, if there exists $M_\Phi \geq 0$ such that

$$\|\Phi(x_1) - \Phi(x_2)\| \leq M_\Phi \|\Phi_I(x_1) - \Phi_I(x_2)\| \quad \forall x_1, x_2 \in \mathcal{X}$$

holds for all $I \subset [p]$ such that $|I| \geq p - k$. □

The condition of redundancy can be understood as follows. Consider the projection map

$$\pi_I : \{\Phi(x)\}_{i=1}^p \mapsto \{\Phi_i(x)\}_{i\in I}$$

defined on the image space $\Phi(\mathcal{X})$ of the map Φ. If Φ is k-redundant, it means that the map π_I is injective and left-invertible on the space $\Phi(\mathcal{X})$, since $\Phi_I(x_1) = \Phi_I(x_2)$ implies $\Phi(x_1) = \Phi(x_2)$. It implies that, although any k-components are removed from the collection $\Phi(x) = \{\Phi_i(x)\}_{i=1}^p$, $x \in \mathcal{X}$, the information of $\Phi(x)$ can be restored from $\{\Phi_i(x)\}_{i\in I}$. Furthermore, the constant M_Φ can be regarded as a Lipschitz constant for the left inverse of π_I for all I, on the image space.

Now, we claim and show that the identification against q-attacks is possible if the collection $\Phi = \{\Phi\}_{i=1}^p$ is $2q$-redundant. For a subset $I \subseteq [p]$ such that $|I| = p - q$, we let the distance from the vector $\hat{z}_I(t) = \{\hat{z}_i(t)\}_{i\in I}$ to the image space $\Phi_I(\mathcal{X})$ of $\Phi_I = \{\Phi_i\}_{i\in I}$ be computed. Then, to inspect if all the partial estimates $\{\hat{z}_i(t)\}_{i\in I}$ can be regarded as uncorrupted, we propose to check if the condition

$$d(\hat{z}_I(t), \Phi_I(\mathcal{X})) \leq \delta(t) \tag{7}$$

holds, where $\delta(t) := \max_{i\in[p]}\{\delta_i(t)\}$ given from (6) is an upper bound of the estimation error for the uncorrupted ones.

If $I \subseteq [p]$ indicates uncorrupted measurements only, i.e., if $I \subseteq I_0$, then it is obvious from (6) that (7) is true. It means that if (7) is violated, then it is clear that there is a compromised estimate in the collection $\hat{z}_I(t) = \{\hat{z}_i(t)\}_{i\in I}$. The following theorem states that the converse is also true, when the function Φ satisfies the redundancy condition; if (7) holds, then it ensures that the estimates in $\hat{z}_I(t) = \{\hat{z}_i(t)\}_{i\in I}$ have not been affected by the attack or its effect is negligible.

Theorem 1: Assuming that Φ is $2q$-redundant, let (7) hold for an index set $I \subset [p]$ such that $|I| = p - q$. Then,

$$\|\hat{z}_I(t) - \Phi_I(x(t))\| \leq (2M_\Phi + 1)\delta(t) \tag{8}$$

holds with some $M_\Phi \geq 0$. □

Proof: The set $\mathcal{X}$ is compact, so (7) implies that there exists $x' \in \mathcal{X}$ such that $\|\hat{z}_I(t) - \Phi_I(x')\| \leq \delta(t)$. It follows that

$$\|\hat{z}_I(t) - \Phi_I(x(t))\| \leq \delta(t) + \|\Phi_I(x(t)) - \Phi_I(x')\|.$$

Since $\Phi = \{\Phi_i\}_{i\in[p]}$ is $2q$-redundant and $|I| = p - q$, note that the map $\Phi_I = \{\Phi_i\}_{i\in I}$ is q-redundant. It implies that

$$\|\Phi_I(x(t)) - \Phi_I(x')\| \leq M_\Phi \|\Phi_{I\cap I_0}(x(t)) - \Phi_{I\cap I_0}(x')\|$$

holds with some $M_\Phi \geq 0$, where $\Phi_{I\cap I_0} = \{\Phi_i\}_{i\in I\cap I_0}$, because $|I \cap I_0| \geq p - 2q$. Proposition 1 for the indexes in I_0 ensures

$$\begin{aligned}&\|\Phi_{I\cap I_0}(x(t)) - \Phi_{I\cap I_0}(x')\| \\ &\leq \|\hat{z}_{I\cap I_0}(t) - \Phi_{I\cap I_0}(x')\| + \delta(t) \leq 2\delta(t)\end{aligned}$$

in which $\hat{z}_{I\cap I_0}(t) = \{\hat{z}_i(t)\}_{i\in I\cap I_0}$. It completes the proof. ■

Theorem 1 implies that, if (7) holds for a collection $\hat{z}_I(t) = \{\hat{z}_i(t)\}_{i\in I}$, the effect of the attack in the partial estimate error in $\|\hat{z}_I(t) - \Phi_I(x(t))\|$ is small enough to be regarded as uncorrupted. Although there may be an index $i \in I$ such that $a_i(t) \not\equiv 0$, its effect cannot be distinguished from the estimation error in (6) for the uncorrupted measurements.

Given the partial estimates $\{\hat{z}_i(t)\}_{i=1}^p$, the attack identification can be performed by applying the inspection (7) for each collection $\{\hat{z}_i(t)\}_{i\in I}$ such that $|I| = p - q$. Since there must exist an estimation set satisfying (7) and such a set will guarantee (8), it can be concluded that Problem 2 is solvable if the map $\Phi = \{\Phi_i\}_{i=1}^p$ is $2q$-redundant. The number of cases to inspect is $\binom{p}{q}$, which can be regarded as the required complexity for this attack identification method described.

Remark 2: We emphasize that the investigated $2q$-redundancy condition does not require left invertibility of the function Φ. For resilient state estimation, it has been investigated as in [5], [6], [7], [8], [10], and [18] that the condition "$2q$-redundant observability" is necessary, which in fact implies that the function Φ is not only $2q$-redundant but also left-invertible. For example, in case the system (1) as well as the function Φ defined in (4) are linear, the results in [6], [7], and [8] assumed that each function $\{\Phi_i\}_{i\in I}$ with $|I| = p - 2q$ is injective as a linear map. However, for attack identification only, we have shown that only the $2q$-redundancy condition is required, and the left invertibility is

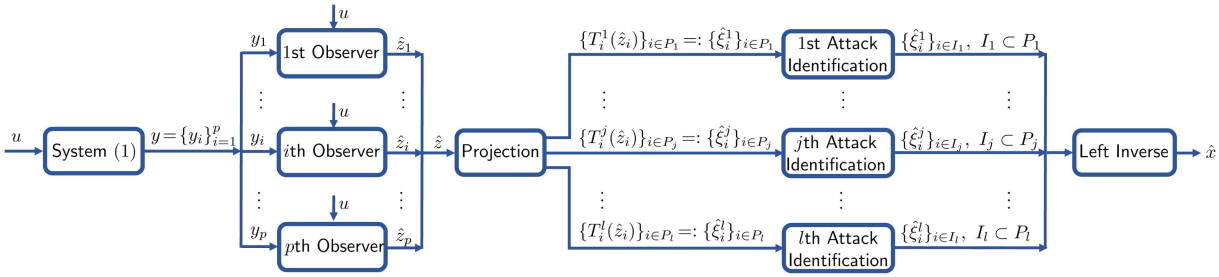

Fig. 1. Configuration of the proposed scheme as three steps: i) partial state observer design, ii) attack identification with projected local estimates, and iii) state reconstruction with identified estimates.

not required. We will exploit this point in the next subsection to reduce the computational effort required for the resilient state estimation. □

C. Resilient State Estimation Via Local Attack Identification

We recall that the complexity of attack identification with respect to the whole measurements is typically $\binom{p}{q}$, which rapidly grows as the number p of measurements increases. In this regard, our approach for reducing the complexity is to divide the set $\{\hat{z}_i(t)\}_{i\in[p]}$ into local sets of estimates, and perform *the attack identification with respect to each local set*.

Let $\{\hat{z}_i(t)\}_{i\in P}$, $P \subset [p]$, be a subset of partial estimates, which corresponds to the subset of functions $\{\Phi_i\}_{i\in P}$. Although it is hard to expect that $\{\Phi_i\}_{i\in P}$ is left-invertible, i.e., the system (1) observable from a local subset of measurements, "local" attack identification with respect to the index set P will be possible if the function $\{\Phi_i\}_{i\in P}$ is $2q$-redundant.

Meanwhile, the following example shows that a function that is not redundant may become redundant, by taking a projection mapping to keep the redundant part only.

Example: Let $x = (r, \theta) \in \mathcal{X} = [1, 2] \times [\pi/8, \pi/4]$, and let

$$\Phi(x) = \begin{bmatrix} \Phi_1(x) \\ \Phi_2(x) \\ \Phi_3(x) \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} r\cos\theta \\ r\sin\theta \end{bmatrix} \\ \tan\theta \\ \theta \end{bmatrix}$$

consist of three components of functions. The function Φ is clearly not 2-redundant, as the map $\pi_{\{3\}} : \Phi(x) \mapsto \Phi_3(x) = \theta$ is not injective and the value of r cannot be restored from θ. But, with a projection $T(z) = z/\sqrt{z^\top z}$, $z \in \mathbb{R}^2$, which removes the information of r from $\Phi_1(x)$ as $T(\Phi_1(x)) = [\cos\theta, \sin\theta]^\top$, the projected function $\Phi'(x) = \{T(\Phi_1(x)), \Phi_2(x), \Phi_3(x)\}$ becomes 2-redundant.

With this observation, the scheme of local attack identification is presented, where a set of projections can be used for refining the sets $\{\Phi_i\}$ of functions to be redundant. Let $P_1, P_2, \ldots, P_l \subset [p]$ be local index sets that cover $[p]$, i.e., $\cup_{j=1}^l P_j = [p]$, and for each $j = 1, \ldots, l$ and $i \in P_j$, let $T_i^j : \mathbb{R}^{n_i} \to \mathbb{R}^{n_i^j}$, $n_i^j \le n_i$, be a Lipschitz function for projecting the estimate $\hat{z}_i(t) \in \mathbb{R}^{n_i}$, with respect to P_j. With

$$\begin{aligned} \hat{\xi}^j(t) = \{\hat{\xi}_i^j(t)\}_{i\in P_j} &:= \{T_i^j(\hat{z}_i(t))\}_{i\in P_j} \\ \Psi^j(x) = \{\Psi_i^j(x)\}_{i\in P_j} &:= \{T_i^j(\Phi_i(x))\}_{i\in P_j} \end{aligned} \tag{9}$$

the property (6) for the uncorrupted estimates is turned into

$$\|\hat{\xi}_i^j(t) - \Psi_i^j(x(t))\| \le L(T_i^j)\delta_i(t) \tag{10}$$

where $L(T_i^j)$ is a Lipschitz constant of T_i^j.

Now, we propose that the attack identification method (7) be applied to each local estimate $\hat{\xi}^j(t)$, as

$$d(\hat{\xi}_{I_j}^j(t), \Psi_{I_j}^j(\mathcal{X})) \le \delta^j(t) := \max_{i\in P_j}\{L(T_i^j)\delta_i(t)\} \tag{11}$$

where for each $I_j \subset P_j$ we define

$$\hat{\xi}_{I_j}^j := \{\hat{\xi}_i^j\}_{i\in I_j} \quad \text{and} \quad \Psi_{I_j}^j := \{\Psi_i^j\}_{i\in I_j}$$

in which the index set $I_j \subset P_j$ removes q elements out of $\hat{\xi}^j(t) = \{\hat{\xi}_i^j(t)\}_{i\in P_j}$, i.e., $I_j \subset P_j$ is such that $|I_j| = |P_j| - q$.

As a result, the following theorem presents our approach to the resilient state estimation, based on the proposed attack identification applied to each local estimate $\hat{\xi}^j(t)$. See Fig. 1 describing the overall configuration of the proposed scheme.

Theorem 2: Consider the local estimates in (9). Assume that the function Ψ^j is $2q$-redundant for $j = 1, 2, \ldots, l$, and the collection $\{\Psi^j\}_{j=1}^l$ is an injective immersion on $\mathcal{X}$. For each $j = 1, 2, \ldots, l$, there exists $I_j \subset P_j$ satisfying (11) with $|I_j| = |P_j| - q$. Then, there exists a Lipschitz map ψ such that

$$\left\|\psi\left(\{\hat{\xi}_{I_j}^j(t)\}_{j=1}^l\right) - x(t)\right\| \le L(\psi)\delta'(t), \quad \forall t \ge 0 \tag{12}$$

where $\delta'(t) := \max_j\{(2M_{\Psi^j} + 1)\delta^j(t)\}$. □

Proof: In case $I_j \subseteq I_0$, the property (6) for $i \in I_j$ implies that (10) holds, thanks to the Lipschitzness of T_i^j. It follows that (11) holds for each $j = 1, 2, \ldots, l$. Thus, this ensures the existence of the sets $\{I_j\}_{j=1}^l$ satisfying (11). Next, since every Ψ^j is $2q$-redundant, according to Theorem 1, it ensures that

$$\|\hat{\xi}_{I_j}^j(t) - \Psi_{I_j}^j(x(t))\| \le (2M_{\Psi^j} + 1)\delta^j(t) \tag{13}$$

holds, where M_{Ψ^j} is a constant such that

$$\|\Psi^j(x_1) - \Psi^j(x_2)\| \le M_{\Psi^j}\|\Psi_{I_j}^j(x_1) - \Psi_{I_j}^j(x_2)\|$$

holds for all $x_1 \in \mathcal{X}$ and $x_2 \in \mathcal{X}$. The $2q$-redundancy of Ψ^j for each j also implies that the map

$$\pi_{I_j} : \Psi^j(x) = \{\Psi_i^j(x)\}_{i\in P_j} \mapsto \Psi_{I_j}^j(x) = \{\Psi_i^j(x)\}_{i\in I_j}$$

is invertible on the image set $\Psi_{I_j}^j(\mathcal{X})$. The inverse of π_{I_j} defined on $\Psi_{I_j}^j(\mathcal{X})$, denoted by $\pi_{I_j}^{-1}$, is a Lipschitz function. According to Kirszbraun's Lipschitz extension theorem [23], there exists[1] a Lipschitz

[1]For a Lipschitz function $\phi : X \to \mathbb{R}^n$, a Lipschitz extension of each component $\phi_i(\cdot)$ of $\phi(\cdot) = \{\phi_i(\cdot)\}_{i=1}^n$ can be found as

$$\overline{\phi}_i(x) := \inf_{x'\in X}\{\phi_i(x') + L(\phi)\|x - x'\|\}.$$

See [23, p. 21] for more details.

function ψ_j for each j that is defined on the whole Euclidean space and satisfies

$$\psi_j(\Psi^j_{I_j}(x)) = \pi^{-1}_{I_j}(\Psi^j_{I_j}(x)) = \Psi^j(x)$$

for every $x \in \mathcal{X}$. And, since the collection $\{\Psi^j\}_{j=1}^l$ is an injective immersion on $\mathcal{X}$, according to [18, Proposition 1], there exists a Lipschitz map Ψ^{-1} such that $\Psi^{-1}(\{\Psi^j(x)\}_{j=1}^l) = x$ for all $x \in \mathcal{X}$. Now, define the function ψ by

$$\psi\left(\{\hat{\xi}^j_{I_j}(t)\}_{j=1}^l\right) := \Psi^{-1}\left(\{\psi_j(\hat{\xi}^j_{I_j}(t))\}_{j=1}^l\right).$$

Clearly, ψ is a Lipschitz function and it satisfies

$$\psi\left(\{\Psi^j_{I_j}(x)\}_{j=1}^l\right) = \Psi^{-1}\left(\{\psi_j(\Psi^j_{I_j}(x))\}_{j=1}^l\right) = x, \forall x \in \mathcal{X}.$$

Therefore, (13) satisfied for all $j = 1, 2, \ldots, l$ ensures that (12) is true. The proof is completed. ■

Implication of Theorem 2 can be understood as follows. If the $2q$-redundant function Φ can be decomposed into $\{\Psi^j\}_{j=1}^l$ which are still $2q$-redundant, although each Ψ^j would not be left-invertible and indicate only a portion of the state $x(t)$, the uncorrupted partial estimates can be identified locally by applying the criterion (11) for each j. And, if the collection $\{\Psi^j\}_{j=1}^l$ of all projected functions retains the observability of the system so that it is still left-invertible, the information of $x(t)$ can be restored from the locally identified estimates.

Remark 3: While the number of cases that the conventional identification considered in the previous section and in the previous results [5], [18], [19] was not less than $\binom{p}{q}$, the proposed scheme by local identification requires that $\sum_{j=1}^l \binom{|P_j|}{q}$ cases are considered, since the number of the cases for the jth local set $\{\hat{\xi}^j_i(t)\}_{i\in P_j}$ is $\binom{|P_j|}{q}$. Thus, the larger number p of the sensors and the smaller number of the elements for each group P_j are considered, the more benefit the proposed scheme takes, in terms of computational efforts and consumed resources for the implementation. ■

IV. Numerical Example

Consider a numerical example for the system (1), given as

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} -x_1 + \frac{1}{2}x_3^2 - x_2 x_3 \cos x_2 \\ -x_2 \\ -x_2 \cos x_2 \end{bmatrix} + \begin{bmatrix} x_3 + x_3 \cos x_2 \\ 1 \\ 1 + \cos x_2 \end{bmatrix} u$$

with $u(t) = 0.25\sin(0.2\pi t)$, where the state $x(t)$ remains in the set $\mathcal{X} = \{x \in \mathbb{R}^3 : \|x\| \le 0.5\}$ when its initial value is given sufficiently small. For the outputs (1b), let there be 20-measurements with the functions $h = \{h_i\}_{i=1}^{20}$ given as

$$h_i(x) = \begin{cases} x_1 - \frac{1}{2}x_3^2 + \frac{i}{10}x_2, & i = 1, \ldots, 10 \\ \frac{1}{2}x_3 - \frac{1}{2}\sin x_2, & i = 11, \ldots, 20 \end{cases}$$

and the noise $v(t) \in \mathbb{R}^{20}$ be given such that $\|v(t)\| \le 0.01$.

In this example, the partial information $\Phi_i(x)$ of the state x obtained from each individual output y_i is no more than $h_i(x)$, i.e., $\Phi_i = h_i$ $\forall i$, as it can be computed from (4) that $L_f h_i = -h_i$ for $i \le 10$, and $L_f h_i = 0$ for $i \ge 11$.

Then, it can be verified that the state $x(t)$ can be restored from any twelve elements of the measurements $\{y_i(t)\}_{i=1}^{20}$, so that the attack identification and the resilient state estimation is possible when up to four measurements are compromised. If the attack identification is performed with all the measurements, then the number of cases is equal to $\binom{20}{4} = 4845$.

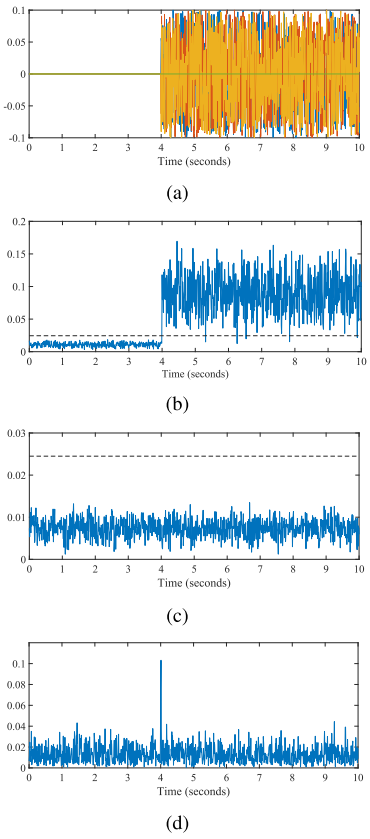

Fig. 2. Simulation results. (a) Injected attack signals $\{a_i(t)\}_{i\in[4]}$. (b) Distance function $\|\{y_i(t)\}_{i\in I_1} - \mathcal{O}_{I_1}\mathcal{O}^\dagger_{I_1}\{y_i(t)\}_{i\in I_1}\|_2$ for $I_1 = \{1, 2, \ldots, 6\}$, and its threshold (black dash-dot). (c) Distance function $\|\{y_i(t)\}_{i\in I_1} - \mathcal{O}_{I_1}\mathcal{O}^\dagger_{I_1}\{y_i(t)\}_{i\in I_1}\|_2$ for $I_1 = \{5, 6, \ldots, 10\}$, and its threshold. (d) Estimation error from $x(t)$.

But, the attack identification can also be performed with the local measurements $\{y_i(t)\}_{i=1}^{10}$ and $\{y_i(t)\}_{i=11}^{20}$, separately, although the state $x(t)$ is not observable from either of them. This is because both the collections $\{\Phi_i\}_{i=1}^{10}$ and $\{\Phi_i\}_{i=11}^{20}$ of functions are 8-redundant on any compact set in $\mathbb{R}^3$; for example, removing any eight elements from $\{\Phi_i\}_{i=1}^{10}$ and given any $\Phi_{i_1}(x)$ and $\Phi_{i_2}(x)$ with $1 \le i_1 < i_2 \le 10$, both the information of $x_1 - \frac{1}{2}x_3^2$ and x_2 can be obtained from the combination so that the information of $\{\Phi_i(x)\}_{i=1}^{10}$ can be restored. As a result, the number of cases for the attack identification is reduced to $2 \times \binom{10}{4} = 420$.

Regarding the criterion (11) for the local inspection for this simple example, we have $\delta^j(t) \equiv 0.01$, $\hat{\xi}^j_{I_j}(t) = \{y_i(t)\}_{i\in I_j}$, and $\Psi^j_{I_j}(\mathcal{X}) = \Phi_{I_j}(\mathcal{X})$, where $I_j \subset \{1, \ldots, 10\}$ for $j = 1$ and $I_j \subset \{11, \ldots, 20\}$ for $j = 2$, with $|I_j| = 6$. For each case of $|I_j| = 6$ with $j = 1$, it satisfies $\Phi(x) = \mathcal{O}_{I_j} \cdot [x_1 - \frac{1}{2}x_3^2, x_2]^\top$ with some matrix $\mathcal{O}_{I_j} \in \mathbb{R}^{6\times 2}$, and it means the set $\Phi_{I_j}(\mathcal{X})$ is included in the image space of $\mathcal{O}_{I_j}$. So we simply check if

$$\left\|\{y_i(t)\}_{i\in I_j} - \mathcal{O}_{I_j}\mathcal{O}^\dagger_{I_j} \cdot \{y_i(t)\}_{i\in I_j}\right\|_2 \le 0.01 \times \sqrt{6} \qquad (14)$$

where $\|\cdot\|_2$ is the Euclidean norm, and $\mathcal{O}^\dagger_{I_j}$ is the Moore–Penrose pseudo inverse of $\mathcal{O}_{I_j}$, so that the violation of (14) implies that (11) is violated. For the cases with $j = 2$, we use that $\Phi_{I_j}(\mathcal{X})$ is included in the image of $[1, 1, \ldots, 1]^\top \in \mathbb{R}^{6\times 1}$.

Fig. 2 shows the simulation results. Fig. 2(a) depicts the attack scenario; a square wave is injected in the first to the fourth sensors, from the time $t \geq 4$. Next, Fig. 2(b) shows the distance function $\|\{y_i(t)\}_{i\in I_1} - \mathcal{O}_{I_1}\mathcal{O}_{I_1}^{\dagger}\{y_i(t)\}_{i\in I_1}\|_2$ for monitoring the collection $\{y_i(t)\}_{i\in I_1}$ with its threshold, where $I_1 = \{1, 2, \ldots, 6\}$. It can be seen that the presence of attack is detected right after it is injected. And, Fig. 2(c) shows the same for the index set $I_1 = \{5, 6, \ldots, 10\}$. As it does not contain a corrupted measurement, the signal does not exceed the threshold. As soon as the attack is detected from a measurement collection, the monitor switches the index set and searches out a set of uncorrupted set of measurements, by inspecting the other combinations and comparing the corresponding distance and threshold. For the switching mechanism, we follow the method in [18]. With the set of identified un-orrupted measurements, the correct estimate of the state $x(t)$ is computed, using left-inverses of the functions. Finally, Fig. 2(d) shows that an accurate estimation for the state $x(t)$ is obtained, except at the moment $t = 4$ when the attack is detected and the attack identification is performed.

V. Conclusion

In this article, we have proposed an approach for reducing the complexity of resilient state estimation for uniformly observable systems, by performing the attack identification with respect to local groups of projected partial state estimates. We have introduced a notion of redundancy for a collection of nonlinear functions and suggested that attack identification is possible even when the state cannot be reconstructed and is not observable from a set of local measurements as long as they are redundant.

References

[1] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, pp. 135–148, 2015.

[2] S. Amin, A. A. Cárdenas, and S. S. Sastry, "Safe and secure networked control systems under denial-of-service attacks," in *Proc. Hybrid Syst., Computation Control*, 2009, pp. 31–45.

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

[4] Q. Zhu and T. Basar, "Game-theoretic methods for robustness, security, and resilience of cyber physical control systems: Games-in-games principle for optimal cross-layer resilient control systems," *IEEE Control Syst. Mag.*, vol. 35, no. 1, pp. 46–65, Feb. 2015.

[5] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Trans. Autom. Control*, vol. 58, no. 11, pp. 2715–2729, Nov. 2013.

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

[7] M. S. Chong, M. Wakaiki, and J. P. Hespanha, "Observability of linear systems under adversarial attacks," in *Proc. Amer. Control Conf.*, 2015, pp. 2439–2444.

[8] C. Lee, H. Shim, and Y. Eun, "On redundant observability: From security index to attack detection and resilient state estimation," *IEEE Trans. Autom. Control*, vol. 64, no. 2, pp. 775–782, Feb. 2019.

[9] Y. Jeong and Y. Eun, "A robust and resilient state estimation for linear systems," *IEEE Trans. Autom. Control*, vol. 67, no. 5, pp. 2626–2632, May 2022.

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

[11] F. Pasqualetti, F. Dörfler, and F. Bullo, "A divide-and-conquer approach to distributed attack identification," in *Proc. 54th IEEE Conf. Decis. Control*, 2015, pp. 5802–5807.

[12] J. Kim, J. G. Lee, C. Lee, H. Shim, and J. H. Seo, "Local identification of sensor attack and distributed resilient state estimation for linear systems," in *Proc. 57th IEEE Conf. Decis. Control*, 2018, pp. 2056–2061.

[13] H. Jeon, S. Aum, H. Shim, and Y. Eun, "Resilient state estimation for control systems using multiple observers and median operation," *Math. Problems Eng.*, vol. 2016, 2016, Art. no. 3750264.

[14] A. Mitra and S. Sundaram, "Secure distributed observers for a class of linear time invariant systems in the presence of Byzantine adversaries," in *Proc. 55th IEEE Conf. Decis. Control*, 2016, pp. 2709–2714.

[15] Y. Chen, S. Kar, and J. M. F. Moura, "Resilient distributed estimation: Sensor attacks," *IEEE Trans. Autom. Control*, vol. 64, no. 9, pp. 3772–3779, Sep. 2019.

[16] J. G. Lee, J. Kim, and H. Shim, "Fully distributed resilient state estimation based on distributed median solver," *IEEE Trans. Autom. Control*, vol. 65, no. 9, pp. 3935–3942, Sep. 2020.

[17] Y. Mao, A. Mitra, S. Sundaram, and P. Tabuada, "On the computational complexity of the secure state-reconstruction problem," *Automatica*, vol. 136, 2022, Art. no. 110083.

[18] 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 having redundant sensors," *IEEE Trans. Autom. Control*, vol. 64, no. 3, pp. 1162–1169, Mar. 2019.

[19] M. S. Chong, H. Sandberg, and J. P. Hespanha, "A secure state estimation algorithm for nonlinear systems under sensor attacks," in *Proc. 59th IEEE Conf. Decis. Control*, 2020, pp. 5743–5748.

[20] Y. Shoukry, P. Nuzzo, N. Bezzo, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada, "Secure state reconstruction in differentially flat systems under sensor attacks using satisfiability modulo theory solving," in *Proc. 54th IEEE Conf. Decis. Control*, 2015, pp. 3804–3809.

[21] H. Shim and A. Tanwani, "Hybrid-type observer design based on a sufficient condition for observability in switched nonlinear systems," Int. J. Robust Nonlinear Control., vol. 24, no. 6, pp. 1064–1089, 2014.

[22] J. P. Gauthier, H. Hammouri, and S. Othman, "A simple observer for nonlinear systems: Applications to bioreactors," *IEEE Trans. Autom. Control*, vol. 37, no. 6, pp. 875–880, Jun. 1992.

[23] J. T. Schwartz, *Nonlinear Functional Analysis*. New York, NY, USA: Gordon and Breach Science, 1969.