TU Delft

Delft University of Technology

Secure state and output estimation for accommodation of false data injection attacks in large-scale systems

Tabatabaei, Hanieh; Gallo, Alexander J.; Al-Dabbagh, Ahmad W.

DOI
10.1016/j.automatica.2025.112460

Publication date
2025

Document Version
Final published version

Published in
Automatica

Citation (APA)
Tabatabaei, H., Gallo, A. J., & Al-Dabbagh, A. W. (2025). Secure state and output estimation for accommodation of false data injection attacks in large-scale systems. *Automatica*, *180*, Article 112460. https://doi.org/10.1016/j.automatica.2025.112460

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Automatica 180 (2025) 112460

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Secure state and output estimation for accommodation of false data injection attacks in large-scale systems ☆

Hanieh Tabatabaei [a], Alexander J. Gallo [b], Ahmad W. Al-Dabbagh [a,*]

[a] School of Engineering, The University of British Columbia, 3333 University Way, Kelowna, V1V 1V7, Canada
[b] Delft Center for Systems and Control, Delft University of Technology, 2628 CD Delft, The Netherlands

ARTICLE INFO

Article history:
Received 11 January 2024
Received in revised form 5 May 2025
Accepted 5 June 2025

Keywords:
Attack accommodation
Attack detection and isolation
Covert cyber attacks
Cyber security
Distributed and networked systems
Large-scale systems
Secure control
Secure state and output estimation
Unknown input reconstruction

ABSTRACT

In this paper, we address the problem of secure estimation in networked systems, by focusing on false data injection attacks in large-scale systems, where malicious attackers alter the original transmitted data between subsystems. We propose a technique that ensures asymptotic secure estimation of the original transmitted data under two attack classes, termed stealthy and non-stealthy, while also providing detection and isolation capabilities. We give conditions under which asymptotic recovery of nominal performance is guaranteed, thus providing the large-scale system with resilience. Furthermore, we demonstrate the effectiveness of the proposed technique through a simulation-based case study.

© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

1. Introduction

Networked systems represent many societal applications, including energy and transportation. Such systems are vulnerable to cyber attacks, specifically on their control systems and transmitted data, which can cause severe consequences that range from disruption of nominal behavior, to reaching unsafe trajectories, and even to complete operational failures. Real-world examples of successful cyber attacks include Stuxnet and Industroyer, which respectively targeted an Iranian uranium enrichment facility and a Ukranian power grid (Hemsley & Fisher, 2018). Thus, ensuring cyber security in networked systems is critical.

In response to the growing threat of cyber attacks on networked systems, researchers have pursued various directions to address different attack types. According to Teixeira et al. (2015), based on the information and resources available to the attacker, the threat can be categorized based on (1) *model knowledge*, (2) *disclosure resources*, and (3) *disruption resources*. Accordingly, attack types are categorized as (1) disclosure attacks, where system information is accessed maliciously, e.g., eavesdropping transmitted data (e.g., Wang et al., 2022), (2) denial-of-service attacks, where system information is intercepted, e.g., jamming data transmission (e.g., Gupta et al., 2010), and (3) integrity/deception attacks, where system information is altered, e.g., stealthy manipulation of transmitted data, like replay attacks (e.g., Barboni et al., 2022; Mo & Sinopoli, 2009) and covert attacks (e.g., Ansari Rad & Al-Dabbagh, 2025).

Researchers have explored different detection and isolation techniques for integrity/deception attacks, for example, based on designing observers (e.g., Al-Dabbagh et al., 2020), modifying system structures (e.g., Teixeira et al., 2012), altering input behaviors and watermarking (e.g., Hoehn & Zhang, 2016; Yang et al., 2023), and deploying moving target algorithms (e.g., Griffioen et al., 2021). Although detection and isolation techniques may offer effective preliminary diagnosis of cyber attacks, they alone do not eliminate the effects of the attacks on system operation and performance. Accommodating the presence of the attacks goes beyond their detection and isolation, and requires corrective actions to compensate for the malicious data alteration.

Attack accommodation can be achieved through reconstruction of the original transmitted data and other system variables/signals, including states, inputs, and outputs in the presence

☆ The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Davide Martino Raimondo under the direction of Editor Alessandro Chiuso.

* Corresponding author.
E-mail addresses: hanieh.tabatabaei@ubc.ca (H. Tabatabaei), alexanderjulian.gallo@polimi.it (A.J. Gallo), ahmad.aldabbagh@ubc.ca (A.W. Al-Dabbagh).

https://doi.org/10.1016/j.automatica.2025.112460
0005-1098/© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

of cyber attacks, namely secure estimation.[1] This can be developed using techniques similar to those that address state and unknown input reconstruction, such as in fault accommodation and fault-tolerant control (e.g., Lan & Patton, 2021; Tan et al., 2023). However, developed techniques must account for the fundamental differences between cyber attacks and faults, as well as the distinct challenges introduced by the presence of attacks. For example, unlike faults, cyber attacks are malicious by design: attacks cannot be easily modeled as faults. Furthermore, while robust control methods can address unknown inputs, they tend to be conservative during nominal system operation, leading to degraded system performance. Thus, customized secure estimation techniques are needed for attack accommodation in networked systems.

Researchers have studied different problems related to secure estimation/control in networked systems, including the presence of denial-of-service attacks (e.g., Yan & Yang, 2023), deception attacks (e.g., Chang et al., 2018; Kazemi et al., 2022), and their combination (e.g., Mousavinejad et al., 2021). However, only a few techniques were developed for the specific class of networked systems, termed Large-Scale System (LSS), which integrates multiple subsystems through physical and cyber interconnections (i.e., couplings), including those that address deception attacks within a subsystem (e.g., Barboni & Parisini, 2020; Ma et al., 2024), while neglecting attacks between interconnected/coupled subsystems.

1.1. Objectives and contributions

In this paper, we consider an LSS composed of multiple subsystems with both physical and cyber couplings. The physical couplings stem from the partitioning of the LSS, while the cyber couplings, in the form of data transmitted over communication links between the subsystems, are deployed for control objectives and allow for distributed control architectures. We suppose the cyber couplings are vulnerable to two classes of integrity/deception attacks, in the form of False Data Injection (FDI) between subsystems; namely, non-stealthy FDI and stealthy FDI (SFDI) attacks. Unlike the former, SFDI attacks are designed and executed by intelligent/resourceful attackers to remain undetected by conventional observers (i.e., detectors) that are deployed locally in subsystems. These include covert attacks, a complex form of integrity/deception attack (Smith, 2015), hence posing additional challenges in secure estimation and attack accommodation.

To address the two classes of cyber attacks in LSSs, we present two distributed schemes to be implemented in each subsystem, in order to achieve the following:

- Accommodate non-stealthy FDI attacks, while achieving full reconstruction/secure estimation of received signals from neighboring subsystems;
- Partially or fully reconstruct local and received signals from neighboring subsystems under SFDI attacks.

We develop the two schemes by introducing customized observer designs that are integrated together, where the observers are constructed based on an Unknown Input Observer (UIO) described in Chen et al. (1996), and a Luenberger observer discussed in Luenberger (1964). In addition to secure estimation, we show how the reconstructed signals can be used for the detection and isolation of cyber attacks. We also demonstrate attack accommodation by utilizing the reconstructed output signals (i.e., sensor measurements) of neighboring subsystems in the local controller design, instead of the received output signals that may be altered by the attacker. Moreover, the first scheme can potentially reduce the number of required sensor measurements for the observer design, compared to other schemes in the literature (e.g., Gallo et al., 2020). We also evaluate the detectability properties of the second proposed scheme given a mismatch between the physical and cyber interconnection topologies, which is a condition that has not been thoroughly explored in the literature, while we provide a solution tailored for specific LSS structures and configurations.

The organization of this paper is as follows. In Section 2, we formulate the problem. Sections 3 and 4 present the proposed schemes for reconstruction/secure estimation and attack accommodation, including stability analysis for the LSS given the proposed schemes. Section 5 discusses the computational complexity of the proposed schemes. In Section 6, simulation results are presented to demonstrate the effectiveness of the proposed schemes, and in Section 7, concluding remarks are provided.

1.2. Notation and definitions

For a square matrix A, $\sigma(A)$ denotes the set of its eigenvalues. Given a set of vectors x_i, $i \in \mathcal{I}$, the operators $\underset{i\in\mathcal{I}}{\text{row}}(x_i)$ and $\underset{i\in\mathcal{I}}{\text{col}}(x_i)$ denote their row or column concatenation; the same operators are also used for matrices of compatible dimensions. For a vector x, its jth component is represented by $x_{[j]}$. For matrices $A_i, i \in \mathcal{I}$, the operator $\underset{i\in\mathcal{I}}{\text{diag}}(A_i)$ denotes the block-diagonal matrix composed of the matrices. For any matrix, the Hermitian concatenation operator is defined as $\text{He}(M) = M + M^\top$, while the operators $\text{Im}(\cdot)$, $\text{Rank}(\cdot)$, and $\ker(\cdot)$ denote its image or range, rank, and kernel. Suppose $A \in \mathbb{R}^{m\times n}$ is column-rank deficient, i.e., $r = \text{Rank}(A) < n$. One can always find two (non-unique) matrices $\bar{A} \in \mathbb{R}^{m\times r}$ and $\check{A} \in \mathbb{R}^{r\times n}$, such that $A = \bar{A}\check{A}$, where $\bar{A}$ is full column rank, and such that $\text{Im}(\bar{A}) \equiv \text{Im}(A)$. Note that, once $\bar{A}$ is fixed, $\check{A}$ is unique. For convenience, we often rewrite $Ax = \bar{A}\bar{x}$, where $\bar{x} = \check{A}x$. This operation also follows for sums of matrix multiplications: given a set of N matrices $A_i \in \mathbb{R}^{m\times n_i}$ and vectors $x_i \in \mathbb{R}^{n_i}$, $i \in \{1, \dots, N\}$, one can define $\bar{A}$ and $\bar{x}$ such that $\bar{A}\bar{x} = \sum_{i=1}^{N} A_i x_i$. In this paper, we imply the above procedure by introducing $\bar{A}$ as the "full column rank counterpart" of A. Given a set of indexed matrices $A_{ij}, i, j \in \mathcal{I}$ of appropriate dimensions, we use the notation $A = [A_{ij}]$ to indicate a block-diagonal matrix with A_{ij} in the ith block-row and jth block-column.

2. Problem formulation

2.1. Modeling large-scale systems

We consider an LSS as a distributed and networked system partitioned/divided into N subsystems. Each subsystem $\mathcal{S}_i, i \in \mathcal{N} = \{1, 2, \dots, N\}$ is composed of a physical process $\mathcal{P}_i$, a local controller $\mathcal{C}_i$, and an estimation module $\mathcal{E}_i$. The dynamics of $\mathcal{P}_i, \forall i \in \mathcal{N}$ are

$$\mathcal{P}_i : \begin{cases} \dot{x}_i = & A_i x_i + B_i u_i + B_{c_i}\bar{x}_i, \\ y_i = & C_i x_i, \end{cases} \tag{1}$$

where $x_i \in \mathbb{R}^{n_i}$, $y_i \in \mathbb{R}^{q_i}$, and $u_i \in \mathbb{R}^{p_i}$ are its state, output, and control input, respectively. The term $B_{c_i}\bar{x}_i$ models the physical couplings between $\mathcal{P}_i$ and a subset of the other subsystems, $\mathcal{N}_i^p \subset \mathcal{N}$. Specifically,

$$B_{c_i}\bar{x}_i = \sum_{k\in\mathcal{N}_i^p} A_{ik} x_k, \tag{2}$$

[1] Reconstruction/secure estimation refers to estimation while being decoupled from cyber attacks, similar to the notion used in the literature (e.g., Keijzer et al., 2023; Ma et al., 2024).

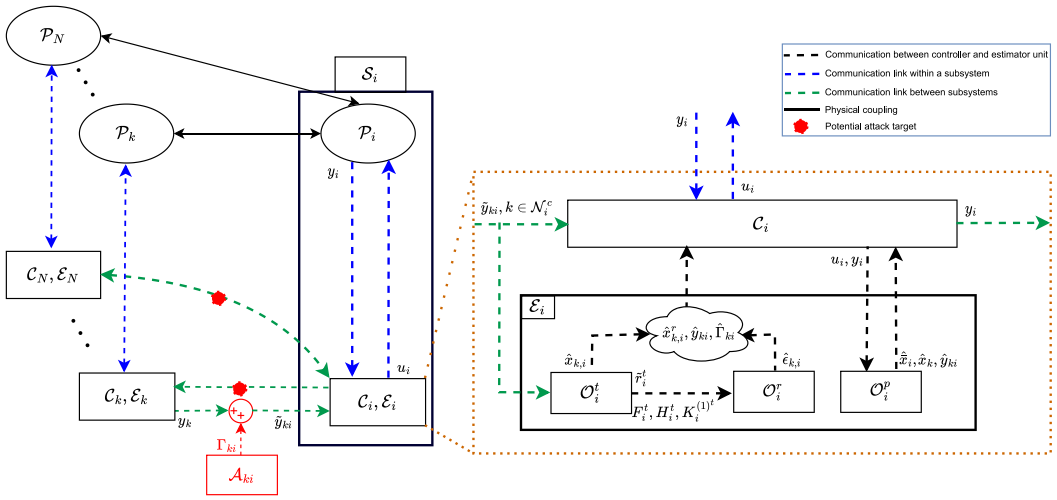

Fig. 1. A schematic of an LSS consisting of N subsystems: each subsystem $\mathcal{S}_i$ includes a process $\mathcal{P}_i$, a local controller $\mathcal{C}_i$, and an estimator unit $\mathcal{E}_i$, $i \in \{1, \dots, N\}$.

where A_{ik} models the physical effect of x_k on the dynamics of $\mathcal{P}_i$ (i.e., physical couplings); $B_{c_i}\bar{x}_i$ is then defined such that $B_{c_i} \in \mathbb{R}^{n_i \times n_{c_i}}$ is full column rank. Subsystems $\mathcal{S}_k$, $k \in \mathcal{N}_i^p$ are referred to as *physical* neighbors of $\mathcal{S}_i$.

Assumption 2.1. For each $\mathcal{P}_i$, $i \in \mathcal{N}$, the pair (A_i, C_i) is detectable and the pair (A_i, B_i) is controllable.

We suppose that u_i is the result of a distributed control architecture, requiring subsystems to exchange sensor measurements over communication links (i.e., cyber couplings), possibly compromised by attacks. To model the communication links, we introduce the set $\mathcal{N}_i^c \subset \mathcal{N}$, consisting of those subsystems with communication links to $\mathcal{S}_i$. Subsystems $\mathcal{S}_k$, $k \in \mathcal{N}_i^c$ are referred to as *cyber* neighbors of $\mathcal{S}_i$.[2] Note that $\mathcal{N}_i^c \neq \mathcal{N}_i^p$ in general.

We give a pictorial representation of the LSS in Fig. 1, illustrating the interconnections for subsystems, with both *physical* and *cyber* couplings. The red circles represent potential target points for attackers, i.e., the communication links between subsystems, where $\mathcal{A}_{ki}$ represents a malicious attacker capable of altering data transmitted from $\mathcal{S}_k$ to $\mathcal{S}_i$. Finally, $\mathcal{E}_i$ indicates an estimation module, including observers to reconstruct the original state and output of $\mathcal{S}_k$.

2.2. False data injection attacks

Following Gallo et al. (2020), Teixeira et al. (2015), false data injection attacks between subsystems are modeled by defining

$$\tilde{y}_{ki} = y_k + \Gamma_{ki}, \tag{3}$$

where $\tilde{y}_{ki}$ denotes the received output (i.e., altered sensor measurements) and Γ_{ki} denotes the malicious signal injected by $\mathcal{A}_{ki}$ to alter the original transmitted output y_k. The injection of Γ_{ki} allows the attacker to disrupt the nominal behavior of $\mathcal{S}_i$ through the action of u_i. Here, $\Gamma_{ki} = 0$ whenever no attack is present on the communication link ki from $\mathcal{S}_k$ and $\mathcal{S}_i$. Otherwise, it is selected by the attacker to achieve its malicious aim.

As we will demonstrate, and in agreement with literature on false data injection attacks (Teixeira et al., 2015), the injected data Γ_{ki} can be defined by a sufficiently sophisticated malicious attacker to avoid detection. We define the following two classes of attacks.

[2] In this paper, we are only interested in the in-neighbors of $\mathcal{S}_i$, i.e., those subsystems transmitting information to $\mathcal{S}_i$.

Definition 2.1 (*Non-stealthy FDI Attack*)**.** An FDI attack manipulating the output transmitted over the communication link ki, $i \in \mathcal{N}$, $k \in \mathcal{N}_i^c$ as in (3) is *non-stealthy* if the residual error $\tilde{y}_{ki} - \hat{y}_{ki}$ does not asymptotically converge to zero, for $\hat{y}_{ki}$ any estimated output using information of $\mathcal{P}_k$.

Definition 2.2 (*SFDI Attack*)**.** An FDI attack that manipulates the output transmitted over the communication link ki, $i \in \mathcal{N}$, $k \in \mathcal{N}_i^c$ as in (3) is *stealthy* if the residual error $\tilde{y}_{ki} - \hat{y}_{ki}$ asymptotically converges to zero, for $\hat{y}_{ki}$ any estimated output using information of $\mathcal{P}_k$.

Remark 2.1. In Definitions 2.1 and 2.2, we do not refer explicitly to the estimation mechanism to produce $\hat{y}_{ki}$, $i \in \mathcal{N}$, $k \in \mathcal{N}_i^c$, but rather only focus on the fact that it is an estimate which relies on information on $\mathcal{P}_k$. Indeed, it is the set of resources possessed by the attacker $\mathcal{A}_{ki}$ compared to those of the estimator unit $\mathcal{E}_i$ that restricts SFDI attacks. Furthermore, inspired by Smith (2015), we introduce covert FDI attacks, a sub-class of SFDI attacks.

Definition 2.3 (*Covert FDI Attack*)**.** An SFDI attack that alters the transmitted output over the communication link ki, $i \in \mathcal{N}$, $k \in \mathcal{N}_i^c$ is *covert* if its design is based on the attacker satisfying the following dynamics:

$$\mathcal{A}_{ki} : \begin{cases} \dot{\tilde{x}}_{ki} = \tilde{A}_k \tilde{x}_{ki} + \tilde{B}_k \alpha_{ki}, \\ \Gamma_{ki} = \tilde{C}_k \tilde{x}_{ki}, \end{cases} \tag{4}$$

where $\tilde{x}_{ki} \in \mathbb{R}^{n_k}$ is its state, α_{ki} is its driving input (i.e., an attacker's design parameter), and $\tilde{A}_k = A_k$, $\tilde{B}_k = \begin{bmatrix} B_k & \underset{j \in \mathcal{N}_k^p}{\text{row}}(A_{kj}) \end{bmatrix}$, and $\tilde{C}_k = C_k$.

2.3. Problem statement

Definition 2.4 (*Full State Reconstruction*)**.** The estimator module $\mathcal{E}_i$, $i \in \mathcal{N}$ achieves full reconstruction of x_k, $k \in \mathcal{N}_i^c$ if its observer can determine an estimate $\hat{x}_k$ such that $\|x_k - \hat{x}_k\| \to 0$, for any $\Gamma_{ki} \neq 0$.

Definition 2.5 (*Partial State Reconstruction*)**.** The estimator module $\mathcal{E}_i$, $i \in \mathcal{N}$ achieves partial reconstruction of x_k, $k \in \mathcal{N}_i^c$ if its observer can determine an estimate $\hat{x}_k$ such that $\|\Pi_{ki} x_k - \Pi_{ki} \hat{x}_k\| \to 0$, for some full row rank matrix Π_{ki} and any $\Gamma_{ki} \neq 0$.

In this paper, we address the following two objectives:

- For non-stealthy FDI attacks, both Γ_{ki} and y_k are reconstructed, and secure state estimation allows to obtain a secure estimate of x_k (Section 3).
- For SFDI attacks, partial or full reconstruction of x_k (and y_k) is achieved via estimation of $\bar{x}_i$ (Section 4).

For attack accommodation, we use the reconstructed y_k in the controller design, rather than the received $\tilde{y}_{ki}$, which may be altered by false data injection. By doing so, each subsystem can asymptotically restore its nominal operation, even under cyber attacks, providing the large-scale system with resilience.

3. Accommodation of non-stealthy FDI attacks

In this section, an FDI attack is considered as in Definition 2.1, which acts as an additive input, Γ_{ki}, to the transmitted output, formulated in (3). The main problem to investigate, then, is to derive asymptotic estimation of x_k, y_k, and Γ_{ki} within $\mathcal{E}_i$ for $i \in \mathcal{N}$, $k \in \mathcal{N}_i^c$.

A cascade of observers, including a UIO, $\mathcal{O}_i^t$, and a Luenberger-like observer, $\mathcal{O}_i^r$, is designed in each $\mathcal{E}_i$. The superscripts t and r denote that $\mathcal{O}_i^t$ is designed to reconstruct transmitted data potentially *targeted* by an attacker and that $\mathcal{O}_i^r$ uses the *residual* of $\mathcal{O}_i^t$. The estimates produced by $\mathcal{O}_i^t$ are designed to be independent from inputs to $\mathcal{P}_k$ unknown to $\mathcal{S}_i$ (i.e., the control input and some physical couplings of $\mathcal{P}_k$). It is shown that these estimates are affected by non-stealthy FDI attacks, and thus $\mathcal{O}_i^t$ on its own cannot address the first objective presented in Section 2.3. To address this, $\mathcal{O}_i^r$ is designed to asymptotically estimate x_k, y_k, and Γ_{ki} by utilizing the output estimation error, or residual, of $\mathcal{O}_i^t$. The reconstructed y_k can then be used by the local controller to asymptotically recover nominal operation.

Assumption 3.1. Each $\mathcal{S}_i$, $i \in \mathcal{N}$ has knowledge of the process model of its cyber neighbors $\mathcal{S}_k$, defined as in (1), i.e., A_k, B_k, $A_{kj}, j \in \mathcal{N}_k^p$, and C_k.

The knowledge of A_k, B_k, and C_k and the locally available sensor measurements $\tilde{y}_{ki}$ are used to reconstruct signals from the cyber neighbors, as assumed in Gallo et al. (2021, 2020), Teixeira et al. (2014). Moreover, knowledge of A_{kj} enhances the performance of the proposed scheme, as discussed in Remarks 3.2 and 4.1.

3.1. Aggregated system

To estimate $x_k, k \in \mathcal{N}_i^c, i \in \mathcal{N}$ in $\mathcal{E}_i$, a new partition of the LSS is introduced, including the dynamics of all cyber neighbors of $\mathcal{S}_i$. A unified state-space representation is presented, by which $\mathcal{O}_i^t$ and $\mathcal{O}_i^r$ are then designed, allowing for looser conditions on detectability than those found in the literature, e.g., in Gallo et al. (2020).

From the viewpoint of $\mathcal{S}_i$, the state dynamics of $\mathcal{P}_k$ remain unvaried compared to (1), but with the measurement defined by (3). Thus, aggregating all $\mathcal{P}_k$, $k \in \mathcal{N}_i^c$,

$$\begin{aligned} \dot{x}_i^t &= A_i^t x_i^t + B_{\varsigma_i}^t \varsigma_i^t, \\ \tilde{y}_i^t &= C_i^t x_i^t + \Gamma_i^t, \end{aligned} \tag{5}$$

where $x_i^t = \underset{k\in\mathcal{N}_i^c}{\text{col}}(x_k)$ is the state vector of all cyber neighbors of $\mathcal{S}_i$, $\tilde{y}_i^t = \underset{k\in\mathcal{N}_i^c}{\text{col}}(\tilde{y}_{ki})$ is the output vector of all received sensor measurements, and $\Gamma_i^t = \underset{k\in\mathcal{N}_i^c}{\text{col}}(\Gamma_{ki})$ is the aggregated vector of all false data injections. We define $A_i^t = [A_{mj}]$, $m, j \in \mathcal{N}_i^c$, where $A_{mm} = A_m$, and $A_{mj} \neq 0$ only if $j \in \mathcal{N}_m^p$. Furthermore, the output distribution matrix is defined as $C_i^t = \underset{k\in\mathcal{N}_i^c}{\text{diag}}(C_k)$. For each cyber neighbor $\mathcal{S}_k$, a full row rank matrix Φ_{ki} is defined such that $x_k = \Phi_{ki} x_i^t$. Similarly, a full row rank matrix Φ'_{ki} is defined such that $\Gamma_{ki} = \Phi'_{ki}\Gamma_i^t$ and $\tilde{y}_{ki} = \Phi'_{ki}\tilde{y}_i^t$.

The term $B_{\varsigma_i}^t \varsigma_i^t$ models the remaining state dynamics of $\mathcal{P}_k$, including the control inputs and the remaining physical couplings between $\mathcal{S}_k$ and its physical neighbors. On one hand, $\mathcal{S}_i$ does not receive any information regarding u_k, yet on the other hand, not all physical neighbors of $\mathcal{S}_k$ are necessarily cyber neighbors of $\mathcal{S}_i$, and therefore their impact is not captured by $A_i^t x_i^t$. The aggregated vector of these inputs to the dynamics of $\mathcal{P}_k$, which are unknown to $\mathcal{E}_i$, is captured by $B_{\varsigma_i}^t \varsigma_i^t$. To formally define ς_i^t, some preliminary definitions are given. The set comprising physical neighbors of $\mathcal{S}_k$ that are not cyber neighbors of $\mathcal{S}_i$ is denoted as $\mathcal{G}_{k,i} = \mathcal{N}_k^p \backslash (\mathcal{N}_k^p \cap \mathcal{N}_i^c)$, where the subscript k,i denotes that the set is a subset of $\mathcal{N}_k^p$ while its definition depends on $\mathcal{N}_i^c$. The interconnection inputs of these subsystems is reformulated as $\sum_{j\in\mathcal{G}_{k,i}} A_{kj} x_j = B_{c_{k,i}} \bar{x}_{k,i}$, where $\bar{x}_{k,i} \in \mathbb{R}^{n_{c_{k,i}}}$. Then, $B_{\varsigma_{k,i}}$ is defined to be the full column rank counterpart of $[B_k \;\; B_{c_{k,i}}]$, with $\varsigma_{k,i}$ such that $B_{\varsigma_{k,i}}\varsigma_{k,i} = B_k u_k + B_{c_{k,i}}\bar{x}_{k,i}$. Finally, $B_{\varsigma_i}^t = \underset{k\in\mathcal{N}_i^c}{\text{diag}}(B_{\varsigma_{k,i}})$ and $\varsigma_i^t = \underset{k\in\mathcal{N}_i^c}{\text{col}}(\varsigma_{k,i})$.

Remark 3.1. Each $\mathcal{E}_i$ has access to $\tilde{y}_{ki}, k \in \mathcal{N}_i^c$, received online from cyber neighbors, as well as knowledge of A_k, B_k, $A_{kj}, j \in \mathcal{N}_k^p$, and C_k according to Assumption 3.1. However, the vectors x_i^t, ς_i^t, and Γ_i^t are unknown to $\mathcal{E}_i$. Thus, $\mathcal{O}_i^t$ is designed such that its estimate of x_i^t is decoupled from ς_i^t, whilst Γ_i^t has an effect on the estimate.

3.2. Design and analysis of $\mathcal{O}_i^t$

Following the design procedure in Gallo et al. (2020), to estimate x_i^t while decoupling the unknown input ς_i^t,

$$\mathcal{O}_i^t : \begin{cases} \dot{z}_i^t = F_i^t z_i^t + K_i^t \tilde{y}_i^t, \\ \hat{x}_i^t = z_i^t + H_i^t \tilde{y}_i^t, \\ \hat{y}_i^t = C_i^t \hat{x}_i^t, \end{cases} \tag{6}$$

where $z_i^t \in \mathbb{R}^{n_i^t}$ is its state, and $\hat{x}_i^t \in \mathbb{R}^{n_i^t}$ and $\hat{y}_i^t \in \mathbb{R}^{q_i^t}$ are the estimated state and output, respectively, with $n_i^t = \sum_{k\in\mathcal{N}_i^c} n_k$ and $q_i^t = \sum_{k\in\mathcal{N}_i^c} q_k$. Its matrices must satisfy

$$\left(H_i^t C_i^t - I\right) B_{\varsigma_i}^t = 0, \tag{7a}$$

$$T_i^t = I - H_i^t C_i^t, \tag{7b}$$

$$F_i^t = T_i^t A_i^t - K_i^{(1)^t} C_i^t, \tag{7c}$$

$$K_i^{(2)^t} = F_i^t H_i^t, \tag{7d}$$

$$K_i^t = K_i^{(1)^t} + K_i^{(2)^t}, \tag{7e}$$

where the matrix H_i^t is defined such that the unknown input ς_i^t is decoupled. The matrix $K_i^{(1)^t}$ is desinged to ensure that F_i^t is Hurwitz. For each cyber neighbor $\mathcal{S}_k, k \in \mathcal{N}_i^c$, we define $\hat{x}_{k,i} = \Phi_{ki}\hat{x}_i^t$.

Assumption 3.2. For each $\mathcal{P}_k$ with $k \in \mathcal{N}_i^c, i \in \mathcal{N}$ and the model as (1), the matrix $\begin{bmatrix} A_k & B_{\varsigma_{k,i}} \\ C_k & 0 \end{bmatrix}$ is full column rank, and $\text{Rank}\left(C_k B_{\varsigma_{k,i}}\right) = \text{Rank}\left(B_{\varsigma_{k,i}}\right)$.

Lemma 3.1. *Assumptions 2.1 and 3.2 are sufficient to ensure the stability of $\mathcal{O}_i^t$, designed according to (7).*

The proof follows from Chen et al. (1996, Thm. 1), noting that the aggregated system (5) maintains the properties of each $\mathcal{S}_k$, described in Assumptions 2.1 and 3.2.

Remark 3.2. The condition $\text{Rank}(C_k B_{\mathcal{S}k,i}) = \text{Rank}(B_{\mathcal{S}k,i})$ in Assumption 3.2 necessitates that $q_i \geq p_i + n_{c_{k,i}}$, where $p_i + n_{c_{k,i}} = p_i + n_{c_k} - n'_{c_{k,i}} \leq p_i + n_{c_k}$. If individual observers were designed for each cyber neighbor of $\mathcal{S}_i$, the number of required sensors would need to satisfy $q_i \geq p_i + n_{c_k}$. However, by leveraging the physical couplings among cyber neighbors, $\mathcal{O}_i^t$ can potentially reduce the required number of sensors, specifically by $n'_{c_{k,i}}$, where $\sum_{j \in \mathcal{G}'_{k,i}} A_{kj} x_j = B'_{c_{k,i}} \bar{x}'_{k,i}$ with $\mathcal{G}'_{k,i} = \mathcal{N}_k^p \cap \mathcal{N}_i^c$, $\bar{x}'_{k,i} \in \mathbb{R}^{n'_{c_{k,i}}}$, and $B'_{c_{k,i}}$ is defined to be the full column rank counterpart of $\underset{j \in \mathcal{G}'_{k,i}}{\text{row}}(A_{kj})$.

To analyze $\mathcal{O}_i^t$, we define the residual $\tilde{r}_i^t = \tilde{y}_i^t - \hat{y}_i^t = C_i^t \epsilon_i^t + \Gamma_i^t$, with $\epsilon_i^t = x_i^t - \hat{x}_i^t$. Thus, applying (7a)–(7e), we obtain

$$\begin{aligned} \dot{\epsilon}_i^t &= F_i^t \epsilon_i^t - K_i^{(1)^t} \Gamma_i^t - H_i^t \dot{\Gamma}_i^t, \\ \tilde{r}_i^t &= C_i^t \epsilon_i^t + \Gamma_i^t. \end{aligned} \tag{8}$$

Under *nominal* system operation when $\Gamma_i^t = 0$, the estimation error dynamics are $\dot{\epsilon}_i^t = F_i^t \epsilon_i^t$, thereby $\epsilon_i^t \to 0$. However, this is not the case in the presence of cyber attacks, as $\epsilon_i^t \nrightarrow 0$, and accordingly $\hat{x}_i^t = x_i^t + \epsilon_i^t \nrightarrow x_i^t$.

Therefore, $\mathcal{O}_i^t$ is not sufficient to provide secure estimation under non-stealthy FDI attacks. To address this, $\mathcal{O}_i^t$ is combined with another observer, $\mathcal{O}_i^r$. This cascade of observers achieves two objectives: (1) the reconstruction of the unknown input Γ_i^t, thus obtaining secure estimate of y_i^t, and (2) the estimation of ϵ_i^t to securely estimate x_i^t, as addressed in the next subsection.

3.3. Design and analysis of $\mathcal{O}_i^r$

To implement a Luenberger-like observer for (8), while considering ϵ_i^t as the states, and Γ_i^t and $\dot{\Gamma}_i^t$ as unknown inputs, we employ an approach similar to those in the literature for unknown inputs reconstruction, such as Hou and Patton (1998), Lan and Patton (2021), Tan et al. (2023). First, the following is introduced:

$$\Psi_i^t : \begin{cases} \dot{\psi}_i^r = A_i^r \psi_i^r + E_i^r \underline{\Gamma}_i, \\ \tilde{r}_i^t = C_i^r \psi_i^r, \end{cases} \tag{9}$$

where $\psi_i^r = \begin{bmatrix} \epsilon_i^{t\top} & \dot{\Gamma}_i^{t\top} & \Gamma_i^{t\top} \end{bmatrix}^\top$ and $\underline{\Gamma}_i = \ddot{\Gamma}_i^t$. In addition,

$$\begin{aligned} A_i^r &= \begin{bmatrix} F_i^t & -H_i^t & -K_i^{(1)^t} \\ 0 & 0 & 0 \\ 0 & I & 0 \end{bmatrix}, \quad E_i^r = \begin{bmatrix} 0 \\ I \\ 0 \end{bmatrix}, \\ C_i^r &= \begin{bmatrix} C_i^t & 0 & I \end{bmatrix}. \end{aligned} \tag{10}$$

Lemma 3.2. *Under Assumptions 2.1 and 3.2, (A_i^r, C_i^r) is undetectable, with unobservable eigenvalues at 0.*

Proof. Let us start by proving that

$$\text{Rank}\left(\begin{bmatrix} sI - A_i^r \\ C_i^r \end{bmatrix}\right) = n_i^t + 2q_i^t, \ \forall s \in \mathbb{C}^+.$$

Given the definition of A_i^r and C_i^r, this is equivalent to the detectability of (F_i^t, C_i^t). Thus, given Lemma 3.1, Assumptions 2.1 and 3.2 are sufficient to prove this condition is satisfied.

Let us now show that (A_i^r, C_i^r) has unobservable eigenvalues at 0. By substituting $s = 0$, we see that

$$\text{Rank}\left(\begin{bmatrix} sI - A_i^r \\ C_i^r \end{bmatrix}\right) = q_i^t + \text{Rank}(\Sigma_i^r),$$

where $\Sigma_i^r = \begin{bmatrix} F_i^t & -K_i^{(1)^t} \\ C_i^t & I \end{bmatrix}$. Given its definition, $\text{Rank}(\Sigma_i^r) = n_i^t + q_i^t - \text{Null}(T_i^t A_i^t) < n_i^t + q_i^t$, where $\text{Null}(T_i^t A_i^t)$ indicates the nullity of $T_i^t A_i^t$. Therefore, the pair (A_i^r, C_i^r) is undetectable, with $\text{Null}(T_i^t A_i^t)$ unobservable eigenvalues of A_i^r at 0. ■

To estimate detectable components of ψ_i^r, a Luenberger-like observer for (9) is designed as

$$\mathcal{O}_i^r : \begin{cases} \dot{\hat{\psi}}_i^r = A_i^r \hat{\psi}_i^r + L_i^r \left(\tilde{r}_i^t - \hat{\tilde{r}}_i^t\right), \\ \hat{\tilde{r}}_i^t = C_i^r \hat{\psi}_i^r, \end{cases} \tag{11}$$

where $\hat{\psi}_i^r \in \mathbb{R}^{n_i^t + 2q_i^t}$ and $\hat{\tilde{r}}_i^t \in \mathbb{R}^{q_i^t}$ are the estimated state and residual, respectively, and with L_i^r such that $F_i^r = A_i^r - L_i^r C_i^r$ is stable, except for the unobservable eigenvalues of A_i^r at 0. Thus, an estimate of ϵ_i^t can be retrieved from $\hat{\psi}_i^r$, and an estimate of the estimation error $\epsilon_{k,i} = x_k - \hat{x}_{k,i}$ of $\mathcal{O}_i^t$ can be computed as:

$$\hat{\epsilon}_{k,i} = \Phi_{ki} \hat{\epsilon}_i^t.$$

Hence, we define the secure estimation of x_k in $\mathcal{E}_i$ by compensating the effect of the cyber attack as

$$\hat{x}_{k,i}^r = \hat{x}_{k,i} + \hat{\epsilon}_{k,i}, \tag{12}$$

while an estimate of y_k is reconstructed in $\mathcal{E}_i$ as

$$\hat{y}_{ki} = C_k \hat{x}_{k,i}^r, \tag{13}$$

and an estimate of Γ_{ki} is obtained in $\mathcal{E}_i$ as

$$\hat{\Gamma}_{ki} = \tilde{y}_{ki} - \hat{y}_{ki}. \tag{14}$$

As noted in Lemma 3.2, there are unobservable modes corresponding to eigenvalues of A_i^r at 0, which may impact $\hat{x}_{k,i}^r$. This limitation is addressed explicitly in the following section, in Lemma 4.1. In the remainder of this section, we make the following assumption, to show the benefits of maintaining the cascade of $\mathcal{O}_i^t$ and $\mathcal{O}_i^r$ within $\mathcal{E}_i$. Here we define T_{a_i} as the earliest start time of the attacks on the communication links to $\mathcal{S}_i$.

Assumption 3.3. $\psi_i^t(T_{a_i})$ is orthogonal to the unobservable subspace of Ψ_i^t. Furthermore, $E_i^r \underline{\Gamma}_i(t)$ is orthogonal to the unobservable subspace of Ψ_i^t for all $t \geq T_{a_i}$.

The above assumption is not strictly necessary to prove the results in the remainder of this section. However, it is sufficient to ensure that the unobservable eigenvalues of A_i^r at 0 do not impact remaining modes of the system. Indeed, depending on the internal structure of Ψ_i^t, it may hold for cases other than Assumption 3.3 that the unobservable modes of ψ_i^t corresponding to eigenvalues at 0 of A_i^r do not influence any of the remaining, detectable modes of the system, and therefore may not influence $\hat{\epsilon}_{k,i}$. We decide, however, to pursue results relying on Assumption 3.3, as it is more general. As we discuss in Section 4, any initial condition $\psi_i^t(T_{a_i})$ can be decomposed into a portion satisfying Assumption 3.3, and a portion which behaves as an SFDI attack.

The following theorem formalizes how secure estimation is achieved for the data injected by the attacker as well as the state and output.

Theorem 3.1. *Let Assumption 2.1, 3.2 and 3.3 hold, and suppose the matrices F_i^t and F_i^r (apart from its unobservable eigenvalues at 0) are designed to be Hurwitz. Furthermore, assume a non-stealthy FDI attack and that $\underline{\Gamma}_i \to 0$. The proposed cascade of observers in $\mathcal{E}_i$, namely $\mathcal{O}_i^t$ and $\mathcal{O}_i^r$, enable reconstruction of x_k, y_k, and Γ_{ki} for $i \in \mathcal{N}$ and $k \in \mathcal{N}_i^c$ according to (12)–(14), respectively.*

Proof. The residual for (11) is defined as $r_i^r = \tilde{r}_i^t - \hat{\tilde{r}}_i^t = C_i^r \epsilon_i^r$, where $\epsilon_i^r = \psi_i^r - \hat{\psi}_i^r$ has the following dynamics:

$$\dot{\epsilon}_i^r = F_i^r \epsilon_i^r + E_i^r \underline{\Gamma}_i. \tag{15}$$

Algorithm 1 Secure Estimation Under Non-stealthy FDI Attacks and Their Detection and Isolation

for $i \leftarrow 1$ to N **do**
  UIO Design in $\mathcal{S}_i$:
  Input: $\tilde{y}_{ki}$ for $k \in \mathcal{N}_i^c$ (online)
    $A_k, B_k, A_{kj}, j \in \mathcal{N}_k^p, C_k$ (offline)
  Aggregated system $\leftarrow$ (1) and (5);
  $K_i^{(1)^t} \leftarrow$ placing eigenvalues of F_i^t;
  $H_i^t, T_i^t, K_i^{(2)^t}, K_i^t \leftarrow$ (7a)–(7e);
  $\hat{x}_i^t, \tilde{r}_i^t \leftarrow \mathcal{O}_i^t$ given in (6).
  Luenberger-like observer Design in $\mathcal{S}_i$:
  Input: $\tilde{r}_i^t$ (online)
    $F_i^t, H_i^t, K_i^{(1)^t}$ (offline)
  Augmented system $\leftarrow$ (9);
  $L_i^r \leftarrow$ placing observable eigenvalues of F_i^r;
  $\hat{\epsilon}_i^t \leftarrow \mathcal{O}_i^r$ given in (11);
  for $k \in \mathcal{N}_i^c$ **do**
    if $\hat{\Gamma}_{ki} \nrightarrow 0$ **then**
      detection and isolation of a non-stealthy FDI attack on the link *ki*.
      $\hat{x}_{k,i}^r \leftarrow \hat{x}_{k,i}$ and $\hat{\epsilon}_{k,i}$ via (12);
      $\hat{y}_{ki} \leftarrow \hat{x}_{k,i}$ via (13);
      $\hat{\Gamma}_{ki} \leftarrow$ (14).
    else
      no non-stealthy FDI attack on the link *ki*.
    end if
  end for
end for

Considering Assumption 3.3 and $\underline{\Gamma}_i \to 0$, then $\epsilon_i^r \to 0$. This in turn implies that $\hat{\epsilon}_i^t \to \epsilon_i^t$, and $\hat{\epsilon}_{k,i} \to \epsilon_{k,i}$. Thus, $\hat{x}_{k,i}^r \to x_k$, and $\hat{y}_{ki} = C_k \hat{x}_{k,i}^r \to y_k$. Therefore, $\hat{\Gamma}_{ki} = \tilde{y}_{ki} - \hat{y}_{ki} = \Gamma_{ki} + C_k \epsilon_{k,i} \to \Gamma_{ki}$. ■

Remark 3.3. In practice, it is reasonable to assume Γ_i^t is finite due to the limited resources of the attacker, justifying the assumption that $\underline{\Gamma}_i \to 0$. However, this may not always be satisfied, potentially leading to biased estimation if $\underline{\Gamma}_i^t \not\to 0$. To address this, as outlined in Gao and Ding (2007), Gao et al. (2007), higher derivatives of Γ_i^t, up to $\Gamma_i^{t^{(m)}}$, can be included in the states of (9), considering them as auxiliary states to be estimated. Then, asymptotic estimation convergence is guaranteed if $\Gamma_i^{t^{(m+1)}} \to 0$, which is reasonable as attacks may typically be designed to be incipient and slowly varying over time to avoid detection, as was assumed for faults (Gao & Ding, 2007; Gao et al., 2007, 2016; Zhao & Polycarpou, 2022). Note that accounting for higher derivatives of Γ_i^t does not affect the detectability of (9).

According to Theorem 3.1, under a non-stealthy FDI attack, as long as Assumption 3.3 holds, $\hat{\Gamma}_{ki} \to \Gamma_{ki}$. Therefore, given that under nominal operation, $\Gamma_{ki} = 0$ holds, having $\hat{\Gamma}_{ki} \nrightarrow 0$ indicates a non-stealthy FDI attack on communication link *ki*, enabling detection and isolation. The proposed distributed scheme for full reconstruction/secure estimation, as well as the detection and isolation logic is summarized in Algorithm 1.

Remark 3.4. Through algebraic manipulations, it can be shown that all the assignable eigenvalues in F_i^t via design of $K_i^{(1)^t}$ are also assignable in F_i^r, by design of L_i^r. Thus, the design of F_i^r can be fully decoupled from that of F_i^t. In Section 6.2, we provide some comments relating to how $K_i^{(1)^t}$ and L_i^r are to be designed to improve the performance of the cascade of $\mathcal{O}_i^t$ and $\mathcal{O}_i^r$ for detection and reconstruction of non-stealthy FDI attacks.

3.4. Stability analysis under secure control

Having shown that asymptotic reconstruction of the original transmitted outputs is possible, we address attack accommodation, showing that closed-loop stability of (1) is maintained. Suppose the LSS is regulated via the static, distributed, output-feedback control law:

$$u_i = K_{ii} y_i + \sum_{k \in \mathcal{N}_i^c} K_{ik} \hat{y}_{ki}, \tag{16}$$

where the feedback gains K_{ii} and K_{ik} are designed to ensure closed-loop stability under nominal conditions (i.e., $\Gamma_{ki} = 0, \forall i, k \in \mathcal{N}$). The following theorem addresses the stability of the LSS under attack accommodation.

Theorem 3.2. *Consider $\mathcal{P}_i, \forall i \in \mathcal{N}$ as given in (1) with the control input specified in (16). Then, $\mathcal{P}_i$ is asymptotically stable, assuming a non-stealthy FDI attack on communication link $ki, \forall k \in \mathcal{N}_i^c$, such that $\underline{\Gamma}_i$ is bounded with an unknown bound while Assumption 3.3 holds.*

Proof. By substituting (16) in (1) and aggregating the dynamics of $\mathcal{P}_i, \forall i \in \mathcal{N}$, we obtain

$$\dot{x} = A_K x + BK\Theta\epsilon^r, \tag{17}$$

where $x = \operatorname{col}_{i \in \mathcal{N}}(x_i)$, $\epsilon^r = \operatorname{col}_{i \in \mathcal{N}}(\epsilon_i^r)$, $A_K = A + BKC$, such that $A = [A_{ik}]$ with $A_{ii} = A_i$, $K = [K_{ik}]$, $B = \operatorname{diag}_{i \in \mathcal{N}}(B_i)$, $C = \operatorname{diag}_{i \in \mathcal{N}}(C_i)$, and Θ is such that

$$K\Theta\epsilon^r = \operatorname*{col}_{i \in \mathcal{N}} \left(\sum_{k \in \mathcal{N}_i^c} K_{ik} (\Gamma_{ki} - \hat{\Gamma}_{ki}) \right).$$

Furthermore, the estimation error dynamics of all $\mathcal{O}_i^r, i \in \mathcal{N}$ can be aggregated as:

$$\dot{\epsilon}^r = F^r \epsilon^r + E^r \underline{\Gamma},$$

where $F^r = \operatorname{diag}_{i \in \mathcal{N}}(F_i^r)$, $E^r = \operatorname{diag}_{i \in \mathcal{N}}(E_i^r)$, and $\underline{\Gamma} = \operatorname{col}_{i \in \mathcal{N}}(\underline{\Gamma}_i)$. Introducing $\eta = \begin{bmatrix} x^\top & \epsilon^{r\top} \end{bmatrix}^\top$, whose dynamics are

$$\dot{\eta} = \begin{bmatrix} A_K & BK\Theta \\ 0 & F^r \end{bmatrix} \eta + \begin{bmatrix} 0 \\ E^r \end{bmatrix} \underline{\Gamma}, \tag{18}$$

Under Assumption 3.3 and given (15), ϵ^r is bounded as long as $\underline{\Gamma}$ is bounded. This together with that A_K is Hurwitz by design of the control gains implies that the system (18) is bounded-input bounded-output stable. Thus, η is bounded for bounded $\underline{\Gamma}$, and $\eta \to 0$ if $\underline{\Gamma} \to 0$. ■

In the case $\underline{\Gamma} \nrightarrow 0$, it can be treated as a disturbance in (18), whose effect can be mitigated through the design of $L_i^r, i \in \mathcal{N}$, for example, using H_∞ optimization.

4. Accommodation of SFDI attacks

4.1. Vulnerability of $\mathcal{O}_i^t$ and $\mathcal{O}_i^r$ under SFDI attacks

In this section, we provide sufficient conditions under which cyber attacks remain stealthy to the cascade of observers $\mathcal{O}_i^t$ and $\mathcal{O}_i^r$, thus highlighting their limitations through Lemmas 4.1 and 4.2.

Lemma 4.1. *Consider (3). If $\Gamma_i^t(T_{a_i})$ is such that $\psi_i^t(T_{a_i})$ lies within the unobservable subspace of Ψ_i^t, and $\underline{\Gamma}_i(t)$ is such that $\psi_i^t(t), t \geq T_{a_i}$ remains in the unobservable subspace of Ψ_i^t, then the attack is stealthy to $\mathcal{O}_i^r$.*

Proof. The first statement is sufficient, for $\underline{\Gamma}_i = 0$, as $\psi_i^t(t), t \geq T_{a_i}$ remains in the unobservable subspace of Ψ_i^t by definition. The latter statement is sufficient, as $\underline{\Gamma}_i$ is such that $\psi_i^t(t)$ does not affect the output. ■

It is possible for ψ_i^t to not satisfy *either* Assumption 3.3 or the conditions in Lemma 4.1. In this case, by superposition, it is possible to decompose $\psi_i^t = \psi_{i,o}^t + \psi_{i,u}^t$, where $\psi_{i,o}^t$ satisfies Assumption 3.3, and thus the corresponding components of Γ_i^t are considered non-stealthy, while $\psi_{i,u}^t$ satisfies the conditions in Lemma 4.1. As such, the combination of $\mathcal{O}_i^t$, $\mathcal{O}_i^r$ and the estimator described in this section can properly accommodate them.

Let us now address covert attacks, as formulated in (4). In Lemma 4.2, we demonstrate that the attacker cannot freely design α_{ki} in (4) if it is to maintain stealthiness, as defined in Definition 2.2, in case of multiple simultaneous cyber attacks. For the statement of Lemma 4.2, we define $\tilde{\Phi}_{\mathcal{G}'_{ki,j}}$ as the full row rank block-matrix that maps the attacker-defined input α_{ki} in (4) to the components corresponding to $A_{kj}, j \in \mathcal{G}'_{k,i}$.[3]

Lemma 4.2. *Consider an FDI attack implemented by injecting data as in (3), where the attacker is specified as in Definition 2.3. The attack is* stealthy *to $\mathcal{O}_i^r$ if*

$$\operatorname*{col}_{j\in\mathcal{G}'_{k,i}}\left(\tilde{\Phi}_{\mathcal{G}'_{k,j}}\alpha_{ki} - \tilde{x}_{ji}\right) \in \ker\left(\operatorname*{row}_{j\in\mathcal{G}'_{k,i}}(A_{kj})\right), \tag{19}$$

leading to $\hat{x}_{k,i}^r \to x_k + \tilde{x}_{ki}$ for $i \in \mathcal{N}$ and $k \in \mathcal{N}_i^c$.

Proof. We rewrite the dynamics in (4) as

$$\dot{\tilde{x}}_{ki} = A_k\tilde{x}_{ki} + B_{\varsigma_{k,i}}\alpha_{\varsigma,ki} + \sum_{j\in\mathcal{G}'_{k,i}} A_{kj}\tilde{\Phi}_{\mathcal{G}'_{ki,j}}\alpha_{ki},$$

where the definition of $\alpha_{\varsigma,ki}$ follows from the definition of $B_{\varsigma_{k,i}}$ as the full column rank counterpart of $[B_k \;\; B_{c_{k,i}}]$. Aggregating these reformulated dynamics yields

$$\dot{\tilde{x}}_i^t = A_i^t\tilde{x}_i^t + B_{\varsigma_i}^t\alpha_{\varsigma,i}^t + \operatorname*{col}_{k\in\mathcal{N}_i^c}\left(\sum_{j\in\mathcal{G}'_{k,i}} A_{kj}(\tilde{\Phi}_{\mathcal{G}'_{ki,j}}\alpha_{ki} - \tilde{x}_{ji})\right),$$

where $\tilde{x}_i^t = \operatorname*{col}_{k\in\mathcal{N}_i^c}(\tilde{x}_{ki})$, $\alpha_{\varsigma,i}^t = \operatorname*{col}_{k\in\mathcal{N}_i^c}(\alpha_{\varsigma,ki})$, and $B_{\varsigma_i}^t = \operatorname*{diag}_{k\in\mathcal{N}_i^c}(B_{\varsigma_{k,i}})$. For an attack to be covert, the last term of the above equation should be zero, so that the dynamics of $\mathcal{S}_k$ can be replicated by the attacker. Thus, cyber attacks must satisfy, for all $k \in \mathcal{N}_i^c$:

$$\sum_{j\in\mathcal{G}'_{ki,j}} A_{kj}\left(\tilde{\Phi}_{\mathcal{G}'_{k,j}}\alpha_{ki} - \tilde{x}_{ji}\right) = 0,$$

which is equivalent to (19). If this holds, (8) turns into

$$\begin{aligned}\dot{\tilde{\epsilon}}_i^t &= F_i^t\tilde{\epsilon}_i^t,\\ \tilde{r}_i^t &= C_i^t\tilde{\epsilon}_i^t,\end{aligned}$$

where $\tilde{\epsilon}_i^t = x_i^t - \hat{x}_i^t + \tilde{x}_i^t$. Having $\tilde{\epsilon}_i^t \to 0$ yields $\tilde{r}_i^t \to 0$, and therefore, the attack is stealthy, as per Definition 2.2. Moreover, given that the resulting dynamics show no trace of an attack, $\mathcal{O}_i^r$ is incapable of reconstructing Γ_i^t, while $\psi_i^r \to 0$, and $\hat{x}_{k,i}^r \to x_k + \tilde{x}_{ki}$. ■

Remark 4.1. The limitation posed by (19) for the attacker's input design stems from the definition of the aggregated system in (5). Since $\mathcal{O}_i^t$ utilizes the physical couplings of cyber neighbors, (19) must be met if multiple communication lines are attacked, limiting the attacker's capability to freely design α_{ki}.

[3] Recall that $\mathcal{G}'_{k,i} = \mathcal{N}_k^p \backslash \mathcal{G}_{k,i}$, where $\mathcal{G}_{k,i} = \mathcal{N}_k^p \backslash (\mathcal{N}_k^p \cap \mathcal{N}_i^c)$.

Having highlighted the vulnerability of the cascade of $\mathcal{O}_i^r$ and $\mathcal{O}_i^t$ against SFDI attacks, the more challenging problem of secure estimation in the presence of this class of attacks remains unresolved. This is addressed in the remainder of this section, where we focus on covert FDI attacks, as defined in Definition 2.3.

4.2. Design and analysis of the observer

To address the limitations of the cascade of observers, $\mathcal{O}_i^t$ and $\mathcal{O}_i^r$, we design an observer in this section to provide a secure state estimation under SFDI attacks. Given SFDI attacks on communication links between subsystems, for achieving secure state estimation, we assume that the attacker cannot alter sensor measurements y_i, obtained locally within a subsystem, and therefore, y_i is considered *secure*. Because of this assumption, it is possible to exploit y_i to reconstruct the physical couplings between subsystems, $\bar{x}_i = \breve{B}_{c_i}\operatorname*{col}_{k\in\mathcal{N}_i^p}(x_k)$, leading to the secure estimation of x_k, either partially or fully, depending on the structure of the physical couplings. Therefore, the objective is to design an observer to estimate $\bar{x}_i$. Next, we define

$$\begin{aligned}\dot{x}_i^p &= A_i^p x_i^p + B_i^p u_i + E_i^p \underline{x}_i,\\ y_i &= C_i^p x_i^p,\end{aligned} \tag{20}$$

where $x_i^p = \begin{bmatrix} x_i^\top & \dot{\bar{x}}_i^\top & \bar{x}_i^\top \end{bmatrix}^\top$, $\underline{x}_i = \ddot{\bar{x}}_i$, and

$$A_i^p = \begin{bmatrix} A_i & 0 & B_{c_i} \\ 0 & 0 & 0 \\ 0 & I & 0 \end{bmatrix}, \quad B_i^p = \begin{bmatrix} B_i \\ 0 \\ 0 \end{bmatrix}, E_i^p = \begin{bmatrix} 0 \\ I \\ 0 \end{bmatrix}, \quad C_i^p = \begin{bmatrix} C_i & 0 & 0 \end{bmatrix}. \tag{21}$$

Then, a Luenberger-like observer is designed to estimate x_i^p, with the following dynamics:

$$\mathcal{O}_i^p : \begin{cases} \dot{\hat{x}}_i^p = A_i^p\hat{x}_i^p + B_i^p u_i + L_i^p(y_i - \hat{y}_i^p),\\ \hat{y}_i^p = C_i^p\hat{x}_i^p, \end{cases} \tag{22}$$

where $\hat{x}_i^p \in \mathbb{R}^{n_i+2n_{c_i}}$ and $\hat{y}_i^p \in \mathbb{R}^{q_i}$ are the estimated state and output, respectively, and L_i^p is a design variable, determined by placing the eigenvalues of $F_i^p = A_i^p - L_i^p C_i^p$ at desired locations to ensure that F_i^p is Hurwitz. Based on this observer, we obtain

$$\hat{\bar{x}}_i = \begin{bmatrix} 0 & 0 & I \end{bmatrix}\hat{x}_i^p. \tag{23}$$

Assumption 4.1. For each $\mathcal{P}_i, i \in \mathcal{N}$ with dynamics as in (1), the matrix $\begin{bmatrix} A_i & B_{c_i} \\ C_i & 0 \end{bmatrix}$ is full column rank.

Lemma 4.3. *Assumptions 2.1 and 4.1 are sufficient to ensure the detectability of the pair (A_i^p, C_i^p), and thereby the stability of $\mathcal{O}_i^p$.*

The proof follows from the detectability of (A_i^p, C_i^p) (Luenberger, 1964), which requires the detectability of (A_i, C_i). This is satisfied under Assumption 2.1. Furthermore, detectability of (A_i^p, C_i^p) requires that $\bar{x}_i$ be reconstructable, which is met as per Assumption 4.1.

Proposition 4.1. *Let Assumptions 2.1 and 4.1 hold, and F_i^p be Hurwitz. In the presence of SFDI attacks on communication links between subsystems, if $\underline{x}_i \to 0$, then $\mathcal{O}_i^p$ enables reconstruction of $\bar{x}_i$.*

Proof. Define the residual of $\mathcal{O}_i^p$ as $r_i^p = y_i - \hat{y}_i^p = C_i^p\epsilon_i^p$, where $\epsilon_i^p = x_i^p - \hat{x}_i^p$, whose dynamics are given as

$$\dot{\epsilon}_i^p = F_i^p\epsilon_i^p + E_i^p\underline{x}_i. \tag{24}$$

If $\underline{x}_i \to 0$, then $\epsilon_i^p \to 0$, thus guaranteeing $\hat{\bar{x}}_i \to \bar{x}_i$. ■

Remark 4.2. As discussed in Remark 3.3, assuming that cyber attacks are bounded, then x_i is bounded, making the assumption $\underline{x}_i \to 0$ reasonable. Moreover, higher derivatives of $\bar{x}_i$ can be considered as auxiliary states to decouple the estimation error from their effect.

Proposition 4.2. *For $k \in \mathcal{N}_i^p$, if $\text{Rank}(A_{ik}) = n_k$ and*

$$\text{Im}(A_{ik}) \cap \Big(\bigcup_{j \in \mathcal{N}_i^p \setminus \{k\}} \text{Im}(A_{ij}) \Big) = 0, \tag{25}$$

x_k can be fully reconstructed from $\bar{x}_i$, and the original transmitted outputs can be reconstructed as $\hat{y}_{ki} = C_k \hat{x}_k$.

Proof. The second condition is equivalent to stating that the image of A_{ik} is column-independent to all other interconnection matrices $A_{ij}, j \in \mathcal{N}_i^p \setminus \{k\}$. This, together with $\text{Rank}(A_{ik}) = n_k$ and the definition of B_{c_i}, implies that $\hat{x}_k$ can be uniquely obtained from $\hat{\bar{x}}_i$. The computation of $\hat{y}_{ki} = C_k \hat{x}_k$ follows. ■

Proposition 4.3. *For $k \in \mathcal{N}_i^p$, if $\text{Rank}(A_{ik}) \neq n_k$ and (25) holds, then x_k can be partially reconstructed.*

Proof. To estimate components of x_k from $\hat{\bar{x}}_i$, $\text{Rank}(A_{ik}) = n_k$ and (25) must hold (see Proposition 4.2). If A_{ik} is not full column rank, then estimates of some components or a linear combination of their associated components of x_k can be derived from $\hat{\bar{x}}_i$ under (25). ■

Remark 4.3. The key difference between the properties of $\mathcal{O}_i^p$ compared to those of the cascade of $\mathcal{O}_i^t$ and $\mathcal{O}_i^r$ is that $\mathcal{O}_i^p$ depends on y_i, which is supposed to be uncorrupted by an attacker. As such, $\mathcal{O}_i^p$ is not susceptible to stealthy and covert FDI attacks, as defined in Definitions 2.2 and 2.3, and the obtained estimate is secure.

4.3. Stability analysis under secure control

Consider the following static, distributed, output-feedback control law:

$$u_i = K_{ii} y_i + \sum_{k \in \mathcal{N}_i^{c,f}} K_{ik} \hat{y}_{ki} + \sum_{k \in \mathcal{N}_i^c \setminus \mathcal{N}_i^{c,f}} K_{ik} \tilde{y}_{ki}, \tag{26}$$

where K_{ii} and K_{ik} are chosen such that under nominal system operation, asymptotic stability of the LSS is guaranteed. Moreover, $\mathcal{N}_i^{c,f} \subseteq \mathcal{N}_i^c$ consists of the cyber neighbors whose states can be *fully* reconstructed according to Proposition 4.2. For any remaining cyber neighbors, their received sensor measurements, which are possibly altered by SFDI attacks, are employed in the controller. The following theorem addresses the stability of the LSS under attack accommodation.

Theorem 4.1. *Consider $\mathcal{P}_i$ as given in (1), with the control input specified in (26). Then, $\mathcal{P}_i$, $\forall i \in \mathcal{N}$ is bounded-input bounded-output stable, assuming an SFDI attack on communication link ki, such that $\underline{x}_i$ and Γ_{ki}, $k \in \mathcal{N}_i^c \setminus \mathcal{N}_i^{c,f}$ are bounded with an unknown bound.*

Proof. Given full reconstruction, define $\hat{x}_k = \Phi_k \hat{\bar{x}}_i$, with full row rank Φ_k. Thus, rewrite (1) as

$$\begin{aligned}\dot{x}_i =& A_i x_i + B_i K_{ii} y_i + B_i \sum_{k \in \mathcal{N}_i^c} K_{ik} y_k - B_i \sum_{k \in \mathcal{N}_i^{c,f}} K_{ik} C_k \Phi_k \begin{bmatrix} 0 & 0 & I \end{bmatrix} \epsilon_i^p \\ &+ B_i \sum_{k \in \mathcal{N}_i^c \setminus \mathcal{N}_i^{c,f}} K_{ik} \Gamma_{ki} + \sum_{k \in \mathcal{N}_i^p} A_{ik} x_k.\end{aligned}$$

Aggregating the state dynamics of $\mathcal{P}_i$, $\forall i \in \mathcal{N}$ yields

$$\dot{x} = A_K x - BK\Theta\epsilon^p + BK\Gamma, \tag{27}$$

where $x = \underset{i \in \mathcal{N}}{\text{col}}(x_i)$, and $\epsilon^p = \underset{i \in \mathcal{N}}{\text{col}}(\epsilon_i^p)$. Moreover, A_K and B are those defined for (17), and Θ is defined such that

$$\begin{aligned}K\Theta\epsilon^p =& \underset{i \in \mathcal{N}}{\text{col}} \left(\sum_{k \in \mathcal{N}_i^{c,f}} K_{ik} C_k \Phi_k \begin{bmatrix} 0 & 0 & I \end{bmatrix} \epsilon_i^p \right), \\ K\Gamma =& \underset{i \in \mathcal{N}}{\text{col}} \left(\sum_{k \in \mathcal{N}_i^c \setminus \mathcal{N}_i^{c,f}} K_{ik} \Gamma_{ki} \right).\end{aligned}$$

Considering (24), the error dynamics for $\mathcal{P}_i$, $\forall i \in \mathcal{N}$ are

$$\dot{\epsilon}^p = F^p \epsilon^p + E\underline{x}, \tag{28}$$

where $F^p = \underset{i \in \mathcal{N}}{\text{diag}}(F_i^p)$, $E = \underset{i \in \mathcal{N}}{\text{diag}}(E_i^p)$, and $\underline{x} = \underset{i \in \mathcal{N}}{\text{col}}(\underline{x}_i)$.

The derivatives of Lyapunov functions $V_x = x^\top P_x x$, and $V_\epsilon = {\epsilon^p}^\top P_\epsilon \epsilon^p$ with $P_x, P_\epsilon > 0$ along (27) and (28) are

$$\begin{aligned}\dot{V}_x =& \dot{x}^\top P_x x + x^\top P_x \dot{x} \\ =& x^\top \text{He}(A_K^\top P_x) x - \text{He}(x^\top P_x BK\Theta\epsilon^p) + \text{He}(x^\top P_x BK\Gamma), \\ \dot{V}_\epsilon =& {\epsilon^p}^\top \text{He}({F^p}^\top P_\epsilon)\epsilon^p + \text{He}({\epsilon^p}^\top P_\epsilon E\underline{x}).\end{aligned}$$

To evaluate the H_∞ performance, under zero initial conditions, let

$$\begin{aligned}J =& \int_0^\infty (z^\top z - \zeta^2 \Gamma^\top \Gamma - \zeta^2 \underline{x}^\top \underline{x} + \dot{V}_x + \dot{V}_\epsilon) dt - \int_0^\infty (\dot{V}_x + \dot{V}_\epsilon) dt \\ \leq& \int_0^\infty \begin{bmatrix} x^\top & {\epsilon^p}^\top & \Gamma^\top & \underline{x}^\top \end{bmatrix} J_1 \begin{bmatrix} x^\top & {\epsilon^p}^\top & \Gamma^\top & \underline{x}^\top \end{bmatrix}^\top,\end{aligned}$$

where $z = C_x x + C_\epsilon \epsilon^p$ is the measured output employed to validate the stability of the system, C_x and C_ϵ are weighting matrices and designed based on the desired performance objectives, $\zeta > 0$ is a scalar,

$$J_1 = \begin{bmatrix} J_{11} & -P_x BK\Theta + C_x^\top C_\epsilon & P_x BK & 0 \\ * & J_{22} & 0 & P_\epsilon E \\ * & * & -\zeta^2 I & 0 \\ * & * & * & -\zeta^2 I \end{bmatrix},$$

$J_{11} = \text{He}(A_K^\top P_x) + C_x^\top C_x$, and $J_{22} = \text{He}({F^p}^\top P_\epsilon) + C_\epsilon^\top C_\epsilon$. Guaranteeing $J_1 < 0$ is a sufficient condition for $J < 0$ and achieving H_∞ performance $\|z\|^2 \leq \zeta^2(\|\Gamma\| + \|\underline{x}\|)$. Ensuring $J_1 < 0$ is equivalent to holding

$$J_1' = \begin{bmatrix} J_{11} & -P_x BK\Theta + C_x^\top C_\epsilon \\ * & J_{22} \end{bmatrix} < 0, \tag{29a}$$

$$-\zeta^2 I < \begin{bmatrix} P_x BK & 0 \\ 0 & P_\epsilon E \end{bmatrix}^\top J_1'^{-1} \begin{bmatrix} P_x BK & 0 \\ 0 & P_\epsilon E \end{bmatrix}, \tag{29b}$$

where (29a) is equivalent to satisfying

$$J_{11} < 0, \tag{30a}$$

$$J_{22} - \left(P_x BK\Theta + C_x^\top C_\epsilon\right)^\top J_{11}^{-1} \left(P_x BK\Theta + C_x^\top C_\epsilon\right) < 0. \tag{30b}$$

Guaranteeing $J_{11} < 0$ is equivalent to $A_K^\top P_x + P_x A_K = -Q_1 - C_x^\top C_x$, where $Q_1 + C_x^\top C_x > 0$. This is a Lyapunov equation and always feasible since A_K is Hurwitz. (30b) is equivalent to ${F^p}^\top P_\epsilon + P_\epsilon F^p = -Q_2 - C_\epsilon^\top C_\epsilon + \left(P_x BK\Theta + C_x^\top C_\epsilon\right)^\top J_{11}^{-1} \left(P_x BK\Theta + C_x^\top C_\epsilon\right)$, where $Q_2 + C_x^\top C_x + \left(P_x BK\Theta + C_x^\top C_\epsilon\right)^\top J_{11}^{-1} \left(P_x BK\Theta + C_x^\top C_\epsilon\right) > 0$. This is a Lyapunov equation and always feasible given $J_{11} < 0$ and that F^p is Hurwitz. Holding both (30a) and (30b) satisfies (29a).

The necessary and sufficient condition to (29b) is $\rho[-J_1'^{-1}] \leq \zeta^2$ for which $\zeta^2 \geq \min(|\lambda(J_1')|)$ is a sufficient condition, where

Algorithm 2 Secure Estimation Under SFDI Attacks and Their Detection and Isolation

for $i \leftarrow 1$ to N **do**
  Secure State Estimation:
  Input: u_i, y_i (online)
    $A_i, B_i, A_{ik}, k \in \mathcal{N}_i^p, C_i$ (offline)
  Augmented system $\leftarrow$ (20);
  $L_i^p \leftarrow$ placing eigenvalues of F_i^p;
  $\hat{\bar{x}}_i \leftarrow$ Luenberger-like observer (22) and (23);
  $\hat{x}_k \leftarrow \hat{\bar{x}}_i$ *(fully or partially, Props. 4.2 and 4.3)*;
  if $\hat{x}_k$ is fully attained **then**
    $\hat{y}_{ki} \leftarrow C_k \hat{x}_k$.
  end if
  SFDI Attack Detection by $\mathcal{S}_i$**:**
  $\hat{\bar{x}}_{1_i} \leftarrow \Phi_{1_i} \hat{\bar{x}}_i$;
  $\hat{x}_{k,i}^r \leftarrow$ (12);
  Input: $\hat{\bar{x}}_{1_i}, \hat{x}_{k,i}^r, \forall k \in \mathcal{N}_i^{pc}$ (online)
  if $\hat{x}_{k,i}^r \nrightarrow \hat{\bar{x}}_{1_i}$ **then**
    SFDI attack is detected in $\mathcal{N}_i^{pc}$.
    SFDI Attack Isolation by $\mathcal{S}_i$**:**
    Input: $\hat{x}_k, \hat{x}_{k,i}^r, k \in \mathcal{N}_i^{pc}$ (online)
    for $k \in \mathcal{N}_i^{pc}$ **do**
      if $\hat{x}_{k,i}^r \nrightarrow \hat{x}_k$ **then**
        SFDI attack is isolated on link *ki*.
      else
        no SFDI attack is isolated on link *ki*.
      end if
    end for
  else
    no SFDI attack in $\mathcal{N}_i^{pc}$.
  end if
end for

$\rho[-J_1'^{-1}]$ is the spectral radius of $-J_1'^{-1}$ and $\lambda(J_1')$ denotes the eigenvalue vector of J_1'. This is always satisfied for some $\zeta > 0$ given any chosen P_x and P_ϵ. Holding (29a) and (29b) leads to $J_1 < 0$ achieving the H_∞ performance. ■

4.4. Detection and isolation

The proposed scheme provides two mechanisms for estimation. One relies on the cascade of observers, $\mathcal{O}_i^t$ and $\mathcal{O}_i^r$, while using sensor measurements transmitted between subsystems, which are potentially altered by attacks. The other one relies on $\mathcal{O}_i^p$, while using secure sensor measurements obtained locally within the subsystem. This concept forms the foundation for SFDI attack detection. However, detection becomes unattainable if any of these observers cannot be implemented. In scenarios where $\mathcal{N}_i^p \nsubseteq \mathcal{N}_i^c$, the outputs of $\mathcal{P}_k$, $k \in \mathcal{N}_i^p \backslash \mathcal{N}_i^c$ are not accessible in $\mathcal{S}_i$, thus preventing the successful implementation of $\mathcal{O}_i^t$ and $\mathcal{O}_i^r$. In order to address this, and before presenting the subsequent assumption and proposition, we define the set $\mathcal{N}_i^{pc} = \mathcal{N}_i^p \cap \mathcal{N}_i^c$. Then, the set of physical neighbors of $\mathcal{S}_i$ can be divided into two disjoint sets as $\mathcal{N}_i^p = \mathcal{N}_i^{pc} \cup (\mathcal{N}_i^p \backslash \mathcal{N}_i^{pc})$. The first set denotes those physical neighbors whose sensor measurements are received by $\mathcal{S}_i$, while the second includes physical neighbors whose sensor measurements are not available to $\mathcal{S}_i$. The interconnection term in (1) is then divided as

$$\sum_{k \in \mathcal{N}_i^p} A_{ik} x_k = \sum_{k \in \mathcal{N}_i^{pc}} A_{ik} x_k + \sum_{k \in \mathcal{N}_i^p \backslash \mathcal{N}_i^{pc}} A_{ik} x_k = B_{c1_i} \bar{x}_{1_i} + B_{c2_i} \bar{x}_{2_i},$$

where B_{c1_i} and B_{c2_i} are full column rank. If $\mathcal{N}_i^p \subseteq \mathcal{N}_i^c$, then $B_{c1_i} = B_{c_i}$, $\bar{x}_{1_i} = \bar{x}_i$, $B_{c2_i} = 0$, and $\bar{x}_{2_i} = 0$.

Assumption 4.2. The interconnection term of $\mathcal{P}_i$ adheres to either of the following conditions: (1) $\mathcal{N}_i^p \subseteq \mathcal{N}_i^c$ or (2) $\mathrm{Im}(B_{c1_i}) \neq \mathrm{Im}(B_{c2_i})$.

In the case $\mathcal{N}_i^p \nsubseteq \mathcal{N}_i^c$, the condition $\mathrm{Im}(B_{c1_i}) \neq \mathrm{Im}(B_{c2_i})$ ensures that $\hat{\bar{x}}_{1_i}$ can be uniquely derived from $\hat{\bar{x}}_i$ as $\hat{\bar{x}}_{1_i} = \Phi_{1_i} \hat{\bar{x}}_i$, where Φ_{1_i} is a full row rank block-diagonal matrix mapping $\hat{\bar{x}}_i$ to $\hat{\bar{x}}_{1_i}$.

Proposition 4.4. *Let Assumption 4.2 hold. An SFDI attack is detected in $\mathcal{N}_i^{pc}$, if $\hat{\bar{x}}_i^r \nrightarrow \hat{\bar{x}}_{1_i}$, where $\hat{\bar{x}}_i^r = \breve{B}_{c1_i} \underset{k \in \mathcal{N}_i^{pc}}{\mathrm{col}} (\hat{x}_{k,i}^r)$, and $\hat{x}_{k,i}^r$ is obtained using (12).*

Proof. Given Assumption 4.2, secure estimation of $\bar{x}_{1_i}$ is derived from $\hat{\bar{x}}_i$. According to Lemma 4.2, under an SFDI attack, $\hat{x}_{k,i}^r \to x_k + \tilde{x}_{ki}$. Therefore, $\hat{\bar{x}}_i^r = \breve{B}_{c1_i} \underset{k \in \mathcal{N}_i^{pc}}{\mathrm{col}} (\hat{x}_{k,i}^r) \to \breve{B}_{c1_i} \big(\underset{k \in \mathcal{N}_i^{pc}}{\mathrm{col}} (x_k) + \underset{k \in \mathcal{N}_i^{pc}}{\mathrm{col}} (\tilde{x}_{ki}) \big)$, while $\hat{\bar{x}}_{1_i} \to \bar{x}_{1_i} = \breve{B}_{c1_i} \underset{k \in \mathcal{N}_i^{pc}}{\mathrm{col}} (x_k)$. Hence, if $\hat{\bar{x}}_i^r \nrightarrow \hat{\bar{x}}_{1_i}$, an SFDI attack is detected in $\mathcal{N}_i^{pc}$. ■

Proposition 4.5. *Let (25) hold. A cyber attack on the communication link ki is isolated if $\hat{x}_k \not\to \hat{x}_{k,i}^r$.*

Proof. If (25) holds, full or partial reconstruction of x_k can be obtained from $\hat{\bar{x}}_i$, based on Propositions 4.2 and 4.3, respectively. Hence, $\hat{x}_k \to x_k$ enables secure estimation despite SFDI attacks, while $\hat{x}_{k,i}^r \to x_k + \tilde{x}_{ki}$. If $\hat{x}_{k,i}^r$ does not converge to $\hat{x}_k$, an SFDI attack is located on the communication link *ki*. The same holds for partial secure estimation as in Proposition 4.3. ■

Remark 4.4. Based on Proposition 4.4, detecting an SFDI attack relies on estimating x_k, $k \in \mathcal{N}_i^p$, derived from $\hat{\bar{x}}_i$. Thus, *SFDI* attack detection and isolation are plausible by $\mathcal{E}_i$ if the attacker targets any sensor measurements received from $\mathcal{S}_k$, $k \in \mathcal{N}_i^{pc}$, but not from $\mathcal{S}_k$, $k \in \mathcal{N}_i^c \backslash \mathcal{N}_i^{pc}$.

Remark 4.5. Note that the detection and isolation approach for SFDI attacks relies on both $\hat{\bar{x}}_i^r$ and $\hat{\bar{x}}_{1_i}$, as outlined in Propositions 4.4 and 4.5. Consequently, even in scenarios where $\mathcal{N}_i^p \subseteq \mathcal{N}_i^c$, $\mathcal{O}_i^p$ is indispensable to enable detection and isolation of SFDI attacks.

The following proposition, based on Gallo et al. (2020), discusses a case of coordinated SFDI attacks, in which an attacker can launch simultaneous attacks, such that they remain undetected by the proposed scheme.

Proposition 4.6. *If simultaneous multiple SFDI attacks are designed such that $\underset{k \in \mathcal{N}_i^{pc}}{\mathrm{col}} (\tilde{x}_{ki}) \subset \ker(\breve{B}_{c1_i})$, then the attacks are not detected by the proposed scheme.*

Proof. $\hat{x}_{k,i}^r \to x_k + \tilde{x}_{ki}$ results in $\breve{B}_{c1_i} \underset{k \in \mathcal{N}_i^{pc}}{\mathrm{col}} (\hat{x}_{k,i}^r) \to \bar{x}_{1_i} + \breve{B}_{c1_i} \underset{k \in \mathcal{N}_i^{pc}}{\mathrm{col}} (\tilde{x}_{ki})$. If $\breve{B}_{c1_i} \underset{k \in \mathcal{N}_i^{pc}}{\mathrm{col}} (\tilde{x}_{ki}) = 0$, then no SFDI attack is detected, and the implemented attacks are undetectable to the proposed scheme. ■

Remark 4.6. Note that the two mechanisms for estimation have complementary properties. Indeed, $\mathcal{O}_i^p$ can provide either *full* or *partial* secure estimation, depending on the structure of the interconnection matrices, but not on the attack, as demonstrated in Propositions 4.2 and 4.3. On the other hand, the combination of $\mathcal{O}_i^t$ and $\mathcal{O}_i^r$ provides *full* secure estimation, irrespective of the coupling, but only under non-stealthy FDI attacks, as established in Theorem 3.1.

The distributed scheme for full or partial reconstruction/secure estimation as well as the detection and isolation logic are summarized in Algorithm 2.

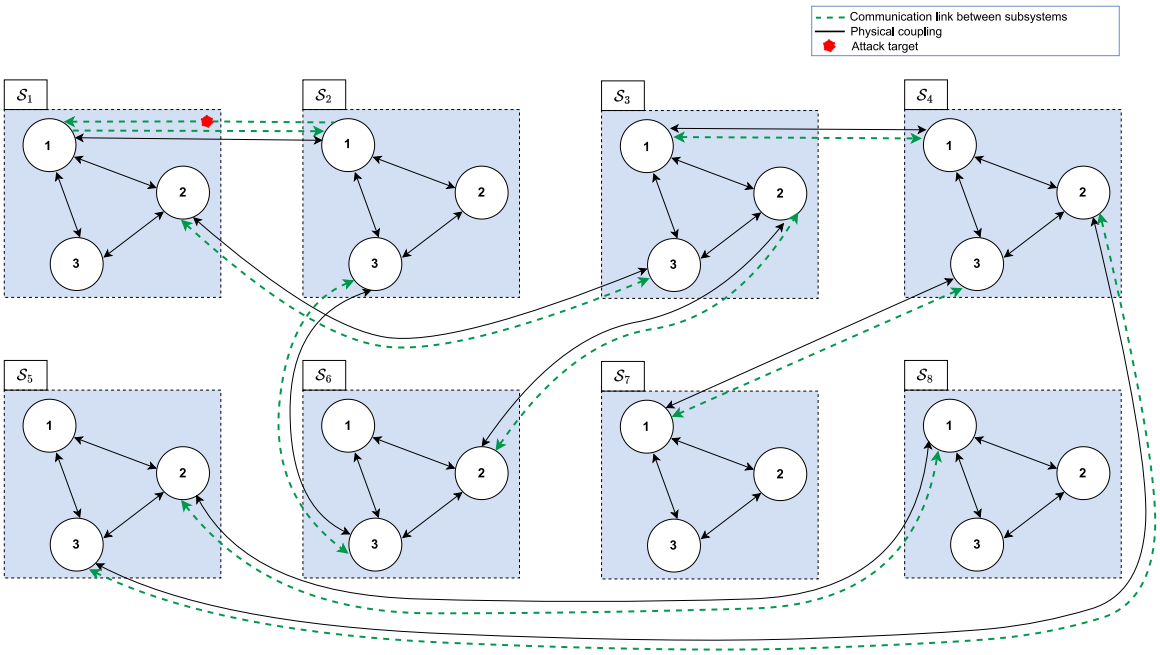

Fig. 2. The LSS with eight subsystems, considered in the simulation-based case study.

5. Computational complexity

For non-stealthy FDI attacks, based on the cascade of observers $\mathcal{O}_i^t$ and $\mathcal{O}_i^r$ in each $\mathcal{S}_i$, $i \in \mathcal{N} = \{1, \ldots, N\}$, the computational complexity of $\mathcal{O}_i^t$ is linearly related to $\sum_{k\in\mathcal{N}_i^c} n_k$, while for $\mathcal{O}_i^r$, it is $\sum_{k\in\mathcal{N}_i^c} n_k + 2\sum_{k\in\mathcal{N}_i^c} q_k$. In a centralized approach, implementing the cascade of observers for all subsystems in the LSS, the computational complexity is a linear function of $\sum_{i\in\mathcal{N}} n_i$ and $\sum_{i\in\mathcal{N}} n_i + 2\sum_{i\in\mathcal{N}} q_i$. Given that $\mathcal{N}_i^c \subset \mathcal{N}$, then $N_i^c = |\mathcal{N}_i^c| < N$. For each subsystem, in turn, the distributed scheme offers lower computational complexity and greater scalability, compared to the centralized approach. However, the total computational complexity for the LSS, considering all subsystems, may be higher. If a subsystem has many cyber neighbors, the computational complexity of the distributed scheme increases, while it remains constant for the centralized approach. The maximum computational complexity for $\mathcal{O}_i^t$ and $\mathcal{O}_i^r$ occurs when $\mathcal{N}_i^c = \mathcal{N}\backslash\{i\}$, resulting in $N_i^c = N - 1$. Therefore, even in such cases, the proposed distributed scheme exhibits lower computational complexity, compared to the centralized approach for each subsystem.

For SFDI attacks, based on $\mathcal{O}_i^p$ in $\mathcal{S}_i$, the computational complexity is linearly related to $n_i + n_{c_i}$. In a centralized approach, implementing the observer for all subsystems in the LSS, the complexity is a linear function of $\sum_{i\in\mathcal{N}} n_i$. Therefore, the distributed scheme offers lower computational complexity and greater scalability since $\sum_{i\in\mathcal{N}} n_i \geq n_i + n_{c_i}, \forall i \in \mathcal{N}$. Note that for high-dimensional processes, where n_i is large, both the distributed scheme and the centralized approach exhibit a linear increase in computational complexity.

Furthermore, while $\mathcal{O}_i^p$ alone enables secure state estimation, whether fully or partially, the detection and isolation of SFDI attacks require the implementation of both schemes (i.e., $\mathcal{O}_i^t$ and $\mathcal{O}_i^r$, in addition to $\mathcal{O}_i^p$), as discussed in Section 4.4. Given that each of the distributed schemes offers lower complexity than the centralized approach for each subsystem, the detection and isolation of SFDI attacks is computationally more efficient.

6. Simulation results

6.1. System definition

We consider an LSS inspired by a model of a DC microgrid that is partitioned into eight subsystems, where $\mathcal{N} = \{1, \ldots, 8\}$, with physical and cyber couplings, as depicted in Fig. 2. Each subsystem consists of three interconnected distributed generation units (DGUs), with a Buck converter connected to a point of common coupling via an RLC circuit, while modeling the Buck converters via their averaged dynamics (Tucci et al., 2018). Each jth DGU in $\mathcal{S}_i$ is described via the dynamics of its voltage at the point of common coupling V_i^j and its terminal current I_{ti}^j:

$$\begin{aligned} C_{ti}\dot{V}_i^j &= I_{ti}^j - I_{Li}^j + \sum_{(k,l)\in\mathcal{N}_{i,j}^p} \frac{V_k^l - V_i^j}{R_{ik}^{jl}}, \\ L_{ti}\dot{I}_{ti}^j &= V_{ti}^j - V_i^j - R_{ti}I_{ti}^j, \end{aligned} \tag{31}$$

where I_{Li}^j is the local load current, V_{ti}^j is the terminal voltage, the control input, and C_{ti}, L_{ti}, and R_{ti} are the capacitance, inductance, and resistance of the RLC circuit, respectively. The set $\mathcal{N}_{i,j}^p = \{(k,l) \in \{\{i\} \cup \mathcal{N}_i^p\} \times \mathcal{N}_k^w : \frac{\partial V_i^j}{\partial V_k^l} \neq 0\}$ is the set of all DGUs in subsystems which are physically coupled with the jth DGU of $\mathcal{S}_i$, where $\mathcal{N}_i^w = \{1, 2, 3\}, \forall i \in \mathcal{N}$. R_{ik}^{jl} is the resistance of the interconnection line between the jth DGU in $\mathcal{S}_i$ and the lth DGU in $\mathcal{S}_k$. Note that $R_{ik}^{jl} = R_{ki}^{jl}$ and $R_{ik}^{jl} = R_{ik}^{lj}$.

Each subsystem is equipped with a control unit combining a primary decentralized PI controller for stability, and a secondary consensus-based distributed controller. The control laws for the jth DGU in $\mathcal{C}_i$ for each $\mathcal{S}_i$ are defined as:

$$\begin{aligned} \dot{\nu}_i^j &= V_{ref} + \xi_i^j - V_i^j, \\ \dot{\xi}_i^j &= k_I \sum_{(k,l)\in\mathcal{N}_{i,j}^c} a_{ik}^{jl}(I_{tk}^l - I_{ti}^j), \\ V_{ti}^j &= k_{i,1}V_i^j + k_{i,2}I_{ti}^j + k_{i,3}\nu_i^j, \end{aligned} \tag{32}$$

where $\mathcal{N}_{i,j}^c \subseteq \{(k,l) \in (\{i\} \cup \mathcal{N}_i^c) \times \mathcal{N}_k^w\}$ indicates all those DGUs whose information is used by $\mathcal{C}_i$ to compute the input V_{ti}^j. The

Table 1
System, controller, and observer parameters.

	$k_{i,1}$	$k_{i,2}$	$k_{i,3}$	R_{ti} (Ω)	C_{ti} (mF)	L_{ti} (mH)	$\sigma(F_i^t)$	$\sigma(F_i^r)$	$\sigma(F_i^p)$
$\mathcal{S}_1$	−2.134	−0.163	13.553	0.2	2.2	1.8	$-5x,\ x=1:18$	$-20x,\ x=1:54$	$-10x\ \ x=1:13$
$\mathcal{S}_2$	−0.869	−0.05	48.285	0.3	1.9	2	$-5x,\ x=1:18$	$-20x,\ x=1:54$	$-10x\ \ x=1:13$
$\mathcal{S}_3$	−0.480	−0.108	30.673	0.1	1.7	2.2	$-5x,\ x=1:27$	$-20x,\ x=1:81$	$-10x\ \ x=1:15$
$\mathcal{S}_4$	−6.990	−0.175	102.96	0.5	2.5	3	$-5x,\ x=1:27$	$-20x,\ x=1:81$	$-10x\ \ x=1:15$
$\mathcal{S}_5$	−0.101	−0.01	16.393	0.4	2	1.3	$-5x,\ x=1:18$	$-20x,\ x=1:54$	$-10x\ \ x=1:13$
$\mathcal{S}_6$	−2.134	−0.163	13.553	0.6	3	2.5	$-5x,\ x=1:18$	$-20x,\ x=1:54$	$-10x\ \ x=1:13$
$\mathcal{S}_7$	−0.869	−0.05	130.285	0.3	2.8	2.1	$-5x,\ x=1:9$	$-20x,\ x=1:27$	$-10x\ \ x=1:11$
$\mathcal{S}_8$	−1.701	−0.06	80.393	0.3	2.7	1.9	$-5x,\ x=1:9$	$-20x,\ x=1:27$	$-10x\ \ x=1:11$

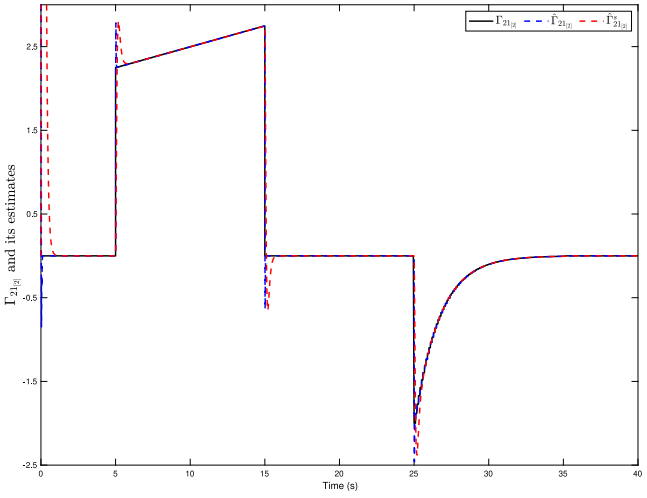

Fig. 3. Non-stealthy FDI attack: $\Gamma_{21_{[2]}}$ and estimates.

interested reader is directed to Tucci et al. (2018) for further details on the design of the controllers. $k_l = 0.5$ and $a_{ik}^{jl} = 1, \forall i \in \mathcal{N}, \forall j \in \mathcal{N}_i^w, \forall (k,l) \in \mathcal{N}_{i,j}^c$ are some of the chosen controller gains. The resistances are $R_{ii}^{12} = 0.04\Omega, R_{ii}^{13} = 0.03\Omega$ and $R_{ii}^{23} = 0.05\Omega, \forall i \in \mathcal{N}$. Also, $R_{12}^{11} = 0.05\Omega, R_{13}^{23} = 0.07\Omega, R_{26}^{33} = 0.1\Omega, R_{34}^{11} = 0.1\Omega, R_{36}^{22} = 0.08\Omega, R_{45}^{23} = 0.06\Omega, R_{47}^{31} = 0.09\Omega$, and $R_{58}^{21} = 0.08\Omega$. The load currents are set on $I_{Li}^j = 10A, \forall i \in \mathcal{N}, \forall j \in \mathcal{N}_i^w$. Table 1 lists the rest of the controller gains and parameters used in the simulation, which are the same for all three DGUs of each subsystem.

Consider $x_i^j = \left[V_i^j \ \ I_{ti}^j \ \ \nu_i^j\right]^\top$, $u_i^j = [V_{ti}^j]$, and $y_i^j = \begin{bmatrix} 1 & 0.8 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} x_i^j$.

The dynamics (31) and (32) can be rewritten in the same form as (1), with $x_i = \underset{j \in \mathcal{N}_i^w}{\text{col}}(x_i^j)$, $u_i = \underset{j \in \mathcal{N}_i^w}{\text{col}}(u_i^j)$, and $y_i = \underset{j \in \mathcal{N}_i^w}{\text{col}}(y_i^j)$. The system matrices are then defined following the introduced system dynamics. Given that a disturbance-free dynamical system is considered in this paper, it is assumed that all disturbance inputs $d_i, i \in \mathcal{N}$ are known to $\mathcal{E}_i$, where d_i is defined in Tucci et al. (2018).

6.2. Design of observers

The system matrices satisfy Assumption 2.1, 3.2, and 4.1, and therefore $\mathcal{O}_i^t$, $\mathcal{O}_i^r$, and $\mathcal{O}_i^p$ can provide secure estimations based on Lemmas 3.1, 3.2, and 4.3.

To evaluate the impact of design parameters on secure estimation, two sets of observers are considered, with eigenvalues of the error system matrices being placed more aggressively in the first simulation than in the second simulation. For the first experiment $\sigma(F_i^t), \sigma(F_i^r)$ and $\sigma(F_i^p)$ are given in Table 1; for the second, the spectra are scaled by 0.2, 0.1, and 0.1, respectively. In the following, a superscript s indicates that a variable is associated to the second simulation.

6.3. Non-stealthy FDI attack

Consider a non-stealthy FDI attack on the communication link between $\mathcal{S}_2$ and $\mathcal{S}_1$, altering the transmitted output $\bar{y}_{21}$ as in (3), which is active during [5, 15]s and [25, 35]s; whereas other transmitted outputs are not altered. It is considered that the attacker alters only the second component of y_2. Fig. 3 presents the profile of the second component of the attacker's injected data, $\Gamma_{21_{[2]}}$, and its estimates, $\hat{\Gamma}_{21_{[2]}}$ and $\hat{\Gamma}^s_{21_{[2]}}$, using Algorithm 1, based on the cascade of observers. The proposed scheme effectively enables $\mathcal{E}_1$ to estimate the injected data.

Fig. 4 presents the nine components of x_2, in addition to $\hat{x}_{2,1}$ and $\hat{x}^s_{2,1}$ associated with $\mathcal{O}_i^t$. As expected, the estimated states do not converge to the true value, due to the attack, as evident from the error dynamics given in (8). To address this, $\mathcal{O}_i^r$ is implemented to provide $\hat{x}^r_{2,1}$ and $\hat{x}^{r,s}_{2,1}$, as given in (12).

It is expected that $\hat{x}^r_{2,1} \rightarrow x_2$, as stated in Theorem 3.1. Fig. 4 indeed confirms this, and demonstrates how secure state estimation is successfully achieved. Finally, because $\hat{\Gamma}_{21_{[1]}}$ does not converge to zero (i.e., as seen in Fig. 3), the diagnosis logic is able to detect the FDI attack and isolate it to link 21. We expect selecting smaller eigenvalues for the second set of parameters to result in a slower response, as also shown in Figs. 3 and 4. Comparing the two obtained estimates, a smoother transient response can be observed by using the second set of parameters, but with longer settling time, resulting in a delay in detection. Also, a larger bias in $\hat{x}^s_{2,1}$ can be noted when attacks are active. Accordingly, it is evident that there is a trade-off between the selection of design parameters and the performance of the state estimation.

6.4. SFDI attack

Consider a cyber attack on the communication link from $\mathcal{S}_2$ to $\mathcal{S}_1$ at time $T_{A_{21}} = 20$ s, lasting for 10 s, implemented by an attacker with the dynamics in (4) as stated in Definition 2.3, which alters y_{21} as in (3). The attacker-designed input is chosen as $\alpha_{21} = [3.5 + 2.7\sin(t-20), 2 + 3(t-20), 1.8 - 0.4(t-20), 0, \ldots, 0]^\top \in \mathbb{R}^{81}$. As only one cyber attack is present, the condition on α_{21} in Lemma 4.2 is met, and therefore the attack is *stealthy* to the cascade of $\mathcal{O}_1^t$ and $\mathcal{O}_1^r$.

The profile of the injected data and its estimate are presented in Fig. 5. Even though $\Gamma_{21} \neq 0$, its estimate, $\hat{\Gamma}_{21}$, obtained by the cascade of $\mathcal{O}_i^t$ and $\mathcal{O}_i^r$, converges to zero (Lemma 4.2). This confirms the need for an additional step to evaluate whether there is any SFDI attack, and if so, to provide secure estimation despite the attack. In Fig. 6, we demonstrate the capability of the procedure illustrated in Algorithm 2. Note that, by definition of A_{12} and A_{13}, the conditions in Proposition 4.5 are met. In Fig. 6, the true value of $\bar{x}_1$ is shown as the black solid line; the estimate obtained by $\mathcal{O}_1^p$, $\hat{\bar{x}}_1$, is shown as the blue dashed line; finally, the

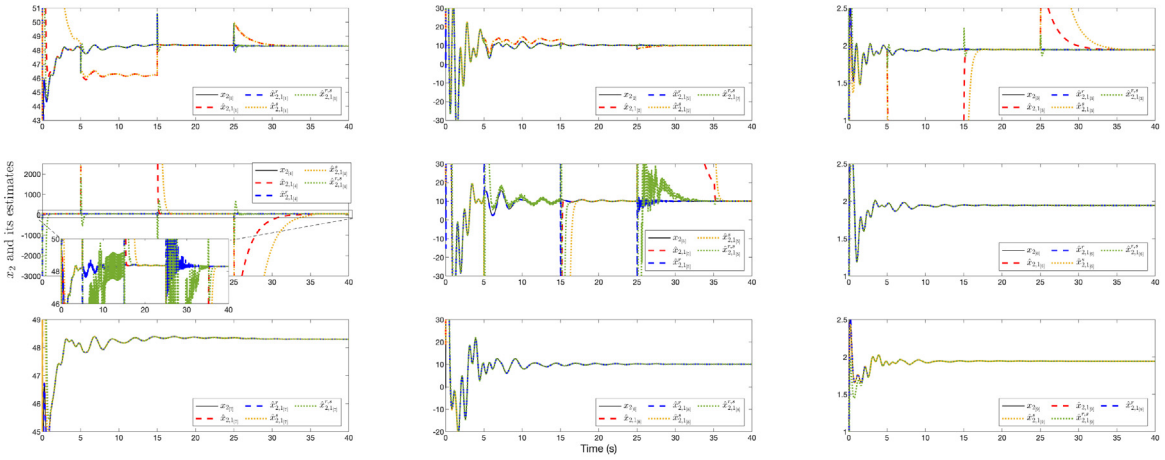

Fig. 4. Non-stealthy FDI attack: x_2, partitioned into its nine components, and their corresponding estimates, $\hat{x}_{2,1}$ and $\hat{x}^s_{2,1}$, obtained by $\mathcal{O}^t_1$, and $\hat{x}^r_{2,1}$ and $\hat{x}^{r,s}_{2,1}$ obtained by the cascade of $\mathcal{O}^t_1$ and $\mathcal{O}^r_1$.

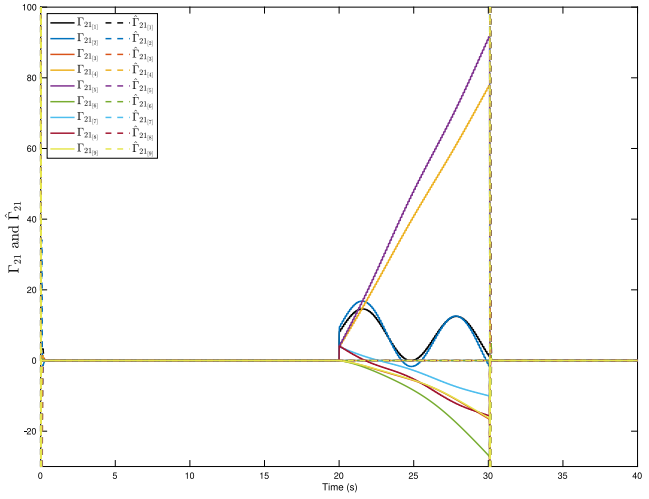

Fig. 5. SFDI attack: Γ_{21}, and estimate.

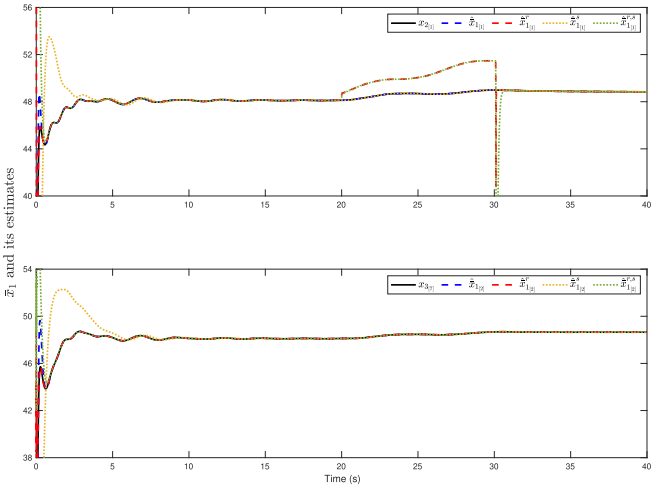

Fig. 6. SFDI attack: $\bar{x}_1$, partitioned into its two components, $\bar{x}_{1_{[1]}}$ (i.e., $x_{2_{[1]}}$) and $\bar{x}_{1_{[2]}}$ (i.e., $x_{3_{[7]}}$), and their estimates, $\hat{\bar{x}}_1$ and $\hat{\bar{x}}^s_1$, obtained by $\mathcal{O}^p_1$, and $\hat{\bar{x}}^r_1$ and $\hat{\bar{x}}^{r,s}_1$, obtained by the cascade of $\mathcal{O}^t_1$ and $\mathcal{O}^r_1$.

estimate $\hat{\bar{x}}^r_1$ is given as the red dashed line. From the figure, it can be seen that $\hat{\bar{x}}_1$ is unaffected by the attacker, thus it converges to the true value $\bar{x}_1$, whilst $\hat{\bar{x}}^r_1$ does not. This is to be expected, following our discussion in Section 4.

Partial estimation of x_2 and x_3 can be obtained (Proposition 4.3), enabling the designed scheme to isolate the attack (Proposition 4.5). While $\hat{\bar{x}}_1$ is a secure estimation of $x_{2_{[1]}}$, $\hat{\bar{x}}^r_{1_{[1]}}$ diverges from the true value as a result of exploiting $\tilde{y}_{21}$ through the cascade of $\mathcal{O}^t_1$ and $\mathcal{O}^r_1$. As the two estimates of $x_{2_{[1]}}$ differ, an SFDI attack on the communication link 21 is diagnosed, as per Proposition 4.4. Given that $\hat{\bar{x}}^r_{1_{[2]}} \to \hat{\bar{x}}_{1_{[2]}}$, the proposed scheme correctly diagnoses that no SFDI attack is present on the communication link 31. As a result of using the second set of parameters (with smaller eigenvalues), Fig. 6 illustrates that larger magnitudes of eigenvalues lead to faster response, though with larger overshoot.

7. Conclusion

In this paper, we address cyber security in large-scale systems, by specifically focusing on reconstruction/secure estimation, given non-stealthy and stealthy false data injection attacks on communication links between subsystems. We present algorithms that address two challenges: (1) reconstruction of system states, outputs, and attacker's injected data in the presence of non-stealthy FDI attacks, and (2) achieving state reconstruction in the case of SFDI attacks. These reconstructed values can be exploited for effective detection and isolation of cyber attacks in both attack classes as well as for providing secure control to enhance the system resilience against FDI attacks. Interesting future research directions include exploring simultaneous attack scenarios, and more complex large-scale system models that account for uncertainties, such as noise, disturbances, and parameter variations. Additionally, investigating nonlinear system's and attacker's dynamics presents a promising direction for further contributions.

Acknowledgment

Hanieh Tabatabaei and Ahmad W. Al-Dabbagh acknowledge financial support from the Natural Sciences and Engineering Research Council of Canada, through Discovery Grant RGPIN-2022-03687. Alexander J. Gallo acknowledges financial support by the

Research Council of Norway through the project AIMWind (grant ID 312486); Alexander J. Gallo is presently with Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy.

References

Al-Dabbagh, A. W., Barboni, A., & Parisini, T. (2020). Distributed detection and isolation of covert cyber attacks for a class of interconnected systems. *IFAC-PapersOnLine*, *53*(2), 772–777.

Ansari Rad, S., & Al-Dabbagh, A. W. (2025). Detection and isolation of covert cyberattacks using decentralized and distributed dynamic observers. *IEEE Transactions on Control of Network Systems*, *12*(1), 736–748.

Barboni, A., Al-Dabbagh, A. W., & Parisini, T. (2022). An event-triggered watermarking strategy for detection of replay attacks. *IFAC-PapersOnLine*, *55*(6), 317–322.

Barboni, A., & Parisini, T. (2020). Towards distributed accommodation of covert attacks in interconnected systems. In *The proceedings of the 59th IEEE conference on decision and control* (pp. 5731–5736).

Chang, Y. H., Hu, Q., & Tomlin, C. J. (2018). Secure estimation based Kalman filter for cyber–physical systems against sensor attacks. *Automatica*, *95*, 399–412.

Chen, J., Patton, R. J., & Zhang, H. (1996). Design of unknown input observers and robust fault detection filters. *International Journal of Control*, *63*(1), 85–105.

Gallo, A. J., Boem, F., & Parisini, T. (2021). Distributed cyber-attack isolation for large-scale interconnected systems. In *2021 European control conference* (pp. 48–53).

Gallo, A. J., Turan, M. S., Boem, F., Parisini, T., & Ferrari-Trecate, G. (2020). A distributed cyber-attack detection scheme with application to DC microgrids. *IEEE Transactions on Automatic Control*, *65*(9), 3800–3815.

Gao, Z., & Ding, S. X. (2007). Actuator fault robust estimation and fault-tolerant control for a class of nonlinear descriptor systems. *Automatica*, *43*(5), 912–920.

Gao, Z., Ding, S. X., & Ma, Y. (2007). Robust fault estimation approach and its application in vehicle lateral dynamic systems. *Optimal Control Applications and Methods*, *28*(3), 143–156.

Gao, Z., Liu, X., & Chen, M. Z. Q. (2016). Unknown input observer-based robust fault estimation for systems corrupted by partially decoupled disturbances. *IEEE Transactions on Industrial Electronics*, *63*(4), 2537–2547.

Griffioen, P., Weerakkody, S., & Sinopoli, B. (2021). A moving target defense for securing cyber-physical systems. *IEEE Transactions on Automatic Control*, *66*(5), 2016–2031.

Gupta, A., Langbort, C., & Başar, T. (2010). Optimal control in the presence of an intelligent jammer with limited actions. In *The proceedings of the 49th IEEE conference on decision and control* (pp. 1096–1101).

Hemsley, K. E., & Fisher, R. E. (2018). History of industrial control system cyber incidents. *Technical Report INL/CON-18-44411-Rev002*, Idaho Falls, United States: Idaho National Laboratory (INL).

Hoehn, A., & Zhang, P. (2016). Detection of covert attacks and zero dynamics attacks in cyber-physical systems. In *The proceedings of the 2016 American control conference* (pp. 302–307).

Hou, M., & Patton, R. J. (1998). Input observability and input reconstruction. *Automatica*, *34*(6), 789–794.

Kazemi, Z., Safavi, A. A., Arefi, M. M., & Naseri, F. (2022). Finite-time secure dynamic state estimation for cyber–physical systems under unknown inputs and sensor attacks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *52*(8), 4950–4959.

Keijzer, T., Ferrari, R. M. G., & Sandberg, H. (2023). Secure state estimation under actuator and sensor attacks using sliding mode observers. *IEEE Control Systems Letters*, *7*, 2071–2076.

Lan, J., & Patton, R. (2021). Asymptotic estimation of state, fault and perturbation for nonlinear systems and its fault-tolerant control application. *International Journal of Control, Automation and Systems*, *19*(3), 1175–1182.

Luenberger, D. G. (1964). Observing the state of a linear system. *IEEE Transactions on Military Electronics*, *8*(2), 74–80.

Ma, R., Hu, Z., Xu, L., & Wu, L. (2024). Distributed secure estimation against sparse false data injection attacks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *54*(5), 2685–2697.

Mo, Y., & Sinopoli, B. (2009). Secure control against replay attacks. In *The proceedings of the 47th annual allerton conference on communication, control, and computing* (pp. 911–918).

Mousavinejad, E., Ge, X., Han, Q.-L., Yang, F., & Vlacic, L. (2021). Resilient tracking control of networked control systems under cyber attacks. *IEEE Transactions on Cybernetics*, *51*(4), 2107–2119.

Smith, R. S. (2015). Covert misappropriation of networked control systems. *IEEE Control System Magazine*, *35*(1), 82–92.

Tan, J., Zheng, H., Meng, D., Wang, X., & Liang, B. (2023). Active input design for simultaneous fault estimation and fault-tolerant control of LPV systems. *Automatica*, *151*, Article 110903.

Teixeira, A., Shames, I., Sandberg, H., & Johansson, K. H. (2012). Revealing stealthy attacks in control systems. In *The proceedings of the 50th annual allerton conference on communication, control, and computing* (pp. 1806–1813).

Teixeira, A., Shames, I., Sandberg, H., & Johansson, K. H. (2014). Distributed fault detection and isolation resilient to network model uncertainties. *IEEE Transactions on Cybernetics*, *44*(11), 2024–2037.

Teixeira, A., Shames, I., Sandberg, H., & Johansson, K. (2015). A secure control framework for resource-limited adversaries. *Automatica*, *51*, 135–148.

Tucci, M., Meng, L., Guerrero, J. M., & Ferrari-Trecate, G. (2018). Stable current sharing and voltage balancing in DC microgrids: A consensus-based secondary control layer. *Automatica*, *95*, 1–13.

Wang, L., Cao, X., Zhang, H., Sun, C., & Zheng, W. X. (2022). Transmission scheduling for privacy-optimal encryption against eavesdropping attacks on remote state estimation. *Automatica*, *137*, Article 110145.

Yan, J.-J., & Yang, G.-H. (2023). Secure state estimation of nonlinear cyber-physical systems against DoS attacks: A multiobserver approach. *IEEE Transactions on Cybernetics*, *53*(3), 1447–1459.

Yang, C., Chu, Z., Ma, L., Wang, G., & Dai, W. (2023). Joint watermarking-based replay attack detection for industrial process operation optimization cyber-physical systems. *IEEE Transactions on Industrial Informatics*, *19*(8), 8910–8922.

Zhao, D., & Polycarpou, M. M. (2022). Distributed fault accommodation of multiple sensor faults for a class of nonlinear interconnected systems. *IEEE Transactions on Automatic Control*, *67*(4), 2092–2099.

Hanieh Tabatabaei received the B.Sc. and the M.Sc. degrees in electrical engineering in 2016 and in 2020, respectively, both from the Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. She is currently working toward the Ph.D. degree in electrical engineering at The University of British Columbia, Kelowna, Canada. Her research interests relate to distributed and networked systems and consists of model-based algorithms for cyber security, fault diagnosis, robust control, and remote estimation.

Alexander J. Gallo received the M.Eng. in Electrical and Electronic Engineering from Imperial College London, London, UK, in 2016. He received his Ph.D. degree in Control Engineering from Imperial College London, London, UK, in 2021. From 2021 to 2024 he was a postdoctoral researcher at the Delft Center for Systems and Control, Delft University of Technology, Delft, the Netherlands. Since Sep 2024, he is a postdoctoral researcher at the Politecnico di Milano, Milan, Italy.

His main research interests include distributed cyber-security and fault tolerant control for large-scale interconnected systems, with a particular focus on energy distribution networks, as well as health-aware and fault tolerant control of wind turbines.

Ahmad W. Al-Dabbagh received the Ph.D. degree in control systems from the University of Alberta, Edmonton, Canada, in 2018. He held postdoctoral fellowships at the University of Alberta, the University of Toronto, and Imperial College London, from 2018 to 2020. He is presently an Assistant Professor and the Principal's Research Chair in Control Systems with the School of Engineering, The University of British Columbia, Kelowna, Canada. He has been serving as a Technical Editor for IEEE/ASME Transactions on Mechatronics, and an Associate Editor for IEEE Transactions on Automation Science and Engineering, IEEE Transactions on Cybernetics, and IEEE Control Systems Society Conference Editorial Board. He is a Senior Member of IEEE and ISA, and is a recipient of ISA's Excellence in Education Award in 2024. His research interests relate to distributed and networked systems, and include fault diagnosis, cybersecurity, and alarm management.