Robust Detection and Reconstruction of State and Sensor Attacks for Cyber-Physical Systems using Sliding Modes

Maria Letizia Corradini, Andrea Cristofaro

Scuola di Scienze e Tecnologie, Università di Camerino, via Madonna delle Carceri, 62032 Camerino (MC), Italy, email: {letizia.corradini, andrea.cristofaro}@unicam.it

Abstract: This paper addresses the problem of detection and reconstruction of cyber-attacks corrupting states and/or outputs of a linear Cyber-Physical System. Robust state/sensor attack observers are designed able both to work as detection monitors with guaranteed performances, and to reconstruct the attacks within a finite-time. Detection and reconstruction are performed robustly with respect to bounded modeling errors possibly affecting the state equation. Compensation of attacks is also addressed for square plants. An extensive simulation study using test-cases taken from the literature is shown to support the theoretical findings.

1. Introduction

The majority of critical infrastructures supporting many domains of modern society is based on CyberPhysical Systems (CPS), i.e. systems integrating physical processes, computational resources and communication capabilities [1] (Fig. 3). Transportation networks, power generation and distribution networks, industrial automation systems are examples of such systems, where control of physical plant is mediated by and integrated with a wireless communication network. If this integration can improve efficiency, it nonetheless make the system more vulnerable to attacks launched in the cyber-domain [2]. Against such attacks, information security techniques may be not enough for controlling systems comprising physical processes [3]. Recent real-world attacks targeting physical plants (such as the StuxNet attack [4] in 2010 that targeted Siemens' supervisory control and data acquisition systems, and the Maroochy water bleach [5]) raised the problem of cyber-physical security, suggesting that information security mechanisms have to be complemented with specifically designed control systems, possibly resilient against attacks and/or equipped with attack monitors [3, 6, 7, 8].

As a matter of fact, resiliency of a control system with respect to faults and failures has been widely addressed in the past, and a vast bibliography is available on fault detection, isolation and tolerance [9] [10]. Though some methods could be partially borrowed, peculiarities of the considered problem of cyberphysical security need to be considered [11]. As discussed in [6], residual based algorithms (within the deterministic framework) could be inappropriate in view of the large number of failure modes potentially available. Residual filters based on standard Luemberger-like observers could be used as attack detection monitors [3], but they show severe limitations, due to their asymptotic performance and their sensitivity to modeling disturbances, naturally bounded [6]. The adoption of a distributed monitoring system has been proposed for power networks in [12], where a bank of Unknown Input Observers is designed to detect and isolate attacks on nodes and attacks on the communication between nodes. These filters are computationally expensive and require some structural conditions. Replay attacks have been addressed in [13], and watermarked input have been proposed for detection. False data injection attacks have been investigated in [14], where it has been shown that undetectable false data attacks can be launched even with limited resources available to the attacker.

In order to increase the robustness of cyberphysical systems, appropriate tools are needed to first understand and detect, and then to protect them against cyber attacks. Reconstruction of attacks is therefore fundamental, beside prompt detection, to guarantee continuity of service of CPSs often supporting critical infrastructures. For this reason, both attacks detection and reconstruction should be focused on. Furthermore, the presence of possible perturbations and/or modeling errors should be explicitly accounted for during detection and reconstruction, in order to achieve a satisfactory level of robustness when dealing with real life cases. Prompt detection and reconstruction using an adaptive approach has been addressed in the very recent paper [15], where adaptive sliding mode observers are designed coupled with a parameter estimator and a robust differentiator. Residual signals are used for detection and ultimately bounded observation errors are shown to be ensured, though the presence of possible external perturbations has not been considered.

Motivation: The previous discussion suggests that attack monitoring devices can be effectively used for protecting cyberphysical systems if finite-time (not asymptotic) performance is ensured, in order to guarantee prompt detection, and if some degree of robustness of the device is ensured with respect to possible perturbations and/or modeling errors affecting the system. Moreover, prompt attack reconstruction can be helpful in counteracting and accounting for the attack itself.

Framework: This article adopts the unified modeling framework for CPSs, based on a control-theoretic approach to cyberphysical security, recently summarized in the tutorial [3]. Cyberphysical systems under attack are modeled as linear systems subject to unknown inputs altering the state (state attack) and the measurements (sensor attack).

Problem Statement: The addressed problem is to design a state/sensor attack observer able to either work as a detection monitor (i.e. a device able to detect attacks in a CPS on the basis of its knowledge of the system matrices and of the measurements) with guaranteed performances, and to reconstruct the attack itself. In addition, bounded external perturbations or modeling errors are assumed to possibly affect the state equation, since attack detection and reconstruction has to be performed robustly with respect to such perturbations.

Main contributions of the paper:

- A design solution is proposed of robust state and/or sensor attack monitors providing finite-time performance. In addition to detection, attack reconstruction is also performed and achieved within an arbitrary finite time with arbitrary precision.
- The proposed attacks detection monitor is built designing an observer for a given state or output equation. The designed observers are proved to show insensitivity with respect to possible bounded perturbations affecting the state equation.
- The proposed attack observers are very easy to be implemented. The chosen methodology is based on sliding-mode based techniques [16], and require the availability of an upper bound, even rough, for the attack.
- An extensive simulation study using test-cases taken from the literature has been performed, showing satisfactory results even in the presence of perturbations affecting the state equation.

The paper is organized as follows: in Section II the model of the system is presented, some technical assumptions are stated and the observer is introduced. Detection and reconstruction of state attacks is addressed in Section 3, where the main technical results are proved. The extension to the case of sensor attacks is reported in Section 4. Finally, compensation of attacks in the case of tracking requirements is shortly addressed in Section 5 for square plants. Each section contains a simulation study supporting the theoretical results and proving the efficiency of the proposed method. Finally, some conclusions are drawn in Section 6.

2. Problem statement

2.1. Motivating example

A large variety of industrial networked structures can be classified as cyber-physical systems. Typical examples are smart grids, power networks, consensus networks, water networks, among many others. A cyber-physical model of a power network is well described by a state $\boldsymbol{x}$ that includes rotor angles and rates for each generator, as well as voltage angles at the buses. In a real-world scenario only a small groups of rotor angles and rates is directly measured, and typical attacks aim at injecting disturbance signals that mainly affect the sensorless generators. In this regard, let us consider the model of the US Western Electricity Coordinating Council (WECC) power system [3] described in Figure 2. The system is characterized by three generators and six buses, with a single generator, i.e. g_1, being directly monitored. Two coordinated deception attacks affect the buses b_4 and b_5. The model of WECC power network under attack can be successfully recast in terms of a state-space representation, this yielding the linear system

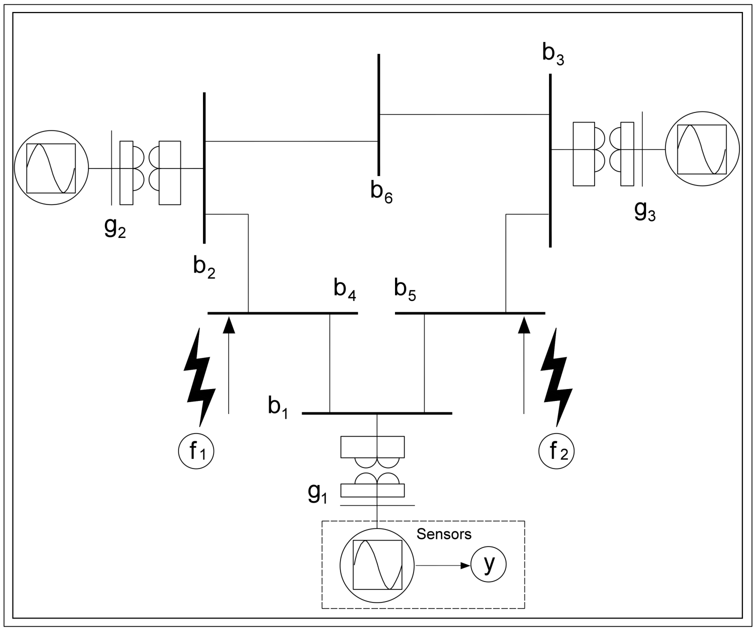


Fig. 1. *US Western Electricity Coordinating Council power system*

$$\begin{aligned} \dot{\delta}(t) &= \omega(t) \\ M_g\dot{\omega}(t) &= -L_{gg}\delta(t) - D_g\omega(t) - L_{gl}\theta(t) + P_\omega \end{aligned} \tag{1}$$

together with the algebraic constraint

$$L_{lg}\delta(t) + L_{ll}\theta(t) = P_\theta, \tag{2}$$

where $\delta, \omega \in \mathbb{R}^3$ denote, respectively, generator rotor angles and frequencies, $\theta \in \mathbb{R}^6$ is the vector of the voltage angles at the buses, $M_g, D_g \in \mathbb{R}^{3\times 3}$ are inertia and damping matrices, $P_\omega \in \mathbb{R}^3$ and $P_\theta \in \mathbb{R}^6$ are known inputs, corresponding to mechanical torque and power demand. The laplacian matrix $\mathcal{L} \in \mathbb{R}^{9\times 9}$ of the network graph is given by

$$\mathcal{L} = \begin{bmatrix} L_{gg} & L_{gl} \\ L_{lg} & L_{ll} \end{bmatrix}. \tag{3}$$

The attacks modify the loads at buses b_4 and b_5, according to

$$\theta(t) = \theta_{nom}(t) + [0\ 0\ 0\ f_1(t)\ f_2(t)\ 0]^T,$$

where θ_{nom} is the vector of nominal voltage angles. The case-study of such power network will be further discussed in Section 3.3, where the proposed framework for state attack detection and reconstruction is tested. It is worth recalling that, again following [3], a water network can be represented analogously, and can be described by state variables such as pressures at reservoirs, tanks and junctions. In this scenario, attacks can be either physical, e.g. subtraction of waters from a reservoir, or cyber, e.g. corruption of pressure measurements.

The remaining part of the paper is devoted at answering the following questions concerning the CPS in the general form (4).

Addressed Problem: *How to design detection filters, robust with respect to bounded perturbations affecting the state equation, able to perceive within a finite time the presence of state or output attacks ? How to reconstruct the profile of attacks?*

2.2. A control theoretic framework

A wide class of CPSs can be described by a linear time invariant continuous-time system [3] of the form

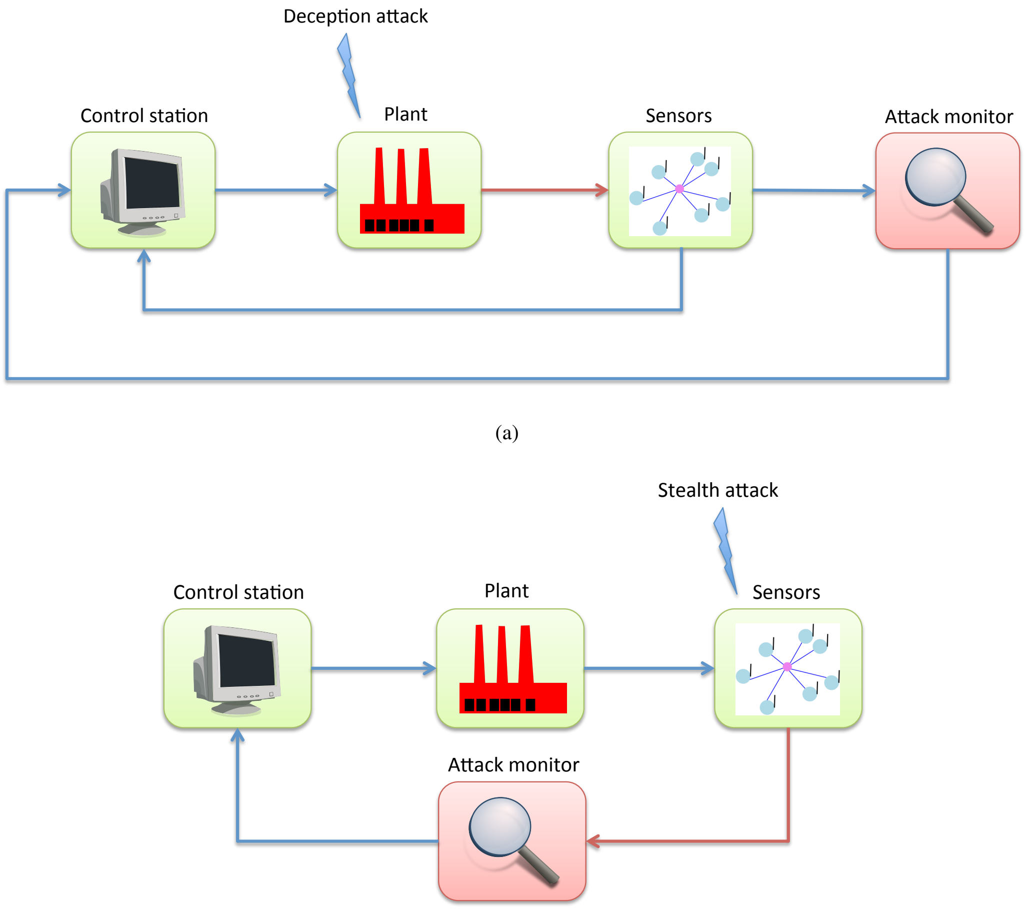
$$\begin{cases} \dot{\boldsymbol{\xi}}(t) = \bar{\boldsymbol{A}}\boldsymbol{\xi}(t) + \bar{\boldsymbol{B}}_u\boldsymbol{u}(t) + \bar{\boldsymbol{B}}_f f(t) + \bar{\boldsymbol{D}}_d\boldsymbol{d}(t) \\ \boldsymbol{y}(t) = \bar{\boldsymbol{C}}\boldsymbol{\xi}(t) + \bar{\boldsymbol{D}}_u\boldsymbol{u}(t) + \bar{\boldsymbol{D}}_f f(t) \end{cases} \tag{4}$$

where $\boldsymbol{\xi} \in \mathbb{R}^n$, $\boldsymbol{u} \in \mathbb{R}^m, y \in \mathbb{R}^p$, $p \geq m$, and the matrices $\bar{\boldsymbol{A}}, \bar{\boldsymbol{B}}_u, \bar{\boldsymbol{B}}_f$ have appropriate dimensions. The term $\bar{\boldsymbol{B}}_f f(t)$ models the effect of state attacks against the CPS, while the term $\bar{\boldsymbol{D}}_f f(t)$ models an output attack corrupting directly the measurement vector. In addition, the term $\bar{\boldsymbol{D}}_d\boldsymbol{d}(t)$ represents possible external perturbations or modelling errors. It is worth noticing that, in the above model, the state $\boldsymbol{\xi}$ includes both physical and cyber variables. The variable $\boldsymbol{u}$ represents known inputs and the vector $\boldsymbol{y}$ is the collection of all measured quantities.

A scalar attack is considered since the proposed approach is aimed at designing an attack detection monitor for a (or a set of) selected state components. However, in the case of multiple attacks, a bank of parallel monitors can be designed in a natural way. The attack and measurements sets are assumed to be selected such that the following assumption is satisfied.

Assumption 1. For the CPS (4) the following holds:

- The system is completely observable;
- $\bar{\boldsymbol{C}}\bar{\boldsymbol{B}}_f \neq 0$;



(a)

(b)

Fig. 2. *Attack monitors in the loop: (a) State attack; (b) Sensor attack.*

- $\bar{\boldsymbol{B}}_u$ is full column rank;
- The invariant zeros of $(\bar{\boldsymbol{A}}, \bar{\boldsymbol{B}}_u, \bar{\boldsymbol{C}}, \bar{\boldsymbol{D}}_u)$ are stable.

Assumption 2. Attacks are assumed detectable. As well known, the existence of undetectable attacks for the system $(\bar{\boldsymbol{A}}, \bar{\boldsymbol{B}}_f, \bar{\boldsymbol{C}}, \bar{\boldsymbol{D}}_f)$ is equivalent to the existence of invariant zeros of the same attack/measurements system.

Under Assumption 1, there exists a linear change of coordinate $\boldsymbol{x} = \boldsymbol{T}\boldsymbol{\xi}$ such that

$$\begin{cases} \dot{\boldsymbol{x}}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}_u\boldsymbol{u}(t) + \boldsymbol{B}_f f(t) + \boldsymbol{D}_d\boldsymbol{d}(t) \\ \boldsymbol{y}(t) = \boldsymbol{C}\boldsymbol{x}(t) + \boldsymbol{D}_u\boldsymbol{u}(t) + \boldsymbol{D}_f f(t) \end{cases} \tag{5}$$

where

$$\begin{aligned} &\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}; \quad \boldsymbol{B}_f = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{B} \end{bmatrix}; \quad \boldsymbol{B}_u = \begin{bmatrix} \boldsymbol{B}_1 \\ \boldsymbol{B}_2 \end{bmatrix}; \\ &\boldsymbol{C} = [0 \quad \boldsymbol{I}_p]; \quad \boldsymbol{D}_d = \begin{bmatrix} \boldsymbol{D}_1 \\ \boldsymbol{D}_2 \end{bmatrix}; \quad \bar{\boldsymbol{D}}_f = \boldsymbol{D}_f, \bar{\boldsymbol{D}}_u = \boldsymbol{D}_u \end{aligned} \tag{6}$$

with $\boldsymbol{B} \in \mathbb{R}^{p\times 1}$ and $\boldsymbol{A}_{11} \in \mathbb{R}^{n-p\times n-p}$ Hurwitz in view of Assumption 1. $\boldsymbol{I}_p$ is the p-dimensional identity matrix.

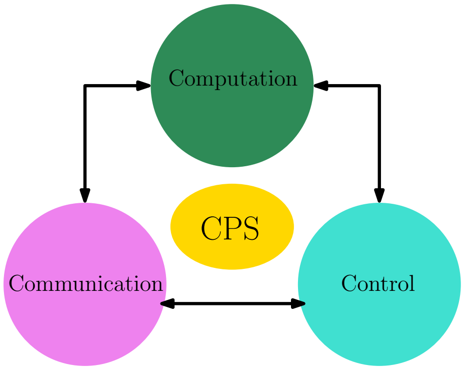


Fig. 3. *Cyber-Physical System*

2.3. Model of attacks

Several types of attacks might compromise the safety of a CPS, involving both the physical and the cyber components of the plant. In this paper we focus on deception attacks, affecting the state of the system, and stealth attacks, involving the sensors.

Deception attacks aim at intentionally compromising the system stability or altering the nominal regime of some of the states for malicious scopes [12, 15]. Coordinated attacks might also occur, and they are often designed in such a way the detection results to be highly challenging or nearly unfeasible. The interested reader may refer to [3], where an extensive coverage of attack detection issues is proposed in terms of geometric control theory. Stealth (or sensor) attacks are instead intended to alter the system measurements [6, 15]: this can be the primary objective of attackers as well as an action performed to hide the effects of deception attacks in order to delay the detection and make them even more severe. Summarizing, and referring to the system model (5), Table 1 below illustrates the scheme of possible attacks that can be handled with the proposed paradigm.

Table 1 State and sensor attacks.

Type of attack	$\mathbf{B}_f$	$\mathbf{D}_f$
Deception attack	$\neq 0$	$= 0$
Stealth attack	$= 0$	$\neq 0$

The following model for attack signals is considered both in the case of state and of sensors attacks :

$$f(t) = \delta(\boldsymbol{x}, t) \tag{7}$$

where $\delta(\boldsymbol{x}, t)$ is un unknown but bounded function satisfying the following assumption:

Assumption 3. Bounds are available for the function $\delta(\boldsymbol{x}, t)$ and for its time derivative:

$$|\delta(\boldsymbol{x}, t)| \leq \rho_1(\boldsymbol{x}, t); \quad |\dot{\delta}(\boldsymbol{x}, t)| \leq \rho_2(\boldsymbol{x}, t) \tag{8}$$

Assumption 4. Deception and stealth attacks do not occur simultaneously, i.e. $\|\mathbf{B}_f\| \cdot \|\mathbf{D}_f\| = 0$.

3. Detection and reconstruction of state attacks

Consider the case of state attacks only ($\boldsymbol{D}_f = 0$). The following standard observer is designed for tackling the case of state attacks:

$$\begin{cases} \dot{\hat{\boldsymbol{x}}}_1(t) & = \boldsymbol{A}_{11}\hat{\boldsymbol{x}}_1(t) + \boldsymbol{A}_{12}\hat{\boldsymbol{x}}_2(t) + \boldsymbol{B}_1\boldsymbol{u}(t) + \boldsymbol{A}_{12}(\boldsymbol{y}(t) \\ & -\hat{\boldsymbol{y}}(t)) \\ \dot{\hat{\boldsymbol{x}}}_2(t) & = \boldsymbol{A}_{21}\hat{\boldsymbol{x}}_1(t) + \boldsymbol{A}_{22}\hat{\boldsymbol{x}}_2(t) + \boldsymbol{B}_2\boldsymbol{u}(t) + \boldsymbol{B}\hat{f}(t) \\ & +\boldsymbol{L}(\boldsymbol{y}(t) - \hat{\boldsymbol{y}}(t)) \\ \hat{\boldsymbol{y}}(t) & = \hat{\boldsymbol{x}}_2(t) \end{cases}$$

where $\boldsymbol{L} \in \mathbb{R}^{p \times p}$ is chosen such that $\boldsymbol{A}_{22} - \boldsymbol{L} = -\boldsymbol{H}$, with $\boldsymbol{H}$ positive definite, and $\hat{f}$ is an estimate of the possible attack, to be designed later. The estimation error dynamics are given by

$$\begin{cases} \dot{\boldsymbol{e}}_1(t) & = \boldsymbol{A}_{11}\boldsymbol{e}_1(t) + \boldsymbol{D}_1\boldsymbol{d}(t) \\ \dot{\boldsymbol{e}}_2(t) & = \boldsymbol{A}_{21}\boldsymbol{e}_1(t) - \boldsymbol{H}\boldsymbol{e}_2(t) + \boldsymbol{B}(f(t) - \hat{f}(t)) + \boldsymbol{D}_2\boldsymbol{d}(t) \end{cases} \tag{9}$$

where $\boldsymbol{e}_1 = \boldsymbol{x}_1 - \hat{\boldsymbol{x}}_1$, $\boldsymbol{e}_2 = \boldsymbol{x}_2 - \hat{\boldsymbol{x}}_2$. In the following section some issues regarding the UIO design will be discussed.

The placement of the state attack monitor in the CPS loop is illustrated in Figure . As shown, the information extracted and processed by the sensor network is fed to the monitor, whose output can be fed back to the controller with the aim of compensating for the attack effects.

3.1. The sliding surface

Consider the following sliding surface

$$\boldsymbol{s}(t) = \boldsymbol{e}_2(t) + \boldsymbol{H}\int_0^t \boldsymbol{e}_2(\tau)d\tau \tag{10}$$

and its time derivative

$$\dot{\boldsymbol{s}}(t) = \dot{\boldsymbol{e}}_2(t) + \boldsymbol{H}\boldsymbol{e}_2(t) \tag{11}$$

The following technical results can be proved.

Lemma 3.1. Consider the function $g(t) = \boldsymbol{s}(t)^T\dot{\boldsymbol{s}}(t)$ with a given initial value $g(t_0) = g_0 < 0$. If the condition $\dot{g}(t) < 0$ is imposed, the variable $\boldsymbol{s}(t)$ decreases in norm and intercepts the layer $||\boldsymbol{s}(t)|| \leq \epsilon$ of arbitrary width ϵ within a finite time $t_\epsilon = \dfrac{\|\boldsymbol{s}(t_0)\|^2 + 2|g_0|t_0 - \epsilon^2}{2|g_0|}$.

Proof. It is straightforward to see that the imposition of the condition $\dot{g}(t) < 0$, starting from a negative initial condition $g_0 = -|\boldsymbol{s}(t_0)^T\dot{\boldsymbol{s}}(t_0)|$, guarantees that the function $g(t) = \boldsymbol{s}(t)^T\dot{\boldsymbol{s}}(t)$ is always negative and decreasing. This implies that, since both $\boldsymbol{s}(t)$ and $\dot{\boldsymbol{s}}(t)$ are bounded away from zero, the variable $\boldsymbol{s}(t)$ is forced to decrease in norm, i.e.

$$\frac{d}{dt}\|\boldsymbol{s}(t)\|^2 = 2g(t) < 2g_0 = -2|g_0| < 0.$$

An upper bound on the time t_ϵ needed to reach the layer of width ϵ can be derived by integrating the latter differential inequality, which gives

$$\|\boldsymbol{s}(t)\|^2 \leq -2|g_0|(t-t_0) + \|\boldsymbol{s}(t_0)\|^2,$$

and then imposing the identity

$$-2|g_0|(t-t_0) + \|\boldsymbol{s}(t_0)\|^2 = \epsilon^2$$

that provides the claimed value $t_\epsilon = \dfrac{\|\boldsymbol{s}(t_0)\|^2 + 2|g_0|t_0 - \epsilon^2}{2|g_0|}$. ■

Proposition 3.1. Consider the function $g(t) = \boldsymbol{s}(t)^T \dot{\boldsymbol{s}}(t)$, with a given initial condition $g(0) = g_0$, and fix $\eta \in \mathbb{R}^+$. As long as $\dot{g}(t) < -\eta$ is imposed for $||\boldsymbol{s}(t)|| \geq \epsilon$, where $\epsilon \in \mathbb{R}^+$ is an arbitrary parameter, the variable $\boldsymbol{s}(t)$ is driven toward the layer $||\boldsymbol{s}(t)|| \leq \epsilon$ within a finite time t_s.

Proof. According to Lemma 3.1, starting from a negative initial condition $g_0 = -|\boldsymbol{s}(0)^T \dot{\boldsymbol{s}}(0)|$ and imposing $\dot{g}(t) < 0$, the variable $\boldsymbol{s}(t)$ is guaranteed to be decreasing in norm and any arbitrary precision $||\boldsymbol{s}(t)|| = \epsilon$ is achieved within a finite time t_ϵ.
Consider now the opposite case when a positive initial condition is given, e.g. $g_0 = \boldsymbol{s}(0)^T \dot{\boldsymbol{s}}(0) > 0$, and assume the uniform bound $\dot{g}(t) < -\eta$. Let us show that the condition $g(t) = 0$ is achieved within a finite time t_0. To this purpose, select an arbitrary time $t_0 > 0$. By the Lagrange Theorem, one has that $g(t_0) - g(0) = \dot{g}(t^*)t_0$, for some $t^* \in (0,\ t_0)$. Now, using the inequality $\dot{g}(t) < -\eta$, one gets $g(t_0) - g(0) \leq -\eta t_0$ and then, taking $t_0 \overset{\text{def}}{=} \dfrac{g(0)}{\eta}$, it has been proved that the condition $g(t) < 0$ is attained for any $t \geq t_0$. From this point, the previous Lemma can be invoked to infer that $\boldsymbol{s}(t)$ enters the layer $\|\boldsymbol{s}(t)\| \leq \epsilon$ within a finite time $t_s \overset{\text{def}}{=} t_0 + t_\epsilon$. ■

Remark 3.1. From the above arguments it follows that the imposition of the condition $\dot{g}(t) < -\eta$ outside a layer of arbitrary width $||\boldsymbol{s}(t)|| > \epsilon$ ensures that the variable $\boldsymbol{s}(t)$ is driven toward the same layer within a finite time and maintained in a neighborhood of it for all the future times. Moreover, the previous discussion proves that the variable $g(t)$ is bounded, since it is not allowed to decrease or increase indefinitely. This in turn guarantees that the variables $\boldsymbol{s}(t)$ and $\dot{\boldsymbol{s}}(t)$ are bounded and, as a consequence, the variable $\dot{\boldsymbol{e}}_2(t)$ is bounded in norm, too, this implying the boundedness of the observation error $f(t) - \hat{f}(t)$ according to (9).

3.2. State Attack Monitor design

The previous results can be used to design an observer to perform an effective monitoring of possible state attacks. Recall that, defining $\aleph = \lambda_{min}(-\boldsymbol{A}_{11})$ and recalling Assumption 1, it holds

$$||\boldsymbol{e}_1(t)|| \leq k_1 e^{-\aleph t}||\boldsymbol{e}_1(0)|| + k_2 \int_0^t e^{-\aleph(t-\tau)}||\boldsymbol{D}_d \boldsymbol{d}(\tau)||d\tau$$

for appropriate constants $k_1, k_2 > 0$, where the initial condition $\boldsymbol{e}_1(0)$ is unknown.

Assumption 5. A bound is available for the initial condition $e_1(0)$:

$$|\boldsymbol{e}_1(0)| \leq \rho_e \tag{12}$$

Moreover, bounded external perturbations are considered, i.e. such that:

$$||\boldsymbol{d}(t)|| \leq \rho_d(t) \quad ||\dot{\boldsymbol{d}}(t)|| \leq \rho_d'(t) \tag{13}$$

It follows that a constant ρ exists such that $||\boldsymbol{e}_1(t)|| \leq \rho$.

Theorem 3.1. It is given the CPS (5) under Assumptions 1-5, subject to state attacks. The following attack observer:

$$\begin{aligned}\dot{\hat{f}}(t) = \gamma & \frac{k}{k\boldsymbol{s}(t)^T\boldsymbol{B} + \epsilon sign(\boldsymbol{s}(t)^T\boldsymbol{B})} \cdot \\ & \Big(\alpha(\boldsymbol{x},t)||\boldsymbol{s}(t)|| + \beta(\boldsymbol{x},t)^2 + ||\boldsymbol{B}||^2\hat{f}^2(t) \\ & +2\beta(\boldsymbol{x},t)||\boldsymbol{B}|| \cdot |\hat{f}(t)| + \eta\Big) \quad if|\boldsymbol{s}(t)^T\boldsymbol{B}| > \epsilon \end{aligned} \tag{14}$$

with $\epsilon > 0, \gamma > 1, \eta > 0$ arbitrary, and

$$\begin{aligned}\alpha(t,\boldsymbol{x}) & \overset{\text{def}}{=} ||\boldsymbol{A}_{21}|| \left(\boldsymbol{A}_{11}||\rho + ||\boldsymbol{D}_1||\rho_d(t)\right) + ||\boldsymbol{D}_2\rho'_d(t)|| \\ & \quad + ||\boldsymbol{B}||\rho_2(\boldsymbol{x},t); \\ \beta(t,\boldsymbol{x}) & \overset{\text{def}}{=} ||\boldsymbol{A}_{21}||\rho + ||\boldsymbol{B}||\rho_1(\boldsymbol{x},t) + ||\boldsymbol{D}_2||\rho_d(t); \end{aligned} \tag{15}$$

ensures that:

- the variable $\boldsymbol{s}(t)$ is driven in norm to a layer of arbitrary width $\dfrac{\epsilon}{||\boldsymbol{B}||}$ within a finite time;
- the observation error $\boldsymbol{e}_2(t)$ is bounded;
- the observation error $f(t) - \hat{f}(t)$ is bounded.

Proof. Consider the function $g(t) = \boldsymbol{s}(t)^T\dot{\boldsymbol{s}}(t)$, with an arbitrary initial condition $g(t_0) = g_0$. Imposing the condition $\dot{g}(t) < -\eta$, $\eta > 0$ arbitrary, one gets

$$\begin{aligned} & \boldsymbol{s}(t)^T\left[\boldsymbol{A}_{21}\left(\boldsymbol{A}_{11}\boldsymbol{e}_1(t) + \boldsymbol{D}_1\boldsymbol{d}(t)\right) + \boldsymbol{B}(\dot{f}(t) - \dot{\hat{f}}(t)) + \boldsymbol{D}_2\dot{\boldsymbol{d}}(t)\right] \\ & \quad < -\Big(\boldsymbol{A}_{21}\boldsymbol{e}_1(t) + \boldsymbol{B}(f(t) - \hat{f}(t))\Big)^T\Big(\boldsymbol{A}_{21}\boldsymbol{e}_1(t) + \boldsymbol{B}(f(t) - \hat{f}(t))\Big) - \eta \end{aligned} \tag{16}$$

i.e.

$$\begin{aligned} & \boldsymbol{s}(t)^T\boldsymbol{B}\dot{\hat{f}}(t) > \boldsymbol{s}(t)^T\left[\boldsymbol{A}_{21}\left(\boldsymbol{A}_{11}\boldsymbol{e}_1(t) + \boldsymbol{D}_1\boldsymbol{d}(t)\right) + \boldsymbol{D}_2\dot{\boldsymbol{d}}(t)\right] \\ & \quad + \Big(\boldsymbol{A}_{21}\boldsymbol{e}_1(t) + \boldsymbol{B}(f(t) - \hat{f}(t))\Big)^T\Big(\boldsymbol{A}_{21}\boldsymbol{e}_1(t) + \boldsymbol{B}(f(t) - \hat{f}(t))\Big) + \eta \end{aligned}$$

Consider the case $||\boldsymbol{s}(t)|| \cdot ||\boldsymbol{B}|| \geq \boldsymbol{s}(t)^T\boldsymbol{B} > \epsilon$, i.e. $||\boldsymbol{s}(t)|| > \dfrac{\epsilon}{||\boldsymbol{B}||} \overset{\text{def}}{=} \epsilon_1$. A strongest condition for $\boldsymbol{s}(t)^T\boldsymbol{B} > \epsilon$ is given by

$$\begin{aligned} (\boldsymbol{s}(t)^T\boldsymbol{B} - \frac{\epsilon}{k})\dot{\hat{f}}(t) & > ||\boldsymbol{s}(t)||\alpha(\boldsymbol{x},t) + \beta(\boldsymbol{x},t)^2 \\ & + B^2\hat{f}^2(t) + 2\beta(\boldsymbol{x},t)B|\hat{f}(t)| + \eta \end{aligned} \tag{17}$$

with $k > 1$. Therefore:

$$\begin{aligned}\dot{\hat{f}}(t) & > \frac{k||\boldsymbol{s}(t)||}{k\boldsymbol{s}(t)^T\boldsymbol{B} - \epsilon}\alpha(\boldsymbol{x},t) + \frac{k}{k\boldsymbol{s}(t)^T\boldsymbol{B} - \epsilon}\big(\beta(\boldsymbol{x},t)^2 \\ & +||\boldsymbol{B}||^2\hat{f}^2(t) + 2\beta(\boldsymbol{x},t)||\boldsymbol{B}|||\hat{f}(t)| + \eta\Big) \end{aligned} \tag{18}$$

and the expression (14) follows. Consider now the case $\boldsymbol{s}(t)^T\boldsymbol{B} < -\epsilon$. From (17) a strongest condition for $\boldsymbol{s}(t)^T\boldsymbol{B} < -\epsilon$ is given by

$$\begin{aligned}
&(\boldsymbol{s}(t)^T\boldsymbol{B} + \frac{\epsilon}{k})\dot{\hat{f}}(t) > ||\boldsymbol{s}(t)||\alpha(\boldsymbol{x},t) + \beta(\boldsymbol{x},t)^2 \\
&\quad + B^2\hat{f}^2(t) + 2\beta(\boldsymbol{x},t)B|\hat{f}(t)| + \eta
\end{aligned} \tag{19}$$

i.e.:

$$\begin{aligned}
&\dot{\hat{f}}(t) < \frac{k||\boldsymbol{s}(t)||}{k\boldsymbol{s}(t)^T\boldsymbol{B} + \epsilon}\alpha(\boldsymbol{x},t) + \frac{k}{k\boldsymbol{s}(t)^T\boldsymbol{B} + \epsilon}\big(\beta(\boldsymbol{x},t)^2 \\
&+||\boldsymbol{B}||^2\hat{f}^2(t) + 2\beta(\boldsymbol{x},t)||\boldsymbol{B}|| \cdot |\hat{f}(t)| + \eta\big)
\end{aligned} \tag{20}$$

and the expression (14) follows. According to Proposition 3.1, the variable $\boldsymbol{s}(t)$ is driven toward the layer $||\boldsymbol{s}(t)|| \leq \epsilon_1$. It crosses the ϵ_1-layer within a finite time t_s and is maintained in a neighborhood of it for $t > t_s$. Moreover, the discussion reported in the proof of Proposition 3.1 proves that the variable $g(t)$ is bounded. As a consequence, the variable $\dot{\boldsymbol{e}}_2(t)$ is bounded in norm, therefore the observation error $f(t) - \hat{f}(t)$ is bounded according to the dynamics (9). ■

Remark 3.2. It is given the CPS (5) under Assumptions 1-5, subject to state attacks. The attack estimator $\hat{f}(t)$ given by (14) approaches the true $f(t)$ after a finite time with bounded observation error and arbitrary precision ϵ. Finite-time detection of possible state attacks targeting a given state variable of the system (5) can be achieved appropriately setting the matrix $\boldsymbol{B}_f$ and using the attack estimator (14). For such an estimator, there exists a threshold $F(\epsilon)$ and a finite time t_d after which the attack is ensured to be detected, according to Lemma 3.1 and Proposition 3.1.

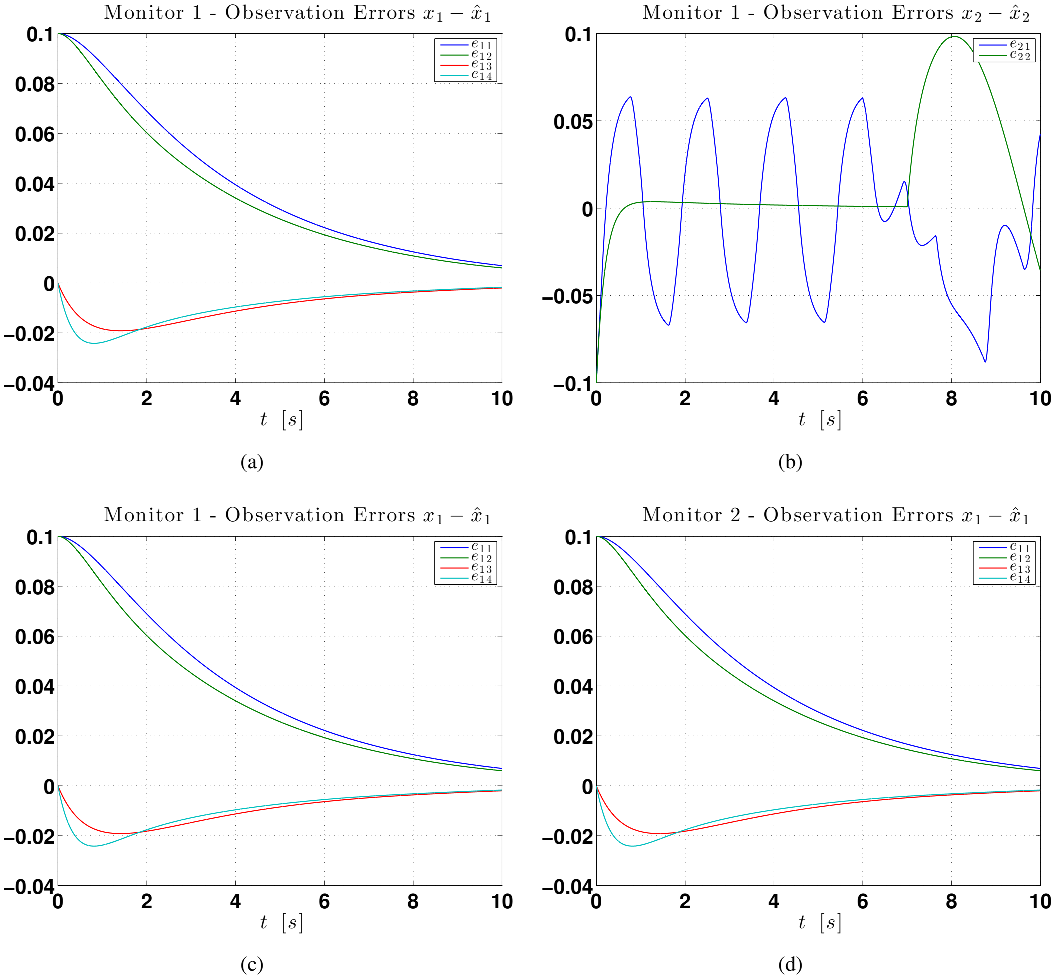
3.3. State attack monitor design for the WECC power system

Consider the US Western Electricity Coordinating Council power network [3] described by (1)-(2). Rearranging equations in order to eliminate the third state variable θ denoting the voltage angles at the buses[1], the dynamic model of the power plant reads:

$$\begin{cases}\dot{\delta}(t) = \omega(t) \\ M_g\dot{\omega}(t) = \left[L_{gg} + L_{gl}L_{ll}^{-1}L_{lg}\right]\{\omega(t) - P_\theta\} - D_g\delta(t) + P_\omega\end{cases} \tag{21}$$

For the sake of completeness, we recall that δ and ω are the generator rotor angles and frequencies, P_θ and P_ω are known inputs, M_g, D_g are diagonal matrices of the inertia and damping coefficients, and the matrices L_{gg},L_{gl}, L_{ll}, L_{lg} are the submatrices of the laplacian matrix $\mathcal{L}$ and have appropriate dimensions. The following numerical values, taken from [3], have been used in the simulation

[1] This can be done for instance using the Kron reduction method [17]



(a)

(b)

(c)

(d)

Fig. 4. *State attack, Monitor 1: (a) Observation error $\mathbf{e}_1(t)$; (b) Observation error $\mathbf{e}_2(t)$. State attack, Monitor 2: (c) Observation error $\mathbf{e}_1(t)$; (d) Observation error $\mathbf{e}_2(t)$.*

study:

$$M_g = \begin{bmatrix} .125 & 0 & 0 \\ 0 & .034 & 0 \\ 0 & 0 & .016 \end{bmatrix}, \quad D_g = \begin{bmatrix} .125 & 0 & 0 \\ 0 & .068 & 0 \\ 0 & 0 & .048 \end{bmatrix},$$

$$L_{gg} = \begin{bmatrix} .058 & 0 & 0 \\ 0 & .063 & 0 \\ 0 & 0 & .059 \end{bmatrix}, \quad L_{gl} = \begin{bmatrix} -0.58 & 0 & 0 & 0 & 0 & 0 \\ 0 & -0.63 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.59 & 0 & 0 & 0 \end{bmatrix},$$

$$L_{lg} = \begin{bmatrix} -.058 & 0 & 0 \\ 0 & .-063 & 0 \\ 0 & 0 & -.059 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \; L_{ll} = \begin{bmatrix} .235 & 0 & 0 & -.085 & -.092 & 0 \\ 0 & .296 & 0 & -.161 & 0 & -.072 \\ 0 & 0 & .330 & 0 & -.170 & -.101 \\ -.085 & -.161 & 0 & .246 & 0 & 0 \\ -.092 & 0 & -.170 & 0 & .262 & 0 \\ 0 & -.072 & -.101 & 0 & 0 & .173 \end{bmatrix}$$

It is supposed that a monitor measures the state variables δ_1 and ω_1 corresponding to the state of generator g_1 , i.e. the output matrix is

$$\bar{\boldsymbol{C}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Such measurement set allows the design of a coupled attack detection monitor for the first and fourth state variables setting $\bar{\boldsymbol{B}}_f = \bar{\boldsymbol{C}}$. The initial condition of the plant (5) has been selected as $\boldsymbol{x}(0) = [0.2\ 0.2\ 0\ 0\ 0\ 0]^T$. Two simultaneous observers (9) have been initialized as follows $\hat{\boldsymbol{x}}_1(0) = \hat{\boldsymbol{x}}_2(0) = [0.1\ 0.1\ 0\ 0\ 0\ 0\ 0.1]^T$. Following [12], two sinusoidal attack inputs of the form $f(t) = 0.5\sin(t)$ have been considered in the simulation, whose objective is to corrupt the first and fourth state variables starting from $t = 6\ s$ and $t = 7\ s$ respectively. The attack detection monitor is allowed to start monitoring at $t = 5\ s$, after an initial dwell-time needed for warming-up the observers.

Some of the performed tests have been reported in the following pictures. In particular, Figs. 4(a),4(b) and Figs. 4(c),4(d) show the observation errors $\boldsymbol{e}_1(t)$ and $\boldsymbol{e}_2(t)$ of the first and second observer respectively. Figs.5(a),5(c) show the behavior of the observers (14). With a width $\epsilon = 0.02$ and setting thresholds of $F = 0.6$, the attack is quickly detected, since detection occurs at times $t_1 = 7.644\ s$, $t_2 = 7.341\ s$ respectively. Finally, Figs.5(b), 5(d) shows the sliding surfaces $\boldsymbol{s}_1(t)$ and $\boldsymbol{s}_2(t)$ respectively.

4. Detection and reconstruction of sensor attacks

Consider the case of sensor attacks only ($\boldsymbol{B}_f = 0$). The following standard observer is designed for tackling the case of state attacks:

$$\begin{cases} \dot{\hat{\boldsymbol{x}}}_1(t) & = \boldsymbol{A}_{11}\hat{\boldsymbol{x}}_1(t) + \boldsymbol{A}_{12}\hat{\boldsymbol{x}}_2(t) + \boldsymbol{B}_1\boldsymbol{u}(t) + \boldsymbol{A}_{12}(\boldsymbol{y}(t) \\ & -\hat{\boldsymbol{y}}(t)) \\ \dot{\hat{\boldsymbol{x}}}_2(t) & = \boldsymbol{A}_{21}\hat{\boldsymbol{x}}_1(t) + \boldsymbol{A}_{22}\hat{\boldsymbol{x}}_2(t) + \boldsymbol{B}_2\boldsymbol{u}(t) + \boldsymbol{B}\hat{f}(t) \\ & +\boldsymbol{L}(\boldsymbol{y}(t) - \hat{\boldsymbol{y}}(t)) \\ \hat{\boldsymbol{y}}(t) & = \hat{\boldsymbol{x}}_2(t) \end{cases}$$

where $\boldsymbol{L} \in \mathrm{I\!R}^{p\times p}$ is chosen such that $\boldsymbol{A}_{22} - \boldsymbol{L} = -\boldsymbol{H}$ is positive definite, as in the previous case, and such that $\boldsymbol{G}_f \overset{\text{def}}{=} \boldsymbol{L}\boldsymbol{D}_f$ is full rank. Again, $\hat{f}$ is an estimate of the possible attack, to be designed later. The estimation error dynamics are given by

$$\begin{cases} \dot{\boldsymbol{e}}_1(t) & = \boldsymbol{A}_{11}\boldsymbol{e}_1(t) + \boldsymbol{D}_1\boldsymbol{d}(t) \\ \dot{\boldsymbol{e}}_2(t) & = \boldsymbol{A}_{21}\boldsymbol{e}_1(t) - \boldsymbol{H}\boldsymbol{e}_2(t) + \boldsymbol{G}_f(f(t) - \hat{f}(t)) + \boldsymbol{D}_2\boldsymbol{d}(t) \end{cases} \tag{22}$$

The interaction of the sensor attack monitor with the plant is shown in Figure 7: the output of the CPS is directly fed to the attack monitor in order to prevent the possible injection of false or corrupted data in the control loop.

Theorem 4.1. It is given the CPS (5) under Assumptions 1-5, subject to sensor attacks. The

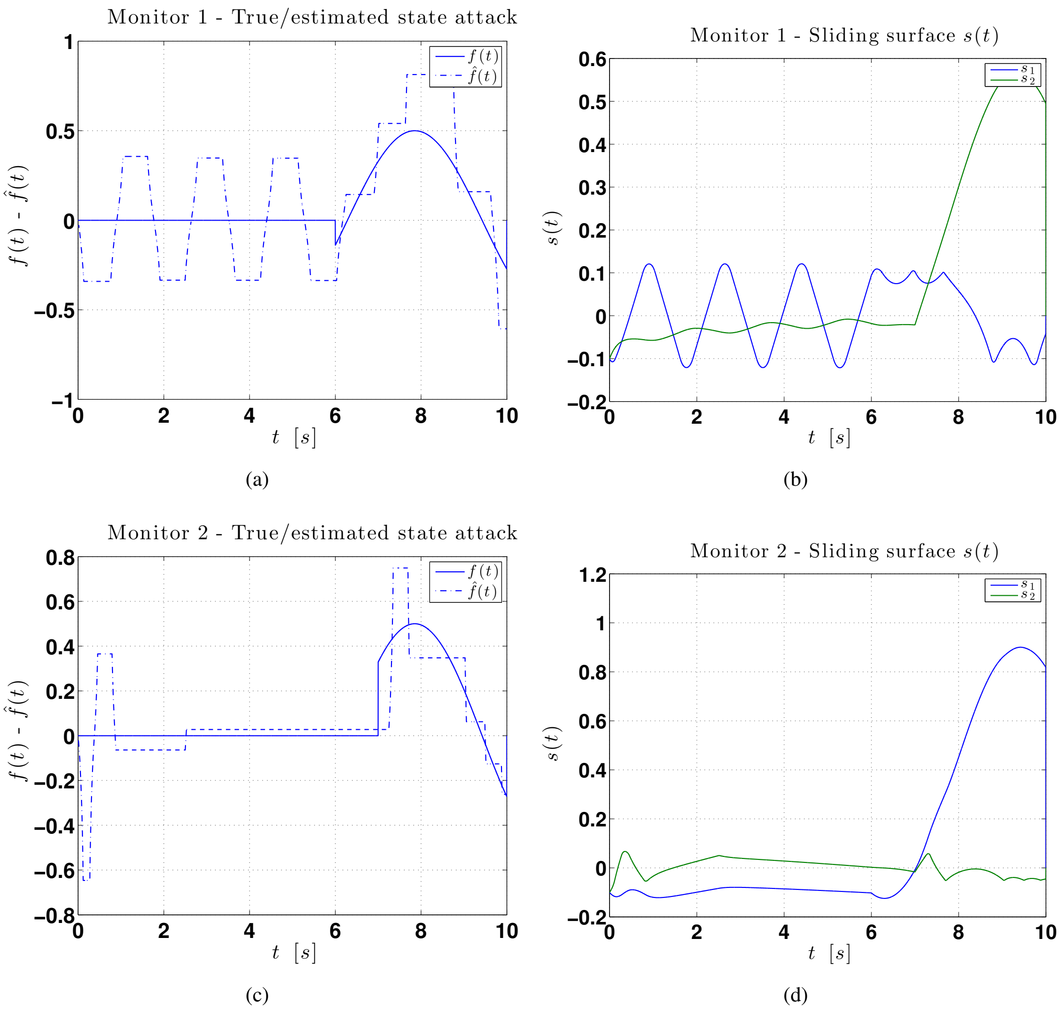


Fig. 5. *State attack, Monitor 1: (a) Estimated ($\hat{f}(t)$) versus true ($f(t)$) state attack; (b) Sliding surface $\boldsymbol{s}(t)$. State attack, Monitor 2: (c) Estimated ($\hat{f}(t)$) versus true ($f(t)$) state attack. (d) Sliding surface $\boldsymbol{s}(t)$.*

following sensor attack observer:

$$\dot{\hat{f}}(t) = -\gamma \frac{k}{k\boldsymbol{s}(t)^T\boldsymbol{G}_f - \epsilon sign(\boldsymbol{s}(t)^T\boldsymbol{G}_f)} \{||\boldsymbol{s}(t)||\,[\alpha_1(\boldsymbol{x},t) + ||\boldsymbol{H}_f|||\hat{f}(t)|\Big] + \Big[\beta_1(\boldsymbol{x},t) + ||\boldsymbol{G}_f|||\hat{f}(t)|\Big]^2 + \eta\Big\} \quad if|\boldsymbol{s}(t)^T\boldsymbol{G}_f| > \epsilon \tag{23}$$

with $\epsilon > 0, \gamma > 1, \eta > 0$ arbitrary, and

$$\begin{aligned}\boldsymbol{H}_f &\overset{\text{def}}{=} \boldsymbol{A}_{21}\boldsymbol{A}_{12}\boldsymbol{D}_f \\ \alpha_1(\boldsymbol{x},t) &\overset{\text{def}}{=} ||\boldsymbol{A}_{21}||\left(\boldsymbol{A}_{11}||\rho + ||\boldsymbol{D}_1||\rho_d(t)\right) + ||\boldsymbol{D}_2\rho'_d(t)|| \\ &+ ||\boldsymbol{H}_f||\rho_1(\boldsymbol{x},t) + ||\boldsymbol{G}_f||\rho_2(\boldsymbol{x},t); \\ \beta_1(\boldsymbol{x},t) &\overset{\text{def}}{=} ||\boldsymbol{A}_{21}||\rho + ||\boldsymbol{G}_f||\rho_1(\boldsymbol{x},t) + ||\boldsymbol{D}_2||\rho_d(t)\end{aligned}$$

ensures that:

- the variable $\boldsymbol{s}(t)$ is driven in norm to a layer of arbitrary width $\dfrac{\epsilon}{||\boldsymbol{G}_f||}$ within a finite time;
- the observation error $\boldsymbol{e}_2(t)$ is bounded;
- the observation error $f(t) - \hat{f}(t)$ is bounded.

Proof. Consider the function $g(t) = \boldsymbol{s}(t)^T\dot{\boldsymbol{s}}(t)$, with an arbitrary initial condition $g(t_0) = g_0$. Imposing the condition $\dot{g}(t) < -\eta$, $\eta > 0$ arbitrary, one gets

$$\begin{aligned}&\boldsymbol{s}(t)^T\Big[\boldsymbol{A}_{21}\Big(\boldsymbol{A}_{11}\boldsymbol{e}_1(t) + \boldsymbol{D}_1\boldsymbol{d}(t) - \boldsymbol{A}_{12}\boldsymbol{D}_f(f(t) - \hat{f}(t))\Big) \\ &-\boldsymbol{G}_f(\dot{f}(t) - \dot{\hat{f}}(t)) + \boldsymbol{D}_2\dot{\boldsymbol{d}}(t)\Big] < -\left(\boldsymbol{A}_{21}\boldsymbol{e}_1(t) - \boldsymbol{G}_f(f(t)\right. \\ &\left.-\hat{f}(t)) + \boldsymbol{D}_2\boldsymbol{d}\right)^T\Big(\boldsymbol{A}_{21}\boldsymbol{e}_1(t) - \boldsymbol{G}_f(f(t) - \hat{f}(t)) + \boldsymbol{D}_2\boldsymbol{d}\Big) - \eta \end{aligned} \tag{24}$$

i.e.

$$\begin{aligned}&\boldsymbol{s}(t)^T\boldsymbol{G}_f\dot{\hat{f}}(t) < -\boldsymbol{s}(t)^T\left[\boldsymbol{A}_{21}\left(\boldsymbol{A}_{11}\boldsymbol{e}_1(t) + \boldsymbol{D}_1\boldsymbol{d}(t) - \boldsymbol{A}_{12}\boldsymbol{D}_f\right.\right. \\ &\left.\left.(f(t) - \hat{f}(t))\right) - \boldsymbol{G}_f\dot{f}(t) + \boldsymbol{D}_2\dot{\boldsymbol{d}}(t)\right] - \left(\boldsymbol{A}_{21}\boldsymbol{e}_1(t) - \boldsymbol{G}_f(f(t)\right. \\ &\left.-\hat{f}(t)) + \boldsymbol{D}_2\boldsymbol{d}\right)^T\Big(\boldsymbol{A}_{21}\boldsymbol{e}_1(t) - \boldsymbol{G}_f(f(t) - \hat{f}(t)) + \boldsymbol{D}_2\boldsymbol{d}\Big) - \eta \end{aligned} \tag{25}$$

Consider the case $||\boldsymbol{s}(t)|| \cdot ||\boldsymbol{G}_f|| \geq \boldsymbol{s}(t)^T\boldsymbol{G}_f > \epsilon$, i.e. $||\boldsymbol{s}(t)|| > \dfrac{\epsilon}{||\boldsymbol{G}_f||} \overset{\text{def}}{=} \epsilon_2$. A strongest condition for $\boldsymbol{s}(t)^T\boldsymbol{G}_f > \epsilon$ is given by

$$\begin{aligned}&(\boldsymbol{s}(t)^T\boldsymbol{G}_f - \frac{\epsilon}{k})\dot{\hat{f}}(t) < -||\boldsymbol{s}(t)||\left[\alpha_1(\boldsymbol{x},t) + ||\boldsymbol{H}_f|||\hat{f}(t)|\right] \\ &-\left[\beta_1(\boldsymbol{x},t) + ||\boldsymbol{G}_f|||\hat{f}(t)|\right]^2 - \eta\end{aligned} \tag{26}$$

with $k > 1$. Therefore

$$\begin{aligned}\dot{\hat{f}}(t) < -\frac{k}{k\boldsymbol{s}(t)^T\boldsymbol{G}_f - \epsilon}&\Big\{||\boldsymbol{s}(t)||\left[\alpha_1(\boldsymbol{x},t) + ||\boldsymbol{H}_f|||\hat{f}(t)|\right] \\ &+\left[\beta_1(\boldsymbol{x},t) + ||\boldsymbol{G}_f|||\hat{f}(t)|\right]^2 + \eta\Big\} < 0\end{aligned} \tag{27}$$

and the expression (23) follows. Consider now the case $\boldsymbol{s}(t)^T\boldsymbol{G}_f < -\epsilon$, i.e. $||\boldsymbol{s}(t)|| > \epsilon_2$. From (25) a strongest condition for $\boldsymbol{s}(t)^T\boldsymbol{G}_f < -\epsilon$ is given by

$$\begin{aligned}(\boldsymbol{s}(t)^T\boldsymbol{B} + \frac{\epsilon}{k})\dot{\hat{f}}(t) < -||\boldsymbol{s}(t)|| \left[\alpha_1(\boldsymbol{x},t) + ||\boldsymbol{H}_f|||\hat{f}(t)|\right] \\ - \left[\beta_1(\boldsymbol{x},t) + ||\boldsymbol{G}_f|||\hat{f}(t)|\right]^2 - \eta \end{aligned} \tag{28}$$

Therefore

$$\begin{aligned}\dot{\hat{f}}(t) > -\frac{k}{k\boldsymbol{s}(t)^T\boldsymbol{G}_f - \epsilon}\Big\{||\boldsymbol{s}(t)|| \left[\alpha_1(\boldsymbol{x},t) + ||\boldsymbol{H}_f|||\hat{f}(t)|\right] \\ + \left[\beta_1(\boldsymbol{x},t) + ||\boldsymbol{G}_f|||\hat{f}(t)|\right]^2 + \eta\Big\} > 0 \end{aligned} \tag{29}$$

and the expression (23) follows. According to Proposition 3.1, the variable $\boldsymbol{s}(t)$ is driven toward the layer $||\boldsymbol{s}(t)|| \leq \epsilon_2$. It crosses the ϵ_2-layer within a finite time t_s and is maintained in a neighborhood of it for $t > t_s$. Moreover, the discussion reported in the proof of Proposition 3.1 proves that the variable $g(t)$ is bounded. As a consequence, the variable $\dot{\boldsymbol{e}}_2(t)$ is bounded in norm, therefore the observation error $f(t) - \hat{f}(t)$ is bounded according to the dynamics (22). ■

Remark 4.1. It is given the CPS (5) under Assumptions 1-5, subject to sensors attacks. The attack estimator $\hat{f}(t)$ given by (23) approaches the true $f(t)$ after a finite time with bounded observation error and arbitrary precision ϵ. Finite-time detection of possible sensor attacks targeting a given output variable of the system (5) can be achieved appropriately setting the matrix $\boldsymbol{D}_f$ and using the attack estimator (23). For such an estimator, there exists a threshold $F(\epsilon)$ and a finite time t_d after which the attack is ensured to be detected, according to Lemma 3.1 and Proposition 3.1.

4.1. Sensors attack monitor design for a consensus network

Consider the undirected consensus network with 8 nodes described in [3, Figure S9] and characterized by a graph with connectivity three. The network evolves according to the unforced state space equation

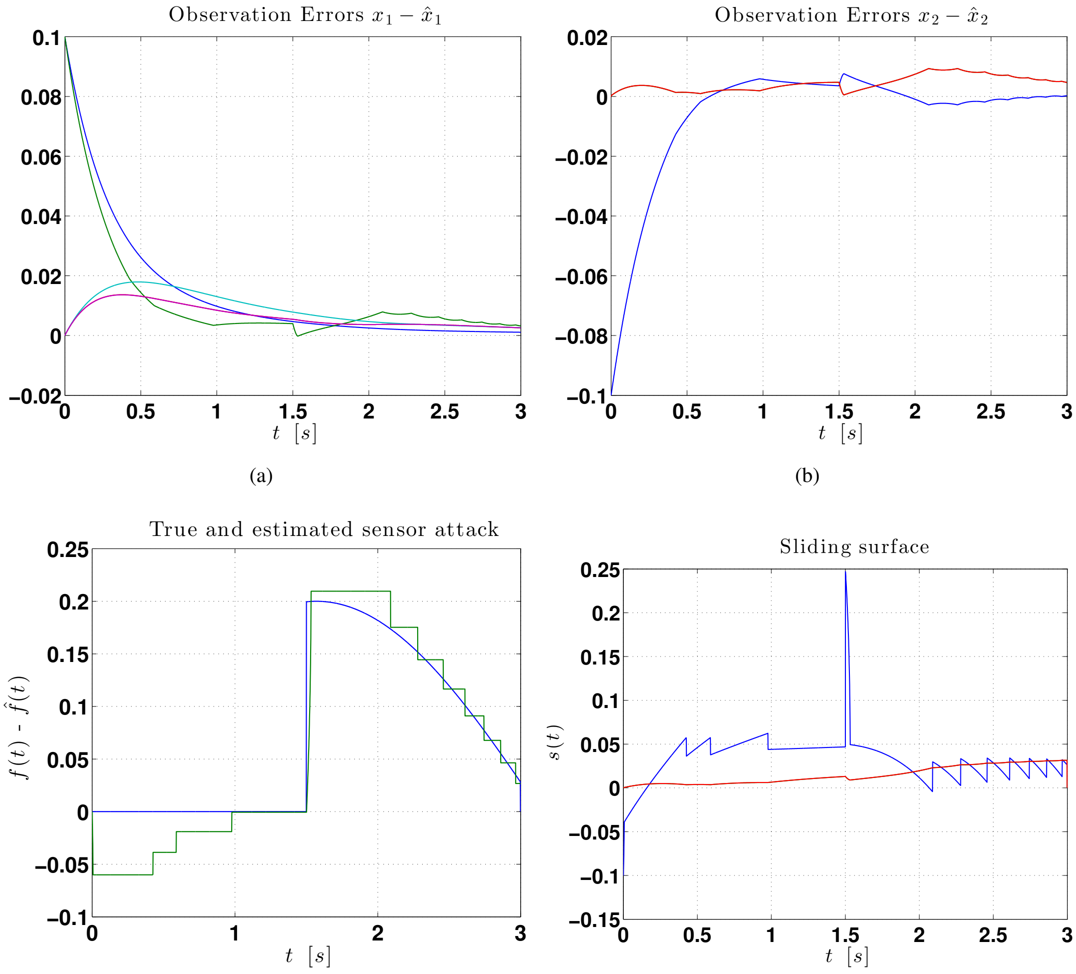
$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x},$$

with graph laplacian matrix

$$\boldsymbol{A} = \begin{bmatrix} -3 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & -3 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & -3 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & -3 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & -3 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & -3 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & -3 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & -3 \end{bmatrix}.$$

An attack detection monitor for the first output variable is designed setting

$$\bar{\boldsymbol{D}}_f = \boldsymbol{D}_f = [\ 1 \quad 0 \quad 0\ ]^T$$



(a)

(b)

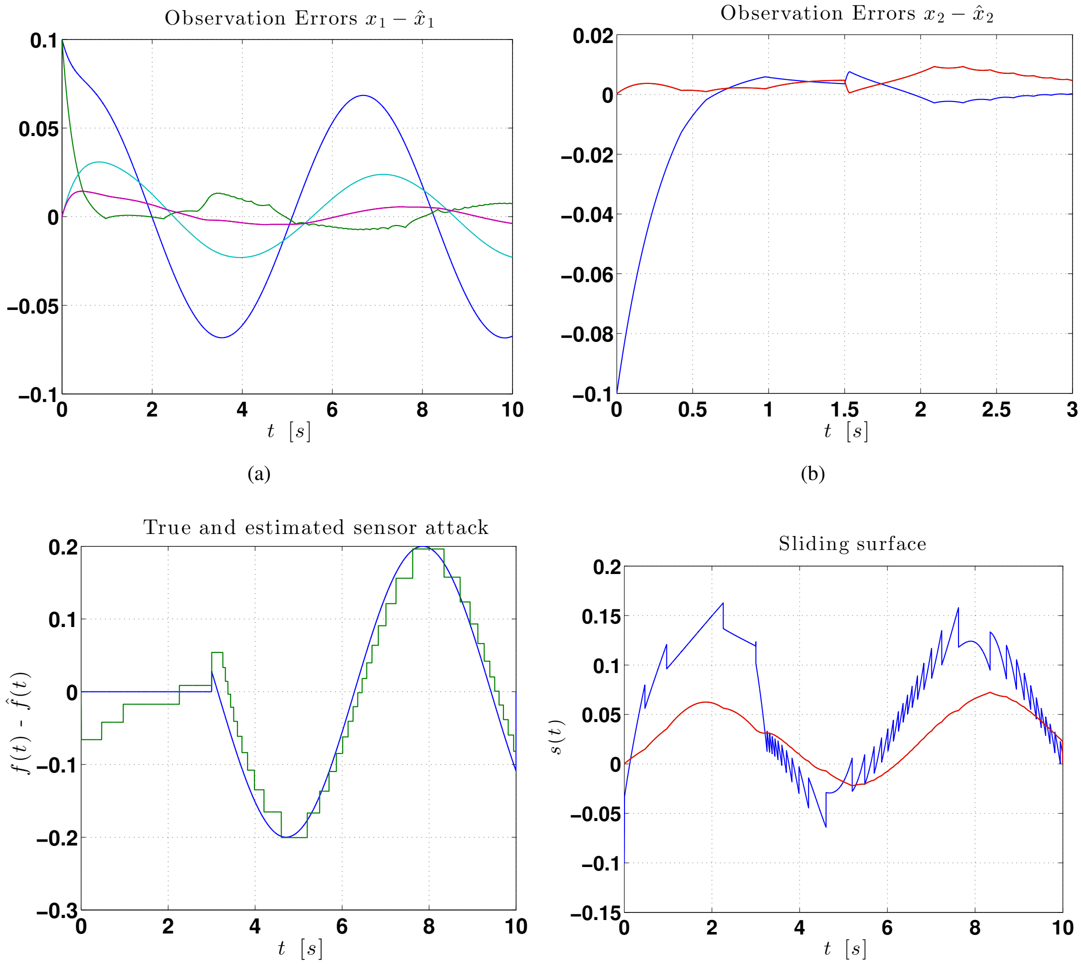
(c)

(d)

Fig. 6. *Sensor Attack: (a) Observation error $\boldsymbol{e}_1(t)$; (b) Observation error $\boldsymbol{e}_2(t)$; (c) Estimated ($\hat{f}(t)$) versus true ($f(t)$) attack; (d) Sliding surface $\boldsymbol{s}(t)$.*

The initial condition of the plant (5) has been selected as $\boldsymbol{x}(0) = [0.2\ 0.2\ 0\ 0\ 0\ 0\ 0.1\ 0]^T$, and the observer (22) has been initialized as follows $\hat{\boldsymbol{x}}(0) = [0.1\ 0.1\ 0\ 0\ 0\ \ 0.1\ 0.1\ 0]^T$. A sinusoidal attack input of the form $f(t) = 0.2\sin(t)$ has been considered to corrupt the first output variable starting from $t = 1.5\ \ s$. The attack detection monitor is allowed to start monitoring at $t = 0.5\ \ s$, after an initial dwell-time needed for warming-up the observer. Some of the performed tests have been reported in Fig. 6(a) and Fig. 6(b), showing the observation errors $\boldsymbol{e}_1(t)$ and $\boldsymbol{e}_2(t)$, respectively. Fig.6(c) shows the behavior of the observer (23). With a width $\epsilon = 0.05$ and setting a threshold of $F = 0.1$, the sensor attack is quickly detected, since detection occurs at time $t = 1.524\ s$. Finally, Fig.6(d) shows the sliding surface $\boldsymbol{s}(t)$.

To test the robustness of the proposed approach, the disturbance term $d(t) = 0.2\cos(t)$ has been added to the plant (4). The sinusoidal sensor attack input has been now considered to corrupt such state variable starting from the time $t = 3\ s$. The obtained results have been reported in Fig. 7(a) - 7(b), showing the observation errors $\boldsymbol{e}_1(t)$ and $\boldsymbol{e}_2(t)$, respectively, and in Fig.6(c) shows the behavior of the observer (23). With the same width $\epsilon = 0.05$ and threshold $F = 0.1$, detection



(a)

(b)

(c)

(d)

Fig. 7. *Sensor Attack, perturbed case: (a) Observation error $\boldsymbol{e}_1(t)$; (b) Observation error $\boldsymbol{e}_2(t)$; (c) Estimated ($\hat{f}(t)$) versus true ($f(t)$) sensor attack; (d) Sliding surface $\boldsymbol{s}(t)$.*

occurs at time $t = 3.820$ s. Finally, Fig.7(d) shows the sliding surface $\boldsymbol{s}(t)$ in the perturbed case. Effectiveness of detection is preserved in face of a perturbing disturbance of almost the same entity.

5. Compensation of attacks

Once a cyber-attack has been detected and reconstructed within a finite time interval, as previously proved, the reconstructed signal can be effectively used for achieving compensation, in order to guarantee continuity of service of the CPS in face of the received attack. The control problem of the output tracking of a given reference variable $\boldsymbol{y}_d(t)$ is considered here, in the presence of a state attack ($\boldsymbol{D}_f = 0$, the extension of the case of sensor attacks is straightforward).

Remark 5.1. Since for the tracking error $\boldsymbol{e}_t(t) \overset{\text{def}}{=} \boldsymbol{y}(t) - \boldsymbol{y}_d(t)$ it holds $\boldsymbol{e}_t(t) = \boldsymbol{e}_2(t) - \boldsymbol{\epsilon}_2(t)$, with $\boldsymbol{\epsilon}_2(t) \overset{\text{def}}{=} \hat{\boldsymbol{y}}(t) - \boldsymbol{y}_d(t)$, finding a control strategy ensuring the asymptotic vanishing of $\boldsymbol{\epsilon}_2(t)$ is enough in order to guarantee the closed loop boundedness of the tracking error, in view of the previous results. The case of square plants ($m = p$) will be shortly addressed in the following.

Define the following sliding surface:

$$\boldsymbol{\sigma}(t) = \boldsymbol{S}\boldsymbol{\epsilon}_2(t) \tag{30}$$

with $\boldsymbol{S} \in \mathbb{R}^{m\times p}$ designed such that $\boldsymbol{S}\boldsymbol{B}_2$ is nonsingular. The following additional assumption is introduced:

Assumption 6. For the CPS (5), it is assumed that the pair $(\boldsymbol{A}_{22}, \boldsymbol{B}_2)$ is controllable.

A control input is designed the following form: $\boldsymbol{u}(t) = \boldsymbol{u}_{eq}(t) + \boldsymbol{v}(t)$, with

$$\boldsymbol{u}_{eq}(t) = -\,(\boldsymbol{S}\boldsymbol{B}_2)^{-1}\boldsymbol{S}(\boldsymbol{A}_{22}\boldsymbol{\epsilon}_2(t) + \boldsymbol{A}_{21}\hat{\boldsymbol{x}}_1(t) + \boldsymbol{A}_{22}\boldsymbol{y}_d(t) + \boldsymbol{L}\boldsymbol{e}_2(t) - \dot{\boldsymbol{y}}_d(t) + \boldsymbol{B}\hat{f}(t)) \tag{31}$$

Accordingly the dynamics of the error $\boldsymbol{\epsilon}_2(t)$ become

$$\dot{\boldsymbol{\epsilon}}_2(t) =\boldsymbol{G}\boldsymbol{A}_{22}\boldsymbol{\epsilon}_2(t) + \boldsymbol{G}(\boldsymbol{A}_{21}\hat{\boldsymbol{x}}_1(t) + \boldsymbol{A}_{22}\boldsymbol{y}_d(t) + \boldsymbol{L}\boldsymbol{e}_2(t) - \dot{\boldsymbol{y}}_d(t) + \boldsymbol{B}\hat{f}(t)) + \boldsymbol{B}_2\boldsymbol{v}(t) \tag{32}$$

where $\boldsymbol{G} \overset{\text{def}}{=} \boldsymbol{I} - \boldsymbol{B}_2(\boldsymbol{S}\boldsymbol{B}_2)^{-1}\boldsymbol{S} \in \mathbb{R}^{p\times p}$. In general, the $m \times p$ matrix $\boldsymbol{S}$ can be assigned such that the matrix $\boldsymbol{G}\boldsymbol{A}_{22}$ has m null eigenvalues and $p - m$ arbitrary eigenvalues (these latter describe the dynamics of the reduced order system once a sliding motion on the surface (30) is enforced). In the square case, a matrix $\boldsymbol{F}$ always exists such that $\boldsymbol{B}_2\boldsymbol{F} = \boldsymbol{G}$. Let $\boldsymbol{v} = \boldsymbol{F}\boldsymbol{\omega} + \boldsymbol{\nu}$, therefore

$$\dot{\boldsymbol{\epsilon}}_2(t) =\boldsymbol{G}\boldsymbol{A}_{22}\boldsymbol{\epsilon}_2(t) + \boldsymbol{G}(\boldsymbol{A}_{21}\hat{\boldsymbol{x}}_1(t) + \boldsymbol{A}_{22}\boldsymbol{y}_d(t) + \boldsymbol{L}\boldsymbol{e}_2(t) - \dot{\boldsymbol{y}}_d(t) + \boldsymbol{B}\hat{f}(t) + \boldsymbol{\omega}(t)) + \boldsymbol{B}_2\boldsymbol{\nu}(t) \tag{33}$$

therefore designing

$$\boldsymbol{\omega}(t) =\dot{\boldsymbol{y}}_d(t) - \boldsymbol{A}_{21}\hat{\boldsymbol{x}}_1(t) - \boldsymbol{A}_{22}\boldsymbol{y}_d(t) - \boldsymbol{L}\boldsymbol{e}_2(t) - \boldsymbol{B}\hat{f}(t) \tag{34}$$

one gets

$$\dot{\boldsymbol{\epsilon}}_2(t) =\boldsymbol{H}\boldsymbol{A}_{22}\boldsymbol{\epsilon}_2(t) + \boldsymbol{B}_2\boldsymbol{\nu}(t) \tag{35}$$

Finally, the finite-time achievement of a sliding motion on the surface $\boldsymbol{\sigma}(t) = 0$ can be ensured imposing the usual condition $\boldsymbol{\sigma}(t)^T\dot{\boldsymbol{\sigma}}(t) < -\eta\boldsymbol{\sigma}(t))^T\boldsymbol{\sigma}(t); \eta > 0$, i.e. designing

$$\boldsymbol{\nu}(t) = -\,\eta(\boldsymbol{S}\boldsymbol{B}_2)^{-1}sign(\boldsymbol{\sigma}(t))) \tag{36}$$

5.1. An example of state attack compensation: the IEEE 39 bus power system

Consider the IEEE 39 bus power system [18, 19] with 10 generators. Following [15, Section 4], the plant model can be transformed in a linear state-space representation by means of Kron reduction [17]. After some additional simplifications, e.g. assuming inertia and damping matrices $\mathrm{M}_g = \mathrm{D}_g = 0.1\boldsymbol{I}_{10}$, and a suitable change of coordinates, the system is expressed in the form (4) with

$$\bar{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{I}_{10} \\ -0.1\boldsymbol{I}_{10} & -\boldsymbol{I}_{10} \end{bmatrix}; \quad \bar{\boldsymbol{B}} = \begin{bmatrix} \boldsymbol{0}_{10} & 10\boldsymbol{I}_{10} \end{bmatrix}^T$$

$$\bar{\boldsymbol{C}} = [\ \boldsymbol{I}_{10} \quad \boldsymbol{I}_{10}\ ]$$

An attack detection monitor for the 11-th state variable has been designed setting

$$\bar{\boldsymbol{B}}_f = [\ \mathbf{0}_{1\times 10} \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0\ ]^T$$

The control objective is the tracking of the reference constants $r_1 = 1$, $r_2 = 1$, $r_3 = 2$ by the first, second and third output variables respectively. Following [15], a sinusoidal attack input of the form $f(t) = 0.5\sin(0.2\pi t)(|x_{11}(t)| + 1)$ has been considered to corrupt the 11-th state variable starting from $t = 2\ s$. The attack detection monitor is allowed to start monitoring at $t = 1\ s$, after an initial dwell-time needed for warming-up the observer. The matrix $\boldsymbol{S}$ has been designed as $\boldsymbol{S} = diag\{1, 1, 1, 1, 1, 1, 1, 1, 1, 1\}$.

Some of the performed tests have been reported in the following pictures. In particular, Fig. 8(a) and 8(b) show the observation errors $\boldsymbol{e}_1(t)$ and $\boldsymbol{e}_2(t)$, respectively. Fig.8(c) shows the behavior of the observer (14). With a width $\epsilon = 0.1$ and setting a threshold of $F = 0.2$, the attack is quickly detected, since detection occurs at time $t = 2.1490\ s$. Finally, Fig.8(d) shows the (true and estimated) output variables and the tracking performances. Note that the first output variable is subject to the above described cyber-attack, which is promptly detected and almost completely compensated.

6. Conclusions

In this paper, a design technique is proposed for building a state/sensor attack observer able to either work as a detection monitor with guaranteed finite-time performance, and to reconstruct the attack itself in the presence of bounded external perturbations or modeling errors possibly affecting the state equation. The proposed attack observers are very easy to be implemented. Compensation of attacks has been also addressed for the case of square plants. The chosen methodology is based on sliding-mode based techniques, and require the availability of an upper bound, even rough, for the function describing the attack. An extensive simulation study using test-cases taken from the literature has been performed to support the theoretical study.

Future developments of the proposed architecture will be focused on enhancing the capabilities of method and improving the design of monitors towards handling more challenging cases, such as the presence of coordinated deception and stealth attacks.

7. References

[1] Poovendran, R., Sampigethaya, K., Gupta, S.K.S., *et al.*: ‘Special issue on cyber-physical systems’, Proceedings of the IEEE, 2012, 100, (1), pp. 1–12

[2] Sandberg, H., Amin, S., Johansson, K.: ‘Cyberphysical security in networked control systems: An introduction to the issue’, IEEE Control Systems, 2015, 35, (1), pp. 1–12

[3] Pasqualetti, F., Dorfler, F., Bullo, F.: ‘Control theoretic methods for cyberphysical security’, IEEE Control Systems, 2015, 35, (1), pp. 110–127

[4] Karnouskos, S.: ‘Stuxnet worm impact on industrial cyber-physical system security’. Proc. 37th Annu. Conf. IEEE Industrial Electronics Society, Melbourne,, 2011, pp. 4490–4494

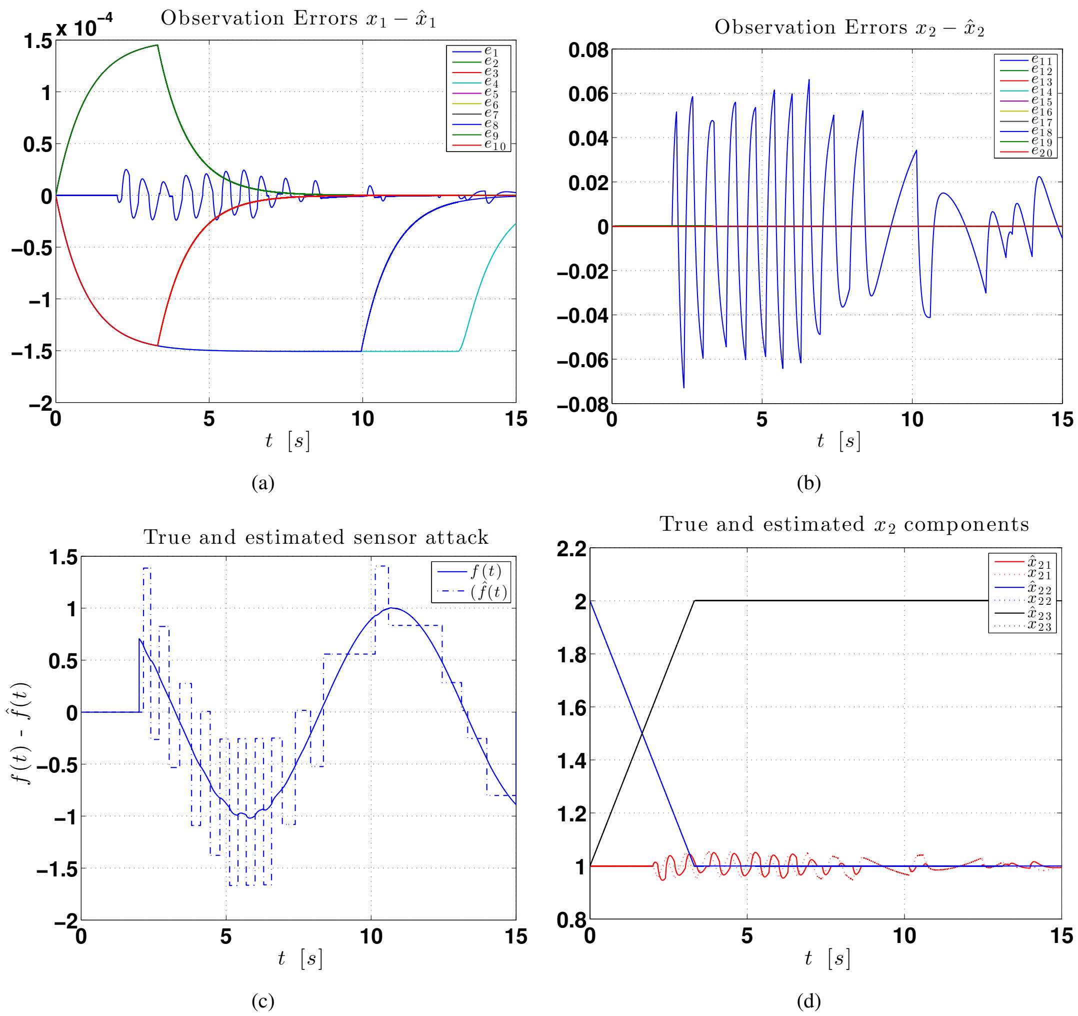


Fig. 8. *Sensor Attack: (a) Observation error* $\mathbf{e}_1(t)$*; (b) Observation error* $\mathbf{e}_2(t)$*; (c) Estimated (*$\hat{f}(t)$*) versus true (*$f(t)$*) attack; (d) True and estimated output variables.*

[5] Slay, J., Miller, M.: 'Lessons learned from the maroochy water breach', (Springer, 2007)

[6] Fawzi, H., Tabuada, P., Diggavi, S.: 'Secure estimation and control for Cyber-Physical Systems under adversarial attacks', IEEE Trans. Autom. Contr., 2014, 59, (6), pp. 1454–1467

[7] Zhang, H., Cheng, P., Shi, L., *et al.*: 'Optimal denial-of-service attack scheduling with energy constraint', IEEE Transactions on Automatic Control, 2015, 60, (11), pp. 3023–3028

[8] Zhang, H., Cheng, P., Shi, L., *et al.*: 'Optimal DoS attack scheduling in wireless networked control system', IEEE Transactions on Control Systems Technology, 2016, 24, (3), pp. 843–852

[9] Chen, J., Patton, R.J.: 'Robust model-based fault diagnosis for dynamic systems', (Kluwer Academic Publishers, Norwell, MS, 1999)

[10] Isermann, R.: 'Fault-Diagnosis Systems', (Springer, 2006)

[11] Aubrun, C., Sauter, D., Yamé, J.: ‘Fault diagnosis of networked control systems’, Int. J. Appl. Math. Comput. Sci., 2008, 18, (4), pp. 525–537

[12] Teixeira, A., Sandberg, H., Johansson, K.: ‘Networked control systems under cyber attacks with applications to power networks’. Proc. 2010 American Control Conference, Baltimore, USA, 2010, pp. 3690–3606

[13] Mo, Y., Weerakkody, S., Sinopoli, B.: ‘Physical autentication of control systems’, IEEE Control Systems, 2015, 31, (1), pp. 93–109

[14] Manandhar, K., Cao, X., Hu, F., *et al.*: ‘Detection of fault and attacks including false data injection attack in smart grid using kalman filter’, IEEE Trans. Contr. Networked Sys., 2014, 1, (4), pp. 370–379

[15] Song, D., Ao, W., Wen, C.: ‘Adaptive CPS attack detection and reconstruction with application to power systems’, IET Contr. Theory and Appl., February 2016

[16] Utkin, V.: ‘Sliding Modes in Control and Optimization’, (Springer, 1992)

[17] Dorfler, F., Bullo, F.: ‘Kron reduction of graphs with applications to electrical networks’, IEEE Transactions on Circuits and Systems I: Regular Papers, 2013, 60, (1), pp. 150–163

[18] Mei, S., Zhang, X., Cao, M.: ‘Power grid complexity’, (Springer Science & Business Media, 2011)

[19] Zimmerman, R.D., Murillo-Sánchez, C.E., Thomas, R.J.: ‘MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education’, IEEE Transactions on power systems, 2011, 26, (1), pp. 12–19