Title

SMT-Based Observer Design for Cyber-Physical Systems Under Sensor Attacks

Permalink

https://escholarship.org/uc/item/01r9n44v

ISBN

978-1-5090-1772-0

Authors

Shoukry, Yasser
Chong, Michelle
Wakaiki, Masashi
et al.

Publication Date

2016-04-01

DOI

10.1109/iccps.2016.7479119

Peer reviewed

SMT-Based Observer Design for Cyber-Physical Systems Under Sensor Attacks

Yasser Shoukry[†§] Michelle Chong[*] Masashi Wakaiki[**] Pierluigi Nuzzo[†]
Alberto L. Sangiovanni-Vincentelli[†] Sanjit A. Seshia[†] João P. Hespanha[**] Paulo Tabuada[§]
[†]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA
[§]Department of Electrical Engineering, University of California, Los Angeles, CA
[*]Department of Automatic Control, Lund University, Sweden
[**]Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA

***Abstract*—We introduce a scalable observer architecture to estimate the states of a discrete-time linear-time-invariant (LTI) system whose sensors can be manipulated by an attacker. Given the maximum number of attacked sensors, we build on previous results on necessary and sufficient conditions for state estimation, and propose a novel multi-modal Luenberger (MML) observer based on efficient Satisfiability Modulo Theory (SMT) solving. We present two techniques to reduce the complexity of the estimation problem. As a first strategy, instead of a bank of distinct observers, we use a family of filters sharing a single dynamical equation for the states, but different output equations, to generate estimates corresponding to different subsets of sensors. Such an architecture can reduce the memory usage of the observer from an exponential to a linear function of the number of sensors. We then develop an efficient SMT-based decision procedure that is able to reason about the estimates of the MML observer to detect at runtime which sets of sensors are attack-free, and use them to obtain a correct state estimate. We provide proofs of convergence for our algorithm and report simulation results to compare its runtime performance with alternative techniques. Our algorithm scales well for large systems (including up to 5000 sensors) for which many previously proposed algorithms are not implementable due to excessive memory and time requirements. Finally, we illustrate the effectiveness of our algorithm on the design of resilient power distribution systems.**

I. Introduction

Large and complex cyber-physical systems (CPSs) (e.g., power grids, water and gas distribution systems) are increasingly being deployed today as a promising response to key infrastructural and societal challenges, ranging from transportation, energy, security, to health-care. In these systems, sensors and cyber components (e.g., digital processors and networks) instrument the physical world to make it "smarter." However, cyber components are also the source of new, unprecedented vulnerabilities to malicious attacks. Striking examples of adversarial attacks include the Stuxnet virus targeting SCADA systems [1] as well as the injection of false data in power systems [2], or the non-invasive sensor spoofing attacks in automotive systems [3]. Independently of the nature of the attack, i.e., whether it originates from the cyber or physical portion of the system, it eventually results into corrupted sensor measurements. Using these measurements to control the operation of the CPS can finally lead to life-threatening situations.

A possible strategy to build *resilient cyber-physical systems* is to equip them with algorithms that are able to recognize the presence of an adversarial attack on their sensors and reconstruct the actual system state independently of the attack, by relying on the measurements collected from uncorrupted sensors. This is the approach advocated in this paper.

The problem of state reconstruction in the presence of attacks, also denoted as *secure state estimation*, has recently attracted considerable attention from the control community [4]–[12]. The first category of works [4], [7]–[10], [12] performs state estimation by analyzing the sensor information collected within a time window of finite length. This is, for instance, the case for the algorithms based on solving an L_0 optimization problem, proposed by Fawzi et al. [4] and Pajic et al. [7], the game-theoretic approach proposed by Mo et al. [9], and the Gramian-based observer by Chong et al. [6]. The second category of results in the literature focuses, instead, on designing recursive observers and filters [5], [6], [8], [11]. Observers and filters show a higher promise of scalability, as they are able to incorporate new information as it becomes available in real time. Moreover, they offer better performance in the presence of noise and modeling errors, since they make use of all the available measurements from the beginning of the observation window to the current time instant t. Observers would be, however, ineffective unless they are coupled with an efficient procedure that can quickly identify and disregard the malicious measurements.

The problem of reconstructing the state under sensor attacks is closely related to fault-tolerant state reconstruction. The robust Kalman filter, described in [13], and the robust principal component analysis (PCA) [14] are approaches to fault-tolerant state reconstruction closer to the results in this paper, at the technical level. In robust Kalman filtering the state estimate updates are obtained by solving a convex L_1 optimization problem that is robust to outliers. Similarly, a PCA robust to outliers is developed in [14]. However, no theoretical guarantees are known regarding the performance of these techniques in the presence of malicious attacks.

This work was partially sponsored by the NSF award 1136174 and CNS-1329650, by DARPA under agreement number FA8750-12-2-0247, by TerraSwarm, one of six centers of STARnet, a Semiconductor Research Corporation program sponsored by MARCO and DARPA, and by the NSF project ExCAPE: Expeditions in Computer Augmented Program Engineering (award 1138996 and 1139138). The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of NSF, DARPA, or the U.S. Government.

The complexity of CPSs is arguably the hardest challenge for the deployment of resilient and secure designs. Identifying which sensors are under attack is combinatorial in the number of sensors; while brute-force strategies show poor scalability [5], [6], [11], problem relaxations tend to weaken the security guarantees [8]. The situation is exacerbated by the "trillion devices" scenario posed by the Internet of Things. Devising scalable algorithmic solutions to secure CPS design is, therefore, highly desirable [15].

In this paper, we address the design of observers that can accurately reconstruct the state of a cyber-physical system under sensor attacks, while being suitable to be deployed in real-time and on large scale systems. At the heart of our approach is a Satisfiability Modulo Theory based algorithm which further improves on the scalability of our previous solutions [10], [16] while addressing both memory and runtime efficiency. In particular, the work in [10], [16] focuses mainly on the case where the state of the system is estimated from data collected within a finite window length. However, when bounded noise exists in the system, estimating the state using finite amount of data leads to poor performance in terms of the state estimation error [10]. Therefore, in this work, we propose designing a Luenberger-like observer that can incorporate new measurements as they become available. Thanks to the recursive nature of the Luenberger observer, noise is "averaged out" as new information become available. Our contributions can be summarized as follows:

- We propose, to the best of our knowledge, the first state estimation algorithm that combines the robustness of an observer to bounded noise with the efficiency of SMT-based detection of corrupted sensors;
- We present a novel observer architecture whose memory usage scales linearly with the number of system states and sensors;
- We demonstrate the scalability of our approach on large CPS examples, showing that it outperforms previously proposed techniques.

The rest of the paper is organized as follows. Section II presents the mathematical formulation of our problem while Section III introduces the overall observer architecture. The design is detailed in Sections IV and V. In Section VI we provide the theoretical analysis of our algorithm and its convergence guarantees. Numerical experiments showing the scalability of the proposed approach for large systems as well as a power system design example are discussed, respectively, in Section VII and Section VIII. Finally, Section IX concludes the paper.

II. Problem Formulation

A. Preliminaries

The symbols $\mathbb{N}, \mathbb{R}$ and $\mathbb{B}$ denote the sets of natural, real, and Boolean numbers, respectively. The symbols $\wedge$ and $\neg$ denote the logical AND and logical NOT operators, respectively. The support of a vector $x \in \mathbb{R}^n$, denoted by $\text{supp}(x)$, is the set of indices of the non-zero elements of x. If S is a set, $|S|$ is the cardinality of S. For a vector $x \in \mathbb{R}^n$, we denote by $\|x\|_2$ the 2-norm of x and by $\|M\|_2$ the induced 2-norm of a matrix $M \in \mathbb{R}^{m \times n}$. We also denote by $M_i \in \mathbb{R}^{1 \times n}$ the ith row of M. Finally, for the set $\Gamma \subseteq \{1, \ldots, m\}$, we denote by $M_\Gamma \in \mathbb{R}^{|\Gamma| \times n}$ the matrix obtained from M by removing all the rows except those indexed by Γ.

B. System and Attack Model

We consider a system under sensor attacks of the form:

$$x^{(t+1)} = Ax^{(t)} + Bu^{(t)} + \mu^{(t)}, \tag{II.1}$$

$$y_i^{(t)} = \begin{cases} C_i x^{(t)} + \psi_i^{(t)} & \text{if } i\text{th sensor is attack-free} \\ C_i x^{(t)} + a_i^{(t)} + \psi_i^{(t)} & \text{if } i\text{th sensor is under attack} \end{cases} \tag{II.2}$$

where $x^{(t)} \in \mathbb{R}^n$ is the system state at time $t \in \mathbb{N}$, $u^{(t)} \in \mathbb{R}^m$ is the system input, and $y_i^{(t)} \in \mathbb{R}$ is the observed output from the ith sensor where $i \in \{1, \ldots, p\}$. Matrices A, B, and $C_1, \ldots, C_p$ represent the system dynamics and have appropriate dimensions. An attacker can corrupt the sensor measurements y_i by either spoofing the sensor output or manipulating the data transmitted from the sensor to the controller. Independently of the nature of the attack, its effect can be described by the attack signal $a_i^{(t)}$. We do not assume bounds, statistical properties, or restrictions on the time evolution of the elements in $a_i^{(t)}$. We only assume that the attacker has access to a fixed subset of sensors of cardinality $s \leq \overline{s}$; whether a specific sensor in this subset is attacked or not may change with time. As shown in [8, Theorem III.2], [4, Proposition 2], and [6, Theorem 1][1], the upper bound on the maximum number of sensors under attack $\overline{s}$ is a characteristic of the system and can be computed *a priori* from the system parameters (the number of sensors p and the A and C_i matrices). We will elaborate on how this upper bound is exploited in the design of an observer in Section III. Finally, the vectors $\mu^{(t)} \in \mathbb{R}^n$ and $\psi^{(t)} = (\psi_1^{(t)}, \ldots, \psi_p^{(t)}) \in \mathbb{R}^p$ represent, respectively, the process (i.e., modeling) noise and the measurement noise, which we assume to be bounded, i.e., there exist bounds $\overline{\psi}_i$ and $\overline{\mu}$ such that:

$$\|\psi_i(t)\|_2 \leq \overline{\psi}_i, \quad \|\mu(t)\|_2 \leq \overline{\mu}, \qquad \forall t \in \mathbb{N}, \, \forall i \in \{1, \ldots, p\}.$$

C. Observer Design Problem

We assume that each sensor can be in one of two modes, i.e., either attack-free or under attack. The attacker has access to at most $\overline{s}$ sensors and the attacker can, at any point of time, decide to switch his attack signal on or off. Since at most $\overline{s}$ sensors can be corrupted, the system can be, at each time, in any of $\sum_{k=0}^{\overline{s}} \binom{p}{p-k}$ modes, corresponding to the specific sets of sensors being under attack.

Since the attacked sensors are unknown a priori, designing a secure Luenberger observer entails two main steps. We first need to detect the system mode by identifying the sensors which are under attack. Then, we can construct the state estimate from the attack-free sensors. We formally define our task as follows.

Problem II.1. ***(Secure Luenberger Observer Design)*** *Given the linear system under attack defined in* (II.1) *and* (II.2)*, construct an estimate $\hat{x}^{(t)}$ such that:*

$$\limsup_{t \to \infty} \left\| x^{(t)} - \hat{x}^{(t)} \right\|_2 \leq \rho\left(\overline{\psi}, \overline{\mu}\right)$$

[1] This result was derived for continuous-time LTI systems.

for some constant $\rho \in \mathbb{R}^+$ *which depends on the noise bounds* $\overline{\psi} = (\sum_{i=1}^{p} \overline{\psi}_i^2)^{\frac{1}{2}}$ *and* $\overline{\mu}$.

In other words, we are interested in an estimate $\hat{x}^{(t)}$ such that the norm of the state estimation error $\left\|x^{(t)} - \hat{x}^{(t)}\right\|_2$ converges to a ball centered at the origin and whose radius is just a function of the noise bound. In particular, in the noiseless case, the observer asymptotically converges to the actual system state, independently of the attack.

We observe that estimating the state of a system in the presence of sensor attacks is not always feasible, in general. To establish conditions under which such an estimation is indeed feasible, hence the secure Luenberger observer design problem can be solved, we resort to the notion of *s-sparse observability* for discrete-time systems [8] (or, similarly, the one of M-attack observability [6]) defined as follows.

Definition II.2. ***(s-Sparse Observable System)*** *The linear control system under attack defined by* (II.1) *and* (II.2) *is said to be* s-sparse observable *if for every set* $\Gamma \subseteq \{1, \ldots, p\}$ *with* $|\Gamma| = p - s$*, the pair* (A, C_Γ) *is observable.*

Informally, a system is s-sparse observable if it remains observable after eliminating any choice of s sensors. In the absence of sensor and process noise, the conditions under which the state can be estimated in spite of sensor attacks were studied in [4], [6], [8], where it is shown that, if $\overline{s}$ is the maximum number of corrupted sensors, $2\overline{s}$-sparse observability is necessary and sufficient for secure state estimation. Therefore, in what follows, we assume that the $2\overline{s}$-sparse observability condition always holds. Moreover, we introduce below a convenient and compact notation to describe the main results of this paper.

D. Notation

For a set of $\tau \in \mathbb{N}$ measurements (with $\tau \leq n$), we can arrange the outputs from the ith sensor at different time instants as follows:

$$\widetilde{Y}_i^{(t)} = \mathcal{O}_i x^{(t)} + E_i^{(t)} + F_i U^{(t)} + \Psi_i^{(t)}$$

where:

$$\widetilde{Y}_i^{(t)} = \begin{bmatrix} y_i^{(t)} \\ y_i^{(t+1)} \\ \vdots \\ y_i^{(t+\tau-1)} \end{bmatrix}, E_i^{(t)} = \begin{bmatrix} a_i^{(t)} \\ a_i^{(t+1)} \\ \vdots \\ a_i^{(t+\tau-1)} \end{bmatrix}, U^{(t)} = \begin{bmatrix} u^{(t)} \\ u^{(t+1)} \\ \vdots \\ u^{(t+\tau-1)} \end{bmatrix},$$

$$\Psi_i^{(t)} = \begin{bmatrix} 0 & 0 & \ldots & 0 & 0 \\ C_i & 0 & \ldots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ C_i A^{\tau-2} & C_i A^{\tau-3} & \ldots & C_i & 0 \end{bmatrix} \begin{bmatrix} \mu^{(t)} \\ \mu^{(t+1)} \\ \vdots \\ \mu^{(t+\tau-1)} \end{bmatrix} + \begin{bmatrix} \psi_i^{(t)} \\ \psi_i^{(t+1)} \\ \vdots \\ \psi_i^{(t+\tau-1)} \end{bmatrix},$$

$$F_i = \begin{bmatrix} 0 & 0 & \ldots & 0 & 0 \\ C_i B & 0 & \ldots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ C_i A^{\tau-2} B & C_i A^{\tau-3} B & \ldots & C_i B & 0 \end{bmatrix}, \mathcal{O}_i = \begin{bmatrix} C_i \\ C_i A \\ \vdots \\ C_i A^{\tau-1} \end{bmatrix}.$$

Since all the inputs in $U^{(t)}$ are known, we can further simplify the output equation as:

$$Y_i^{(t)} = \mathcal{O}_i x^{(t)} + E_i^{(t)} + \Psi_i^{(t)}, \tag{II.3}$$

where $Y_i^{(t)} = \widetilde{Y}_i^{(t)} - F_i U^{(t)}$. We also define:

$$Y^{(t)} = \begin{bmatrix} Y_1^{(t)} \\ \vdots \\ Y_p^{(t)} \end{bmatrix}, E^{(t)} = \begin{bmatrix} E_1^{(t)} \\ \vdots \\ E_p^{(t)} \end{bmatrix}, \Psi^{(t)} = \begin{bmatrix} \Psi_1^{(t)} \\ \vdots \\ \Psi_p^{(t)} \end{bmatrix}, \mathcal{O} = \begin{bmatrix} \mathcal{O}_1 \\ \vdots \\ \mathcal{O}_p \end{bmatrix} \tag{II.4}$$

to denote, respectively, the vector of outputs, attacks, noise, and observability matrices related to all sensors over the same time window of length τ. With some abuse of notation, Y_i, E_i and $\mathcal{O}_i$ are also used to denote the ith block of Y, E, and $\mathcal{O}$. Similarly, we denote with Y_Γ, E_Γ, Ψ_Γ, and $\mathcal{O}_\Gamma$ the blocks indexed by the elements in the set Γ.

Because of our assumptions on the system noise, there exists a uniform upper bound on its magnitude, denoted by $\overline{\Psi}_i \in \mathbb{R}^+$, i.e., the following inequality $\left\|\Psi_i^{(t)}\right\|_2 \leq \overline{\Psi}_i$ holds for all time $t \in \mathbb{N}$. With some abuse of notation, for the set $\Gamma \subseteq \{1, \ldots, p\}$ we denote with $\overline{\Psi}_\Gamma \in \mathbb{R}^+$ the bound on the noise for the set of sensors indexed by Γ, i.e.,

$$\left\|\Psi_\Gamma^{(t)}\right\|_2^2 = \sum_{i \in \Gamma} \left\|\Psi_i^{(t)}\right\|_2^2 \leq \sum_{i \in \Gamma} \overline{\Psi}_i^2 = \overline{\Psi}_\Gamma^2.$$

By the same abuse of notation, we drop the subscript Γ for the special case in which Γ is the set of all sensors, i.e., $\overline{\Psi} = \overline{\Psi}_\Gamma$ when $\Gamma = \{1, \ldots, p\}$.

III. Observer Architecture

In this section we detail the overall architecture of the proposed observer. For ease of presentation, we focus on the noiseless case (i.e., when $\mu^{(t)}$ and $\psi^{(t)}$ are equal to zero for all $t \in \mathbb{N}$). We extend our results to the noisy case in Section VI-C.

A. Exhaustive-Search-Based Observer

We recall that the states of the attacked system (II.1) can be estimated if and only if, for every subset Γ of $\{1, \ldots, p\}$ with at least $p - 2\overline{s}$ elements, the pair (A, C_Γ) is observable. We could exploit this result to construct an observer for every set Γ with $p - \overline{s}$ elements (which is greater than $p - 2\overline{s}$) as follows:

$$\begin{aligned} \hat{x}_\Gamma^{(t+1)} &= A\hat{x}_\Gamma^{(t)} + Bu^{(t)} + L_\Gamma(y_\Gamma^{(t)} - \hat{y}_\Gamma^{(t)}) \\ \hat{y}_\Gamma^{(t)} &= C_\Gamma \hat{x}_\Gamma^{(t)}, \end{aligned} \tag{III.1}$$

where $\hat{x}_\Gamma$ denotes the state estimate generated from the input u and output y_i, $i \in \Gamma$. Note that L_Γ can be chosen such that the eigenvalues of $A - L_\Gamma C_\Gamma$ are strictly within the unit disk since the pair (A, C_Γ) is observable. Clearly, since at least one subset of $p - \overline{s}$ sensors are attack-free, we expect the output error dynamics $\|Y_\Gamma - \mathcal{O}_\Gamma \hat{x}_\Gamma\|_2$ of at least one of these observers to decay. Our aim would then be to select the state estimate whose estimation error is no worse than the one generated by the attack-free sensors. We refer to this approach as the "exhaustive-search-based observer."

Such a brute force observer would have, however, two major disadvantages:

- **Memory complexity:** running $\binom{p}{p-\overline{s}}$ estimators as defined in (III.1), each of which produces an estimate $\hat{x}_{\Gamma_i} \in \mathbb{R}^n$, results in updating a vector of length $n\binom{p}{p-\overline{s}}$ at each sample time. This requires an amount of memory that is exponential in the number p of sensors.
- **Computational complexity:** after producing all the $\binom{p}{p-\overline{s}}$ estimates, they must still be analyzed to select the best state estimate based on some performance criterion. This analysis further adds to the computational complexity of the estimators.

Our main goal is, therefore, to *develop a new, scalable observer architecture* that overcomes the disadvantages above.

B. Multi Modal SMT-Based Observer

To reduce memory complexity, we propose to replace the bank of $\binom{p}{p-\overline{s}}$ observers with a single *multi-modal Luenberger (MML)-observer* which is still able to produce all the estimates of the naive observer. The MML-observer uses the input $u^{(t)}$ and measurements $y^{(t)}$ collected from all the sensors to update an extended state estimate $\hat{z}^{(t)}$. Whenever needed, the extended state estimate $\hat{z}^{(t)}$ can be transformed into a state estimate $\hat{x}_{\Gamma}^{(t)}$ that matches the data corresponding to the set of sensors indexed by Γ. We show in Section IV that the memory usage for the extended state estimate $\hat{z}^{(t)}$ scales linearly with p as opposed to the exponential scaling of the exhaustive-search-based observer (III.1).

Although the MML-observer reduces the memory requirements, we would still need to analyze all the $\binom{p}{p-\overline{s}}$ estimates to detect the attack-free sensors. We harness the underlying combinatorial nature of this problem by leveraging techniques from efficient satisfiability solving. To do so, we reformulate the estimation problem as a satisfiability problem as follows.

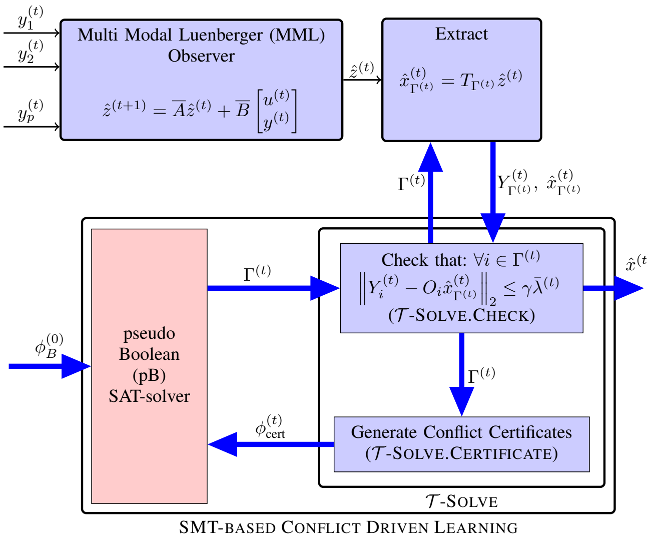
First, we recall that there is at least one set of sensors $\Gamma^{\star}$ with cardinality $|\Gamma^{\star}| \geq p - \overline{s}$ such that all the sensors indexed by this set are attack-free. Then, by Proposition A.2 in the appendix, we can conclude that the output error $\left\| Y_i^{(t)} - \mathcal{O}_i \hat{x}_{\Gamma^{\star}}^{(t)} \right\|_2^2$ decays exponentially over time, i.e.,

$$\left\| Y_i^{(t)} - \mathcal{O}_i \hat{x}_{\Gamma^{\star}}^{(t)} \right\|_2^2 \leq \gamma_i \bar{\lambda}^t, \qquad \forall i \in \Gamma^{\star}$$

where γ_i and $\bar{\lambda}$ are design parameters independent of the specific set $\Gamma^{\star}$. By defining a binary indicator variable $b_i \in \mathbb{B}$ such that $b_i = 0$ when the ith sensor is attack-free and $b_i = 1$ otherwise, the problem of constructing a secure Luenberger observer can be formulated as the search for an estimate $\eta^{(t)} = (\hat{x}^{(t)}, b^{(t)}) \in \mathbb{R}^n \times \mathbb{B}^p$ such that $\eta^{(t)} \models \phi^{(t)}$ $\forall t \in \mathbb{N}$, where $\phi^{(t)}$ is defined as:

$$\phi^{(t)} ::= \bigwedge_{i=1}^{p} \left(\neg b_i^{(t)} \Rightarrow \left\| Y_i^{(t)} - \mathcal{O}_i \hat{x}^{(t)} \right\|_2^2 \leq \gamma_i \bar{\lambda}^t \right) \wedge \left(\sum_{i=1}^{p} b_i^{(t)} \leq \overline{s} \right).$$

The first part of $\phi^{(t)}$ asks for an estimate $\hat{x}^{(t)}$ and an assignment for the attack indicator variables $b^{(t)} = (b_1^{(t)}, \ldots, b_p^{(t)})$ such that the discrepancy between the state estimate and the measured outputs decreases exponentially with time. The second clause requires, instead, that the number of attacked sensors be no greater than $\overline{s}$. As denoted by the time argument in $\phi^{(t)}$, at each time t, a new formula must be satisfied.



$y_1^{(t)}$
$y_2^{(t)}$
$y_p^{(t)}$
Multi Modal Luenberger (MML) Observer
$\hat{z}^{(t+1)} = \overline{A}\hat{z}^{(t)} + \overline{B}\begin{bmatrix} u^{(t)} \\ y^{(t)} \end{bmatrix}$
$\hat{z}^{(t)}$
Extract
$\hat{x}_{\Gamma^{(t)}}^{(t)} = T_{\Gamma^{(t)}} \hat{z}^{(t)}$
$\Gamma^{(t)}$
$Y_{\Gamma^{(t)}}^{(t)}, \hat{x}_{\Gamma^{(t)}}^{(t)}$
$\phi_B^{(0)}$
pseudo Boolean (pB) SAT-solver
$\Gamma^{(t)}$
Check that: $\forall i \in \Gamma^{(t)}$
$\left\| Y_i^{(t)} - O_i \hat{x}_{\Gamma^{(t)}}^{(t)} \right\|_2 \leq \gamma \bar{\lambda}^{(t)}$
($\mathcal{T}$-SOLVE.CHECK)
$\hat{x}^{(t)}$
$\Gamma^{(t)}$
$\phi_{\text{cert}}^{(t)}$
Generate Conflict Certificates ($\mathcal{T}$-SOLVE.CERTIFICATE)
$\mathcal{T}$-SOLVE
SMT-BASED CONFLICT DRIVEN LEARNING

Fig. 1. Architecture of the proposed observer. The observer consists of two main blocks: a Multi-Modal Luenberger (MML) observer that computes an extended estimate $\hat{z}^{(t)}$, which can then be transformed into the state estimate $\hat{x}_{\Gamma}^{(t)}$, for any set Γ, and an SMT-based conflict-driven learning algorithm that searches for the correct set of sensors Γ.

Following the *lazy* approach to Satisfiability Modulo Theory solving, our architecture uses a pseudo-Boolean[2] satisfiability (pB-SAT) solver to reason about possible assignments for the Boolean variables $b^{(t)}$. The pB-SAT leverages the David-Putnam-Logemann-Loveland (DPLL) algorithm [17] to suggest a set of sensors that are attack-free. The sensor choice is then passed to the MML-observer to transform the extended state estimate $\hat{z}^{(t)}$ into a corresponding state estimate $\hat{x}^{(t)}$, which is used to check the satisfiability of the formula $\phi^{(t)}$. If $\phi^{(t)}$ is not satisfied, the selected estimate (and the related sensor set) is incorrect. The observer will then implement a learning procedure to produce a succinct explanation for the infeasibility, i.e., to highlight which sensors are responsible for it. This conflict-driven learning mechanism is instrumental to speed-up the process of detecting and isolating the attacked sensors. The overall architecture is summarized in Figure 1. In the following sections, we give details for each of the two building blocks, i.e., the MML-Observer and the SMT-based conflict-driven learning.

IV. Multi-Modal Luenberger (MML) Observer

In this section we explain how to replace the bank of $\binom{p}{p-\overline{s}}$ observers (III.1) with only one observer which is able to produce the estimates computed by all those observers.

Step 1: We start by rewriting the observer (III.1) with initial state $\hat{x}_{\Gamma}^{(0)} = 0$ as:

$$\hat{x}_{\Gamma}^{(t+1)} = \widetilde{A}_{\Gamma} \hat{x}_{\Gamma}^{(t)} + \widetilde{B}_{\Gamma} \widetilde{u}^{(t)}, \quad \hat{x}_{\Gamma}^{(0)} = 0 \tag{IV.1}$$

$$\eta_{\Gamma}^{(t)} = \hat{x}_{\Gamma}^{(t)}. \tag{IV.2}$$

[2] A pseudo-Boolean constraint is a linear constraint over Boolean variables with integer coefficients

where $\widetilde{A}_\Gamma := A - L_\Gamma C_\Gamma$, $\widetilde{B}_\Gamma := [B \quad \mathcal{L}_\Gamma]$, $y^{(t)} := [y_1^{(t)} \ \ldots \ y_p^{(t)}]^\top$, and $\widetilde{u}^{(t)} := [u^{(t)} \ y^{(t)}]^\top$. The columns of $\mathcal{L}_\Gamma$ corresponding to the output y_i, $i \in \Gamma$, are equal to those of L_Γ and the other columns are zero.

Step 2: The next step is to choose the observer gain L_Γ such that $\widetilde{A}_\Gamma = A - L_\Gamma C_\Gamma$ has the same characteristic polynomial:

$$d(s) := s^n + a_1 s^{n-1} + \cdots + a_n \qquad \text{(IV.3)}$$

for all Γ. We note that this step can be always done thanks to the $2\overline{s}$-sparse observability property.

Step 3: The final step is to find a linear change of coordinates T_Γ which transforms the observer (IV.1) and (IV.2) into the following Controllable Canonical Form (CCF)[3]:

$$\hat{z}_\Gamma^{(t+1)} = \overline{A}\hat{z}_\Gamma^{(t)} + \overline{B}\widetilde{u}^{(t)}, \quad \hat{z}_\Gamma^{(0)} = 0 \qquad \text{(IV.4)}$$

$$\eta_\Gamma^{(t)} = \overline{C}_\Gamma \hat{z}_\Gamma^{(t)}, \qquad \text{(IV.5)}$$

where $\hat{z}_\Gamma^{(t+1)} \in \mathbb{R}^{nl}$, $l = m + p$, and:

$$T_\Gamma \overline{A} = \widetilde{A}_\Gamma T_\Gamma, \quad T_\Gamma \overline{B} = \widetilde{B}_\Gamma, \quad \overline{C}_\Gamma = T_\Gamma,$$

such that:

$$\overline{A} = \begin{bmatrix} -a_1 I_\ell & -a_2 I_\ell \ldots & -a_{n-1} I_\ell & -a_n I_\ell \\ I_\ell & 0_\ell \ \ldots & 0_\ell & 0_\ell \\ 0_\ell & I_\ell \ \ldots & 0_\ell & 0_\ell \\ \vdots & \vdots \ \ddots & \vdots & \vdots \\ 0_\ell & 0_\ell \ \ldots & I_\ell & 0_\ell \end{bmatrix}, \overline{B} = \begin{bmatrix} I_\ell \\ 0_\ell \\ \vdots \\ 0_\ell \\ 0_\ell \end{bmatrix}$$

where $a_1, \ldots, a_n$ are the coefficients of the characteristic polynomial (IV.3). Note that we dropped the subscript Γ from $\overline{A}$ and $\overline{B}$ in (IV.4). This follows from step 2 which ensures that all observers, for all sets Γ, have the same characteristic polynomial and hence they all have the same matrix $\overline{A}$ along with the fact that the definition of the matrix $\overline{B}$ does not depend on the set Γ.

To find such transformation, we use Proposition 2.3 in [18] on the realization of linear time-invariant systems to obtain the matrix T_Γ as:

$$\mathcal{R}_\Gamma = \begin{bmatrix} \widetilde{B}_\Gamma & \widetilde{A}_\Gamma \widetilde{B}_\Gamma & \ldots & \widetilde{A}_\Gamma^{n-1} \widetilde{B}_\Gamma \end{bmatrix}$$

$$\mathcal{R}' = \begin{bmatrix} I_\ell & a_1 I_\ell & a_2 I_\ell & \ldots & a_{n-1} I_\ell \\ 0_\ell & I_\ell & a_1 I_\ell & \ldots & a_{n-2} I_\ell \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0_\ell & \ldots & 0_\ell & I_\ell & a_1 I_\ell \\ 0_\ell & \ldots & 0_\ell & 0_\ell & I_\ell \end{bmatrix}$$

$$T_\Gamma = \mathcal{R}_\Gamma \mathcal{R}'. \qquad \text{(IV.6)}$$

Finally, by noticing that all observers are initialized to the same initial condition and they all share the same state update equation, we can rewrite all observers (IV.4) and (IV.5) as:

$$\hat{z}^{(t+1)} = \overline{A}\hat{z}^{(t)} + \overline{B}\widetilde{u}^{(t)}, \quad \hat{z}^{(0)} = 0 \qquad \text{(IV.7)}$$

$$\eta_\Gamma^{(t)} = \overline{C}_\Gamma \hat{z}^{(t)}, \qquad \text{(IV.8)}$$

$$\begin{tikzcd}
 & \hat{x}_\Gamma \in \mathbb{R}^n \arrow[rr, "\tilde{A}_\Gamma"] & & \hat{x}_\Gamma \in \mathbb{R}^n \arrow[dr, "I"] & \\
\tilde{u} \in \mathbb{R}^{m+p} \arrow[ur, "\tilde{B}_\Gamma"] \arrow[dr, "\bar{B}"'] & & & & \eta_\Gamma \in \mathbb{R}^n \\
 & \hat{z} \in \mathbb{R}^{n(m+p)} \arrow[uu, "T_\Gamma"'] \arrow[rr, "\bar{A}"'] & & \hat{z} \in \mathbb{R}^{n(m+p)} \arrow[uu, "T_\Gamma"'] \arrow[ur, "\bar{C}_\Gamma"'] &
\end{tikzcd}$$

Fig. 2. Commutative diagram of the observers (IV.1), (IV.2) and (IV.7), (IV.8).

where we dropped the dependency on the set Γ in (IV.7). In other words, the observer defined by (IV.7) and (IV.8) updates the intermediate state $\hat{z}^{(t)}$ based on all sensor measurements while still being able to transform $\hat{z}^{(t)}$ in the sensor dependent output $\hat{x}_\Gamma^{(t)}$ for any sensor set Γ.

Remark IV.1. *The size of $\hat{z}$ is $n(m+p)$, i.e., it grows linearly with the number of sensors p. This eliminates the need for a state vector that is exponential in p as in the case of the exhaustive-search-based observer.*

The discussion in this section is summarized in Figure 2 and in the following result.

Theorem IV.2. *Let the linear dynamical system defined by* (II.1) *and* (II.2) *be $2\overline{s}$-sparse observable. The observer defined by* (IV.7) *and* (IV.8) *generates the same output η_Γ as the original bank of observers* (IV.1), (IV.2) *for every input $\widetilde{u}$ and for any set of sensors Γ. Moreover, the size of $\hat{z}$ grows linearly with the number of outputs p.*

Proof. Since the linear system, defined by (II.1) and (II.2), is $2\overline{s}$-sparse observable, the pair (A, C_Γ) is observable for all Γ. Hence, we can always choose L_Γ such that every matrix $\widetilde{A}_\Gamma$ satisfies the same characteristic polynomial (IV.3). Routine calculations show that:

$$T_\Gamma \overline{A} = \widetilde{A}_\Gamma T_\Gamma, \quad T_\Gamma \overline{B} = \widetilde{B}_\Gamma, \quad \overline{C}_\Gamma = T_\Gamma, \qquad \text{(IV.9)}$$

for all Γ. Finally, the claim on the size of $\hat{z}$ follows from Remark IV.1. □

V. Satisfiability Modulo Theory (SMT)-Based Engine

As discussed in Section III-B, the SMT-based engine has three objectives: (i) hypothesize which sensors are attack-free and hence select the mode of the MML-observer; (ii) check whether the selected set of sensors is, indeed, attack-free; and (iii) generate conflicts (counterexamples) to speed up the search over the possible sensor sets. In this section, we give details on these tasks.

A. Hypothesizing the Attack-free Sensors

Searching for the attack-free sensors is combinatorial in nature. At each time instance t, we need to select a set $\Gamma^{(t)}$ containing at most $p - \overline{s}$ sensors. To do this, we use the indicator variable $b^{(t)} = (b_1^{(t)}, \ldots, b_p^{(t)}) \in \mathbb{B}^p$, where we use $b_i^{(t)} = 0$ to denote that sensor i is considered attack-free at

[3] The purpose of using the CCF is to obtain an observer system with a state equation that is independent of the set Γ (as shown in Step 3), which allows the multi-observer to be implemented as a family of systems that share a single state equation, but with different output equations.

Algorithm 1 $\mathcal{T}$-SOLVE.CERTIFICATE$\left(\mathcal{I}, \hat{x}_{\mathcal{I}}^{(t)}\right)$

```
Compute the residues for $i \in \mathcal{I}$
    $r_i := \left\| Y_i^{(t)} - \mathcal{O}_i \hat{x}_{\mathcal{I}}^{(t)} \right\|_2^2 - \gamma_i \bar{\lambda}^t$
Normalize the residues
    $r_i := r_i / \left\| \mathcal{O}_i \right\|_2^2$,
Sort the residues in ascending order
    $r\_sorted :=$ sortAscendingly($\{r_i | i \in \mathcal{I}\}$);
Choose sensor indices of $p - 2\bar{s}$ smallest residues
    $r\_min :=$ Index $(r\_sorted[1 : p - 2\bar{s}])$;
Search linearly for the certificate
status = UNSAT; counter = 1;   $\mathcal{I}' = \mathcal{I}$
while status == UNSAT do
    $\mathcal{I}' := \mathcal{I}' \setminus r\_min$[counter];
    $\hat{x}_{\mathcal{I}'}^{(t)} := T_{\mathcal{I}'} \hat{z}^{(t)}$
    if $\exists i \in \mathcal{I}'$  s.t.  $\left\| Y_i^{(t)} - \mathcal{O}_i \hat{x}_{\mathcal{I}'}^{(t)} \right\|_2^2 > \gamma_i \bar{\lambda}^t$  then
        $\phi_{\text{conf-cert}} := \sum_{i \in \mathcal{I}'} b_i \geq 1$;
        counter := counter + 1;
    else
        status := SAT;
return $\phi_{\text{conf-cert}}$
```

time t. At any point in time, the set of hypothesized attack-free sensors $\Gamma^{(t)}$ can then be extracted from $b^{(t)}$ using

$$\Gamma^{(t)} = \{1, \ldots, p\} \setminus \text{supp}\left(b^{(t)}\right).$$

We then ask the PB-SAT-SOLVE for an assignment over $b_i^{(0)}$ (at time $t = 0$) that satisfies the following constraint:

$$\phi_B^{(0)} := \sum_i^p b_i^{(0)} \leq \bar{s},$$

ensuring that at most $\bar{s}$ sensors are regarded as attacked. If the state estimate produced by the MML observer from $\Gamma^{(t)}$ does not satisfy

$$\left\| Y_i - \mathcal{O}_i \hat{x}_{\Gamma^{(t)}}^{(t)} \right\|_2^2 \leq \gamma_i \bar{\lambda}^t \qquad \forall i \in \Gamma^{(t)}, \qquad \text{(V.1)}$$

i.e., it is not bounded by the exponentially decaying envelope, then a new conflict clause is generated by $\mathcal{T}$-SOLVE.CERTIFICATE. This clause takes the form

$$\phi_{\text{cert}}^{(t)} = \sum_{i \in \mathcal{I}^{(t)}} b_i^{(t)} \geq 1$$

for some set $\mathcal{I}^{(t)} \subseteq \Gamma^{(t)}$, providing a certificate that at least one of the sensors indexed by the set $\mathcal{I}^{(t)}$ is under attack. Details on how to select $\mathcal{I}^{(t)}$ are given in the next subsection. The certificate generated at time t is then conjoined to the formula $\phi_B^{(t-1)}$ to create a new Boolean formula $\phi_B^{(t)} = \phi_B^{(t-1)} \wedge \phi_{\text{cert}}^{(t)}$ that needs to be satisfied by the pB-SAT solver, thus leading to a new candidate set of attack-free sensors.

B. Learning a Conflict Clause

Whenever the set of hypothesized attack-free sensors $\Gamma^{(t)}$ does not satisfy (V.1), we need to generate a compact Boolean certificate that explains the conflict. A *trivial certificate* could be easily generated as mentioned above:

$$\phi_{\text{triv-cert}}^{(t)} = \sum_{i \in \Gamma^{(t)}} b_i^{(t)} \geq 1, \qquad \text{(V.2)}$$

indicating that at least one of the sensors, which was initially assumed as attack-free (i.e., for which $b_i = 0$), is actually under attack; one of the b_i variables should then be set to one in the next assignment of the pB-SAT solver. However, $\phi_{\text{triv-cert}}$ does not provide much information, since it only excludes the current assignment from the search space, and can still lead to exponential execution time [10], [16]. In fact, the generated certificates heavily affect the overall execution time of an SMT solver. Smaller certificates prune the search space faster [10].

To find such a certificate, we need to search for a subset $\mathcal{I}^{(t)} \subseteq \Gamma^{(t)}$ whose elements cannot all be attack-free. To this end, we start by removing one sensor at a time from the original set and re-run the test (V.1) on the set of remaining sensors. This procedure repeats as long as the residual sensor set is conflicting. Finally, we generate the certificate:

$$\phi_{\text{conf-cert}}^{(t)} = \sum_{i \in \mathcal{I}^{(t)}} b_i^{(t)} \geq 1. \qquad \text{(V.3)}$$

Termination of the above procedure is guaranteed regardless of the order in which the sensors are chosen. In practice, different orderings may lead to different execution times. In Algorithm 1 we describe a heuristic based on the procedure in [10].

C. Strict versus Relaxed Conflict Clause Learning

Whenever a set of sensors does not pass the test (V.1), we need to learn a conflict clause and search for a new set of sensors. As the estimation algorithm must run in real time, a natural question is when to terminate the iterative process between hypothesizing a new set of sensors and learning a conflict clause. In particular, we propose two termination schemes, namely, *strict conflict clause learning* and *relaxed conflict clause learning*.

In the strict conflict clause learning, as shown in Algorithm 2, new sets of candidate attack-free sets are repetitively generated until we find a set $\Gamma^{(t)}$ that satisfies (V.1). This scheme maintains the following property invariant for all t:

$$\left\| Y_i - \mathcal{O}_i \hat{x}_{\Gamma^{(t)}}^{(t)} \right\|_2^2 \leq \gamma_i \bar{\lambda}^t, \qquad \forall i \in \Gamma^{(t)} \qquad \text{(V.4)}$$

On the other hand, in the relaxed conflict clause learning in Algorithm 3, a new set of candidate attack-free sensors and conflict clause are generated only once per time step. We call this scheme "relaxed" since it allows for (V.4) to be violated at some times.

Figure 3 emphasizes the difference between the two learning schemes using the IEEE-14 power bus example discussed in Section VIII. As shown a the top, the norm of the output estimation error $\|Y_i - \mathcal{O}_i x\|_2$ is guaranteed to be always below the decaying bound $\gamma_i \bar{\lambda}^t$. However, this occasionally comes at the cost of large execution time. The relaxed scheme allows, instead, the output estimation error to exceed the bound but achieves a constant execution time performance, which may be much better than the one of the strict observer. In the

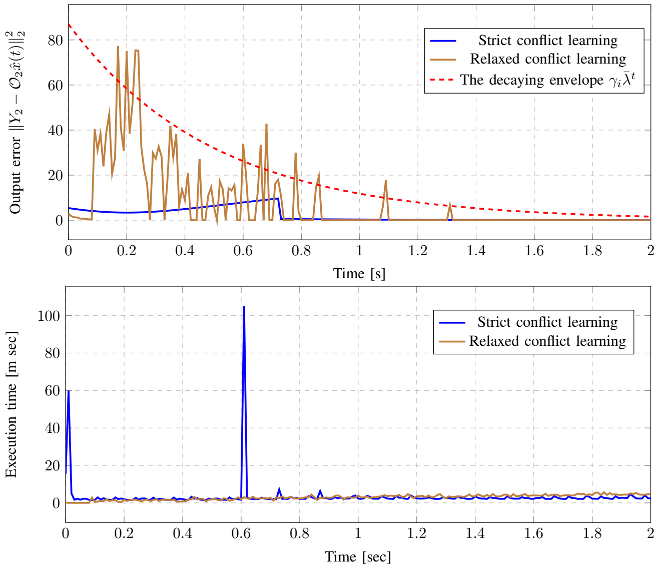


Fig. 3. Difference between strict and relaxed conflict clause learning. The top figure represents the output estimation error $\|Y_i - \mathcal{O}_i x\|_2$ as a function of the decaying upper bound $\gamma_i\bar{\lambda}^t$. The bottom figure shows, instead, the execution time of both the schemes.

Algorithm 2 Strict Secure Luenberger Observer

```
status := UNSAT,   $\phi_B^{(t)} := \phi_B^{(t-1)}$;
while status == UNSAT do
    $b^{(t)} :=$ pB-SAT-Solve$\left(\phi_B^{(t)}\right)$;
    $\Gamma^{(t)} = \{1, \ldots, p\} \setminus \text{supp}\left(b^{(t)}\right)$
    $\hat{x}_{\Gamma^{(t)}}^{(t)} := T_{\Gamma^{(t)}}\hat{z}^{(t)}$;
    if $\exists i \in \Gamma^{(t)}$ s.t. $\left\|Y_i - \mathcal{O}_i\hat{x}_{\Gamma^{(t)}}^{(t)}\right\|_2^2 > \gamma_i\bar{\lambda}^t$ then
        $\phi_{\text{cert}}^{(t)} := \mathcal{T}$-Solve.Certificate$\left(\Gamma^{(t)}, \hat{x}_{\Gamma^{(t)}}^{(t)}\right)$;
        $\phi_B^{(t)} := \phi_B^{(t)} \wedge \phi_{\text{cert}}^{(t)}$;
    else
        status := SAT;
$\hat{z}^{(t+1)} :=$ MML-Observer-Update$\left(\hat{z}^{(t)}, u^{(t)}, y^{(t)}\right)$;
```

next section, we discuss the theoretical guarantees of both the schemes.

VI. Convergence Analysis

We provide proofs of convergence for the two observer architectures introduced in Section V-C. We start by recalling the following result.

Proposition VI.1 (Proposition II.6 [10]). *Let the linear dynamical system defined in* (II.1) *and* (II.2) *be $2\bar{s}$-sparse observable. There exists a $\delta_{2\bar{s}} \in \mathbb{R}^+$ such that:*

$$0 < \delta_{2\bar{s}} \leq \lambda_{\min}\left\{\mathcal{O}_{\Gamma_{2\bar{s}}}^T \mathcal{O}_{\Gamma_{2\bar{s}}}\right\}$$

for any set $\Gamma_{2\bar{s}} \subset \{1, \ldots, p\}$ with $|\Gamma_{2\bar{s}}| \geq p - 2\bar{s}$.

In the rest of this paper, we refer to $\delta_{2\bar{s}}$ as the $2\bar{s}$-restricted eigenvalue of the system defined in (II.1) and (II.2). Using this notion of restricted eigenvalue, we can characterize the convergence of the proposed observers.

Algorithm 3 Relaxed Secure Luenberger Observer

```
$\phi_B^{(t)} := \phi_B^{(t-1)}$;
$b^{(t)} :=$ pB-SAT-Solve$\left(\phi_B^{(t)}\right)$;
$\Gamma^{(t)} = \{1, \ldots, p\} \setminus \text{supp}\left(b^{(t)}\right)$
$\hat{x}_{\Gamma^{(t)}}^{(t)} := T_{\Gamma^{(t)}}\hat{z}^{(t)}$;
if $\exists i \in \Gamma^{(t)}$ s.t. $\left\|Y_i - \mathcal{O}_i\hat{x}_{\Gamma^{(t)}}^{(t)}\right\|_2^2 > \gamma_i\bar{\lambda}^t$ then
    $\phi_{\text{cert}}^{(t)} := \mathcal{T}$-Solve.Certificate$\left(\Gamma^{(t)}, x^{(t)}\right)$;
    $\phi_B^{(t)} := \phi_B^{(t)} \wedge \phi_{\text{cert}}^{(t)}$;
$\hat{z}^{(t+1)} :=$ MML-Observer-Update$\left(\hat{z}^{(t)}, u^{(t)}, y^{(t)}\right)$;
```

A. Convergence of the Strict Secure Observer

Theorem VI.2. *Let the linear dynamical system defined in* (II.1) *and* (II.2) *be $2\bar{s}$-sparse observable. There exist constants $0 < \bar{\lambda} < 1$ and $\kappa \in \mathbb{R}^+$ such that the state estimation error $\left\|x^{(t)} - \hat{x}^{(t)}\right\|_2$ produced by the secure Luenberger Observer defined in Algorithm 2 at time t is bounded from above as*

$$\left\|x^{(t)} - \hat{x}^{(t)}\right\|_2^2 \leq \kappa\bar{\lambda}^t.$$

Moreover the state estimation error satisfies

$$\lim_{t\to\infty}\left\|x^{(t)} - \hat{x}^{(t)}\right\|_2^2 = 0$$

Proof. First, it follows from the $2\bar{s}$-sparse observability and Proposition A.2 in the Appendix that there exist observer gains such that the error dynamics of the attack-free observer is stable. In other words, there exist constants γ_i and $0 < \bar{\lambda} < 1$ and sets $\Gamma^{(t)}$ such that (V.1) holds. Hence, we conclude that the following inequality hold:

$$\left\|Y_i - \mathcal{O}_i\hat{x}_{\Gamma^{(t)}}^{(t)}\right\|_2^2 \leq \gamma_i\bar{\lambda}^t, \qquad \forall i \in \Gamma^{(t)} \;\; \forall t \in \mathbb{N} \tag{VI.1}$$

We now define $\mathcal{I}^{(t)} = \Gamma^{(t)} \setminus \text{supp}(E^{(t)})$, and $\mathcal{I}'^{(t)} = \text{supp}(E^{(t)})$. Then, the following holds:

$$\begin{aligned}\left\|Y_{\Gamma^{(t)}}^{(t)} - \mathcal{O}_{\Gamma^{(t)}}\hat{x}^{(t)}\right\|_2^2 &= \left\|\mathcal{O}_{\Gamma^{(t)}}(x^{(t)} - \hat{x}^{(t)}) + E_{\Gamma^{(t)}}^{(t)}\right\|_2^2 \\ &= \left\|\mathcal{O}_{\mathcal{I}^{(t)}}(x^{(t)} - \hat{x}^{(t)})\right\|_2^2 + \left\|\mathcal{O}_{\mathcal{I}'^{(t)}}(x^{(t)} - \hat{x}^{(t)}) + E_{\mathcal{I}'^{(t)}}^{(t)}\right\|_2^2 \\ &\overset{(a)}{\leq} \sum_{i\in\Gamma^{(t)}} \gamma_i\bar{\lambda}^t \quad \overset{(b)}{\leq} \gamma\bar{\lambda}^t \end{aligned} \tag{VI.2}$$

where inequality (a) follows from (VI.1), which ensures that Algorithm 2 always returns an estimate $\eta^{(t)} = (\hat{x}^{(t)}, b^{(t)})$ that satisfies $\phi^{(t)}$ for all t. Inequality (b) is, instead, obtained by setting $\gamma = \sum_{i=1}^{p}\gamma_i$.

We now observe that, because the attacker can corrupt at most $\bar{s}$ sensors, the cardinality of $\text{supp}(E^{(t)})$ is bounded by $\bar{s}$, i.e., $|\text{supp}(E^{(t)})| \leq \bar{s}$. Then, because the pB-SAT solver assumes at most that $\bar{s}$ sensors can be under attack, then the set $\Gamma^{(t)} = \{1, \ldots, p\} \setminus \text{supp}(b^{(t)})$ has a cardinality bounded by $p - \bar{s}$, i.e., $|\Gamma^{(t)}| \geq p - \bar{s}$. Using these two facts we conclude that the cardinality of $\mathcal{I}^{(t)} = \Gamma^{(t)} \setminus \text{supp}(E^{(t)})$ is bounded by $p - 2\overline{s}$, i.e. $|\mathcal{I}^{(t)}| \geq p - 2\overline{s}$. Hence, by using (VI.2), we conclude

$$\left\|\mathcal{O}_{\mathcal{I}^{(t)}}(x^{(t)} - \hat{x}^{(t)})\right\|_2^2 \leq \gamma\bar{\lambda}^t$$
$$\Rightarrow \delta_{2\overline{s}} \left\|x^{(t)} - \hat{x}^{(t)}\right\|_2^2 \leq \gamma\bar{\lambda}^t \Rightarrow \left\|x^{(t)} - \hat{x}^{(t)}\right\|_2^2 \leq \frac{\gamma}{\delta_{2\overline{s}}}\bar{\lambda}^t$$

and the result holds with $\kappa = \frac{\gamma}{\delta_{2\overline{s}}}$. □

B. Convergence of the Relaxed Secure Observer

Theorem VI.3. *Let the linear dynamical system defined in* (II.1) *and* (II.2) *be $2\overline{s}$-sparse observable. The state estimation error produced by the secure Luenberger Observer defined in Algorithm 3 satisfies*

$$\lim_{t\to\infty} \left\|x^{(t)} - \hat{x}^{(t)}\right\|_2^2 = 0$$

Proof. It follows from the $2\overline{s}$-sparse observability and Proposition A.2 in the Appendix that there exist observer gains such that error dynamics of the attack-free observer is stable. Therefore, there exist constants γ_i and $0 < \bar{\lambda} < 1$, and sets $\Gamma^{(t)}$ such that (V.1) holds. However, since Algorithm 3 may terminate before finding such $\Gamma^{(t)}$, the condition (V.1) may not always hold in the case of the relaxed observer.

On the other hand, we note that: (i) there are finitely many choices for the set $\Gamma^{(t)}$ since the indicator Boolean variable $b^{(t)}$ can take at most a finite number of values; (ii) whenever a certificate $\phi^{(t)}_{\text{conf-cert}}$ is learnt, at least one assignment of $b^{(t)}$ is ruled out of the search space. Therefore, there exists a time t' such that the following holds:

$$\left\|Y_i - \mathcal{O}_i\hat{x}^{(t)}_{\Gamma^{(t)}}\right\|_2^2 \leq \gamma_i\bar{\lambda}^t, \qquad \forall i \in \Gamma^{(t)} \;\; \forall t \geq t'. \quad \text{(VI.3)}$$

We can the follow the same line of reasoning as in the proof of Theorem VI.2 starting with time t' to conclude on the convergence of the relaxed scheme. □

C. Extension to the Noisy Case

Our convergence results can be extended to the noisy case by replacing (V.1) with the following inequality:

$$\left\|Y_i - \mathcal{O}_i\hat{x}^{(t)}_{\Gamma^{(t)}}\right\|_2^2 \leq \gamma_i\bar{\lambda}^t + \alpha_i \sum_{m=0}^{t-1} \bar{\lambda}^{t-m-1} + \beta_i \quad \forall i \in \Gamma^{(t)}, \qquad \text{(VI.4)}$$

where γ_i, α_i and β_i are defined in Proposition A.2 in the Appendix. It is crucial to note that all the constants γ_i, α_i and β_i do not depend on the knowledge of the attack-free set of sensors and can be computed offline. The following result is a direct extension of the argument used in Theorems VI.2 and VI.3. For this reason, we omit its proof.

Theorem VI.4. *Let the linear dynamical system defined in* (II.1) *and* (II.2) *be $2\overline{s}$-sparse observable. The state estimation error produced by the secure Luenberger Observers defined in Algorithm 2 and Algorithm 3, where the tests in line 6 and 5, respectively, are replaced with tests that check whether* (VI.4) *holds, satisfies*

$$\limsup_{t\to\infty} \left\|x^{(t)} - \hat{x}^{(t)}\right\|_2^2 \leq \rho(\overline{\psi}^2, \overline{\mu}^2),$$

where $\rho(\overline{\psi}^2, \overline{\mu}^2)$ is given by:

$$\rho(\overline{\psi}^2, \overline{\mu}^2) = \frac{1}{\delta_{2\overline{s}}}\left(\overline{\Psi} + \sqrt{\frac{\alpha}{1-\bar{\lambda}} + \beta}\right)^2, \alpha = \sum_{i=1}^{p} \alpha_i, \beta = \sum_{i=1}^{p} \beta_i,$$

and α_i, β_i and $\bar{\lambda}$ are as defined in Proposition A.2.

VII. Numerical Evaluation

As shown by Theorem IV.2, the memory requirement of our observer architecture grows linearly with the number of sensors and system states, a substantial improvement over traditional architectures. In fact, by assuming a number of sensors ranging from 500 to 5000 with 100 sensors being under attack, previously proposed observer- or filter-based algorithms [5], [11] would not be directly implementable, since they would require a bank of $\binom{500}{400} = 2.0417 \times 10^{107}$ observers or filters for a system with 500 sensors. Since each observer or filter updates a vector of n elements, this results into $2.0417 \times 10^{107} \times n$ memory units to implement a traditional architecture. Similarly, for a system with 5000 sensors, previous algorithms require $\binom{5000}{4900} \times n = 3.1201 \times 10^{211} \times n$ memory units. Our MML observer uses, instead, only $500 \times n$ memory units for a system with 500 sensors or $5000 \times n$ memory units for a system with 5000 sensors, which is a substantial decrease in terms of memory requirements.

Therefore, in this section, we focus on the evaluation of the time required to search for the attack-free sensors, and compare the performance of the proposed observer against the solution obtained by: (i) our previously proposed IMHOTEP-SMT algorithm, which uses data collected over a finite window length; (ii) the $L_1 \backslash L_r$ algorithm [4]. Since the search space increases exponentially with the number of sensors p, we generate a set of random systems (i.e., random matrices A, B and C) for an increasing number of sensors p, assuming that the number of states n is fixed. For each system, we generate a random initial state $x^{(0)}$ and input sequence $u^{(t)}$.

We run the experiments multiple times, by randomly selecting each time 100 sensors as being under attack. We report the worst case execution time in Figure 4. All the experiments were executed on an Intel Core i7 3.4-GHz processor with 8 GB of memory. The proposed observers are implemented in MATLAB while the pseudo-Boolean SAT solver is implemented using the SAT4J solver [19].

As shown in Figure 4, the relaxed conflict clause learning algorithm outperforms the other algorithms by at least an order of magnitude. Moreover, as the number of sensors increases, the gap between the relaxed conflict clause learning algorithm and other algorithm increases. The relaxed learning algorithm takes approximately 274 s, in the worst case, for a system with 5000 sensors, which shows that our approach is indeed suitable to be deployed on large scale systems. On the other hand, the worst case execution time of the strict learning algorithm is comparable to the one of the finite-window-length algorithm. This is mostly due to the fact that, in the worst case, the strict learning algorithm may end up with finding all the needed conflict clauses to reveal the attack-free set, after which no further learning is needed.

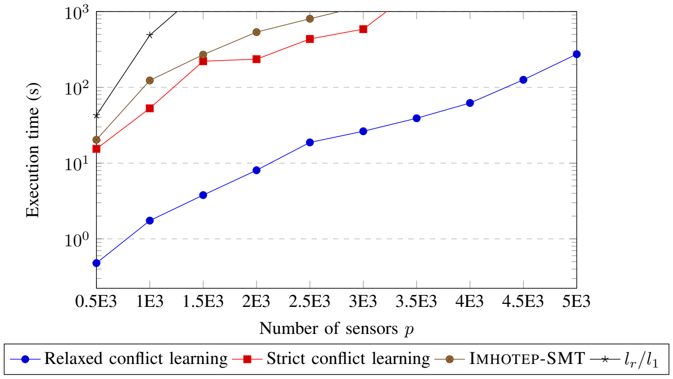


Fig. 4. Worst case execution time for different algorithms.

VIII. Case Study: Power Systems

Power systems are an important example of CPSs for which attacks have been recently documented [2]. In this paper, we consider the IEEE 14-bus power network shown in Figure 5 (left), composed of 5 synchronous generators and 14 buses. The state of each generator includes rotor angle and frequency. The overall system has 35 sensors: 14 sensors measure the real power injections at every bus, 20 sensors measure the real power flows along every branch, and one sensor measures the rotor angle of generator 1. The matrices A, B, and C modeling the power network are derived in [5]. While the IEEE 14-bus can model the power generation and distribution within a small geographical area, our objective is to test the performance of the proposed algorithms for large scale systems. To emulate the power generation and distribution over a larger geographical area, we instantiate the IEEE 14-bus system multiple times and connect these instances together, as shown in Figure 5 (right).

Since the scalability of our algorithm is evaluated in detail in the previous section, in this case study we focus on the performance of the proposed observer in terms of the estimation error $\left\|x^{(t)} - \hat{x}^{(t)}\right\|_2$ when the model of the system is imperfect (the process noise $\overline{\mu}$ is set to 0.5) and the sensors are noisy ($\overline{\psi}_i = 0.5 \; \forall i \in \{1, \ldots, p\}$) as the size of the power grid increases. In fact, a major concern in secure state estimation is the ability of an intruder to use the uncertainty in the model and the noise in sensors to mount its attack [7]. We compare the estimation performance of the proposed observer against the performance of the previously proposed IMHOTEP-SMT solver [10], [16], which uses sensor data collected over a finite window length. As shown in Table I, the proposed observer performs better in terms of estimation error, which is to be expected as the observer averages out the noise over time. This advantage becomes substantial as the size of the system (hence the norm bound $\overline{\psi}^2 = \sum_{i=1}^{p} \overline{\psi}_i$) increases.

IX. Conclusions

We addressed the problem of designing large-scale cyber-physical systems that are resilient to sensor attacks. We proposed a novel, scalable, multi-modal Luenberger (MML) observer that can isolate the sensors under attack and estimate the state of the underlying dynamics from the remaining sensors. Our architecture has a memory requirement that scales linearly with the system size and adopts an efficient SMT-based search algorithm to harness the computational complexity of identifying the set of attacked sensors. Numerical results show that the proposed observer outperforms other state-of-the-art algorithms in terms of memory requirements and computational complexity, and is suitable to be deployed on large scale systems. When applied to a power grid case study, the MML observer allows for a better estimation error in the presence of model uncertainties and sensor noise.

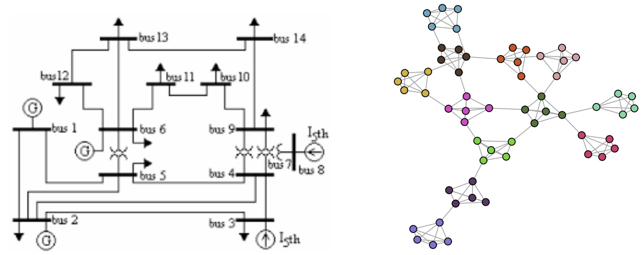


Fig. 5. The IEEE 14-bus power network (left) and twelve connected instances of the IEEE 14 bus (right). Each set of colored nodes represent the five generators in an IEEE 14 bus.

#Generators	#Sensors	#Attacks	SMT-based Observer	IMHOTEP-SMT
5	35	10	0.520	0.461
10	70	20	0.447	1.185
15	105	30	0.571	2.322
20	140	40	0.614	2.733
25	175	50	0.599	3.889
30	210	60	0.332	4.250
35	245	70	0.697	4.754
40	280	80	0.643	5.504
45	315	90	0.702	6.291
50	350	100	0.829	6.797
55	385	110	1.363	9.297

TABLE I
Norm of estimation error $\left\|x^{(t)} - \hat{x}^{(t)}\right\|_2$ (evaluated at the end of simulation time) in the power grid test case for a finite-window-length algorithm and the proposed observer.

References

[1] R. Langner, “Stuxnet: Dissecting a cyberwarfare weapon,” *IEEE Security and Privacy Magazine*, vol. 9, no. 3, pp. 49–51, 2011.

[2] Y. Liu, P. Ning, and M. K. Reiter, “False data injection attacks against state estimation in electric power grids,” in *Proceedings of the 16th ACM conference on Computer and communications security*, ser. CCS ’09. New York, NY, USA: ACM, 2009, pp. 21–32.

[3] Y. Shoukry, P. D. Martin, P. Tabuada, and M. B. Srivastava, “Non-invasive spoofing attacks for anti-lock braking systems,” in *Workshop on Cryptographic Hardware and Embedded Systems*, ser. G. Bertoni and J.-S. Coron (Eds.): CHES 2013, LNCS 8086. International Association for Cryptologic Research, 2013, pp. 55–72.

[4] 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, June 2014.

[5] F. Pasqualetti, F. Dorfler, and F. Bullo, “Attack detection and identification in cyber-physical systems,” *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, Nov 2013.

[6] M. Chong, M. Wakaiki, and J. Hespanha, “Observability of linear systems under adversarial attacks,” in *American Control Conference (ACC), 2015*, September 2014, pp. 2439–2444.

[7] M. Pajic, J. Weimer, N. Bezzo, P. Tabuada, O. Sokolsky, I. Lee, and G. Pappas, “Robustness of attack-resilient state estimators,” in *ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS)*, April 2014, pp. 163–174.

[8] Y. Shoukry and P. Tabuada, "Event-triggered state observers for sparse sensor noise/attacks," *IEEE Transactions on Automatic Control*, 2016. [Online]. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7308014

[9] Y. Mo and R. Murray, "Multi-dimensional state estimation in adversarial environment," in *34th Chinese Control Conference (CCC)*, July 2015, pp. 4761–4766.

[10] Y. Shoukry, P. Nuzzo, A. Puggelli, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada, "Secure State Estimation Under Sensor Attacks: A Satisfiability Modulo Theory Approach," *ArXiv e-prints*, Dec. 2014, [online] http://arxiv.org/abs/1412.4324.

[11] 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 *IEEE International Symposium on Information Theory (ISIT)*, June 2015.

[12] A. Teixeira, K. C. Sou, H. Sandberg, and K. Johansson, "Secure control systems: A quantitative risk management approach," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 24–45, Feb 2015.

[13] J. Mattingley and S. Boyd, "Real-time convex optimization in signal processing," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 50–61, May 2010.

[14] S. Farahmand, G. B. Giannakis, and D. Angelosante, "Doubly robust smoothing of dynamical processes via outlier sparsity constraints," *IEEE Trans. on Signal Processing*, vol. 59, no. 10, pp. 4529–4543, Oct. 2011.

[15] P. Nuzzo, A. Sangiovanni-Vincentelli, D. Bresolin, L. Geretti, and T. Villa, "A platform-based design methodology with contracts and related tools for the design of cyber-physical systems," *Proc. IEEE*, vol. 103, no. 11, Nov. 2015.

[16] Y. Shoukry, A. Puggelli, P. Nuzzo, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada, "Sound and complete state estimation for linear dynamical systems under sensor attack using satisfiability modulo theory solving," in *Proc. IEEE American Control Conference*, 2015, pp. 3818–3823.

[17] S. Malik and L. Zhang, "Boolean satisfiability from theoretical hardness to practical success," *Commun. ACM*, vol. 52, no. 8, pp. 76–82, Aug. 2009.

[18] B. De Schutter, "Minimal state-space realization in linear system theory: an overview," *Journal of Computational and Applied Mathematics*, vol. 121, no. 1–2, pp. 331–354, 2000.

[19] D. L. Berre and A. Parrain, "The Sat4j library, release 2.2," *Journal on Satisfiability, Boolean Modeling and Computation*, vol. 7, pp. 59–64, 2010.

[20] D. Liberzon, "Finite data-rate feedback stabilization of switched and hybrid linear systems," *Automatica*, vol. 50, no. 2, pp. 409–420, Feb. 2014.

Appendix

Proposition A.1. *Consider the following linear dynamical system:*

$$x^{(t+1)} = Ax^{(t)} + u^{(t)} \tag{A.1}$$

where A is a stable matrix. Then, there exist constants $0 < \overline{\lambda} < 1$ and $\kappa \in \mathbb{R}^+$ such that the function $V\left(x^{(t)}\right) = x^{(t)^T} P x^{(t)}$, with $P = P^T > 0$, satisfies

$$V\left(x^{(t+1)}\right) \leq \overline{\lambda} V\left(x^{(t)}\right) + \kappa \left\|u^{(t)}\right\|_2^2$$

and $A^T PA - P = -I$.

Proof. The proof follows a similar argument as the proof of Lemma 1 in [20] and hence is omitted for brevity. □

Using Proposition A.1, we can study the convergence of the Luenberger observer that uses the attack-free set of sensors. In particular, we can bound the norm of the estimation error as shown in the following result.

Proposition A.2. *Let $\Gamma^\star$ denote the set of $p - \overline{s}$ sensors which are attack-free for all times $t \in \mathbb{N}$. Let $\hat{x}^{(t)}_{\Gamma^\star}$ be the state estimated by the Luenberger observer that uses only the sensors indexed by $\Gamma^\star$. The following holds for any sensor $i \in \Gamma^\star$ and for all $t \in \mathbb{N}$:*

$$\left\|Y_i^{(t)} - \mathcal{O}_i \hat{x}^{(t)}_{\Gamma^\star}\right\|_2^2 \leq \gamma_i \bar{\lambda}^t + \alpha_i \sum_{m=0}^{t-1} \bar{\lambda}^{t-m-1} + \beta_i$$

where:

$$(A - L_\Gamma C_\Gamma)^T P_\Gamma (A - L_\Gamma C_\Gamma) - P_\Gamma = -I$$

$$\overline{\lambda}_P = \max_{\substack{\Gamma \subset \{1,\ldots,p\} \\ |\Gamma| = p - \overline{s}}} \lambda_{\max}\{P_\Gamma\}, \qquad \underline{\lambda}_P = \min_{\substack{\Gamma \subset \{1,\ldots,p\} \\ |\Gamma| = p - \overline{s}}} \lambda_{\min}\{P_\Gamma\}$$

$$\gamma_i = \frac{2 \left\|\mathcal{O}_i\right\|_2^2 \overline{\lambda}_P \left\|Y^{(0)}\right\|_2^2}{\underline{\lambda}_P \, \delta_{2\overline{s}}^2}, \qquad \overline{\overline{\Psi}} = \overline{\mu} + \overline{\psi} \max_{\substack{\Gamma \subset \{1,\ldots,p\} \\ |\Gamma| = p - \overline{s}}} \left\|L_\Gamma\right\|_2,$$

$$\alpha_i = \frac{2 \left\|\mathcal{O}_i\right\|_2 \kappa \overline{\overline{\Psi}}^2}{\underline{\lambda}_P}, \qquad \beta_i = 2\overline{\Psi}_i^2 \qquad \bar{\lambda} = 1 - \frac{1}{2\overline{\lambda}_P}$$

$$\kappa = \max_{\substack{\Gamma \subset \{1,\ldots,p\} \\ |\Gamma| = p - \overline{s}}} \left\|P_\Gamma (A - L_\Gamma C_\Gamma)\right\|_2^2 + \left\|P_\Gamma\right\|_2$$

Proof. We start by writing the error dynamics of the Luenberger observer corresponding to $\Gamma^\star$ as:

$$e_x^{(t+1)} = x^{(t+1)} - \hat{x}^{(t+1)}_{\Gamma^\star} = A_e e_x^{(t)} + \mu^{(t)} + L_{\Gamma^\star} \psi^{(t)}_{\Gamma^\star}, \tag{A.2}$$

where $e_x^{(t)} = x^{(t)} - \hat{x}^{(t)}_{\Gamma^\star}$ and $A_e = A - L_{\Gamma^\star} C_{\Gamma^\star}$. It follows from the $2\overline{s}$-sparse observability that the pair $(A, C_{\Gamma^\star})$ is observable and hence there exists an observer gain $L^\star$ such that the error dynamics (A.2) is stable. Therefore, by applying Proposition A.1 we conclude that there exists a function $V\left(e_x^{(t)}\right) = e_x^{(t)^T} P e_x^{(t)}$ with $P = P^T$ such that:

$$\begin{aligned} V\left(e_x^{(t+1)}\right) &\leq \bar{\lambda} V\left(e_x^{(t)}\right) + \kappa \left\|\mu^{(t)} + L_{\Gamma^\star} \psi^{(t)}_{\Gamma^\star}\right\|_2^2 \\ &\leq \bar{\lambda}^t V\left(e_x^{(0)}\right) + \kappa \overline{\overline{\Psi}}^2 \sum_{m=0}^{t-1} \bar{\lambda}^{t-m-1}, \end{aligned}$$

from which we conclude that:

$$\left\|e_x^{(t)}\right\|_2^2 \leq \frac{\lambda_{\max}\{P\}}{\lambda_{\min}\{P\}} \bar{\lambda}^t \left\|e^{(0)}\right\|_2^2 + \frac{\kappa \overline{\overline{\Psi}}^2}{\lambda_{\min}\{P\}} \sum_{m=0}^{t-1} \bar{\lambda}^{t-m-1}. \tag{A.3}$$

Accordingly, simple algebraic manipulations show that the following bound holds for each sensor $i \in \Gamma^\star$:

$$\begin{aligned} \left\|Y_i^{(t)} - \mathcal{O}_i \hat{x}^{(t)}\right\|_2^2 &= \left\|\mathcal{O}_i e_x^{(t)} + \Psi_i\right\|_2^2 \overset{(a)}{\leq} 2 \left\|\mathcal{O}_i e_x^{(t)}\right\|_2^2 + 2\overline{\Psi}_i^2 \\ &\overset{(b)}{\leq} \gamma_i \bar{\lambda}^t + \alpha_i \sum_{m=0}^{t-1} \bar{\lambda}^{t-m-1} + \beta_i, \end{aligned}$$

where (a) follows from Cauchy-Schwarz inequality along with the definition of $\overline{\Psi}_i$ and (b) follows from (A.3) and from the fact that the Luenberger observer is initialized at zero, and therefore the following inequality holds

$$\left\|Y^{(0)}_{\Gamma^\star} - \mathcal{O}_{\Gamma^\star} \hat{x}^{(0)}\right\|_2 = \left\|Y^{(0)}_{\Gamma^\star}\right\|_2 = \left\|\mathcal{O}_{\Gamma^\star} x^{(0)}\right\|_2 \geq \delta_{2\overline{s}} \left\|e_x^{(0)}\right\|_2,$$

combined with the fact that $\left\|Y^{(0)}\right\|_2 \geq \left\|Y^{(0)}_{\Gamma^\star}\right\|_2$. □