Cyber-Attacks and Faults Reconstruction using Finite Time Convergent Observation Algorithms: Electric Power Network Application

S. Nateghi[a*], Y. Shtessel[a], C. Edwards[b]

[a] *Department of Electrical Engineering, University of Alabama, Huntsville, AL, USA*
[b] *Structures & Dynamics Research Group, CEMPS, University of Exeter, Exeter, UK*

Abstract

In this work, linear (linearized) cyber-physical systems with output feedback control, whose sensors are experiencing faults or are under cyber-attack, are studied. Two different cases are investigated. First, when all sensors are attacked, then, when some sensors are protected from the attacks. Finite time convergent observers, specifically the sliding mode ones, including the observers with gain adaptation, are employed for on-line reconstruction of the cyber-attacks. The corrupted measured outputs are "cleaned" from cyber-attacks, and feedback control that uses the "cleaned" outputs is shown to provide elevated cyber-physical system performance close to the one without attack. Finally, the proposed methodology is applied to an electric power system under cyber-attack. Simulation results illustrate the efficacy of the proposed observers.

Keywords: Cyber-physical systems, Finite Time Convergent (Sliding mode) observer, Adaptive sliding mode observer

1. Introduction

Cyber Physical Systems (CPS) are the integration of the cyber-world of computing and communications with the physical world. In many systems, control of a physical plant is integrated with a wireless communication network, for example transportation networks, electric power generation and distribution networks, integrated biological systems, industrial automation systems, and economic systems [1-3].

[*] *Corresponding author: 301 Sparkman Drive, University of Alabama in Huntsville, Huntsville, AL 35899*
Email address: sb0086@uah.edu (S. Nateghi)

Since CPSs use open computation and communication platform architectures, they are vulnerable to suffering adversarial physical faults or cyber-attacks. Faults and cyber-attacks are referred to as *attacks* throughout the paper. We have to acknowledge that cyber-attacks are relatively new phenomena that cause systems faults. Recent real-world cyber-attacks, including multiple power blackouts in Brazil [4], the StuxNet attack [5] in 2010 and the Maroochy water bleach in 2000 [6], showed the importance of providing security to CPSs. It is suggested in [7] that information security mechanisms have to be complemented by specially designed resilient control systems. In the other words, information security techniques [8] may be not sufficient for protecting systems from sophisticated cyber-attacks.

Controlling systems with measurement sensors and actuators, which are hijacked/corrupted remotely or physically by the attackers is challenging. The use of novel control/observation algorithms is proposed in this paper for recovering CPS performance on-line if an attacker penetrates the information security mechanisms.

Cyber security of CPS must provide three main security goals: *availability*, *confidentiality*, and *integrity* [8]. This means that the CPS is to be accessible and usable upon demand, the information has to be kept secret from unauthorized users, and the trustworthiness of data has to be guaranteed. Lack of availability, confidentiality, and integrity yields denial of service, disclosure, and deception respectively. A specific kind of deception attack called a *replay attack* has been investigated when the system model is unknown to the attackers but they have access to the all sensors [9-10]. *Replay attacks* are carried out by “hijacking” the sensors, recording the readings for a certain time, and repeating such readings while injecting them as exogenous signal into the system’s sensors. In the case when the system’s dynamic model is known to the attacker, another kind of deception attack, called *covert attack*, has been studied in [11], and the proposed algorithm allows cancelling out the effect of this attack on the system dynamics. In systems with unstable modes, false data injection attacks are applied to make some unstable modes unobservable [12]. Denial of service attacks assaults data availability through blocking information flows between different components of the CPS. The attacker can jam the communication channels, modify devices and prevent them from sending data, violate the routing protocols, etc. [13]. In a stealth attack, the attacker modifies some sensor readings by physically tampering with the individual meters or by getting access to some communication channels [14]. As a result, detecting and isolating cyber-attacks in CPSs has received immense attention [15], and there is a significant amount of publications, which have focused on keeping the system safe from attacks. However, how to ensure the CPS control system can continue functioning properly if the cyber-attack has happened is another serious problem that should be investigated.

To increase the robustness of CPS to sensor attacks, a lot of effort has been applied to counteract the attacks via developing sophisticated software [16-17]. However, if despite this an attack still happens, they must be detected, isolated, and, if possible, reconstructed and compensated. Therefore, beside prompt detection, on-line attack reconstruction is crucial for guaranteeing continuity of service of CPS, often supporting critical infrastructures. Thus, both on-line detection and reconstruction of attacks on SPC with a consequent compensation are of prime interest in terms of CPS research.

In [18], new adaptive control architectures that can foil malicious sensor and actuator attacks are developed without reconstructing the attacks, by means of feedback control only. Another approach to protect CPS from attack using on-line attack reconstruction/estimation techniques is proposed in [19], [20]. A sparse recovery algorithm is applied to reconstruct cyber-attacks in [20]. Sparse recovery algorithms are effective when only limited number of sensors is under attacks. In [19], a finite time convergent higher order sliding mode (HOSM) observer based on a HOSM differentiator and a sparse recovery algorithm are used to reconstruct on-line the cyber-attack in a nonlinear system.

Detection and observation of a scalar attack by a sliding mode observer (SMO) has been accomplished for a linearized differential-algebraic model of an electric power network when plant and sensor attacks do not occur simultaneously [21].

A probabilistic risk mitigation model for cyber-attacks against Phasor Measurement Unit (PMU) networks is presented in [22], where a risk-mitigation technique determines whether a certain PMU should be kept connected to the network or removed, while minimizing the maximum threat level for all connected PMUs. In [23] the sliding mode-based observation algorithm is used to reconstruct the attacks asymptotically. This reconstruction is approximate only, since pseudo-inverse techniques are used.

In this paper, linear/linearized CPSs controlled by output feedback subject to sensor attacks are considered. The corrupted/attacked measurements propagate the attack signals to the CPS through the feedback controllers causing CPS performance degradation. Two CPS configurations are studied: a scenario when all sensors are prone to get attacked and a scenario when some sensors are protected from the attacks.

The main challenge that is addressed in the paper is on-line exact reconstruction of the sensor attacks with an application to an electric power network.

The contributions of this are as follows

- Two CPS scenarios are investigated: all sensors are prone to attack; and some sensors are protected from the attack.
- Fixed-gain and adaptive-gain SMOs are proposed for the on-line reconstruction of sensor attacks. Specially, dynamic filters that address the attack propagation dynamics are proposed and employed for attack reconstruction for the first time.
- The measurements corrupted by the attacks are "cleaned" on-line in order to stop the sensor attack propagation to the CPS through output feedback control. After a finite time transient, needed for the attack reconstruction the CPS closed loop performance results as attack free performance.
- The proposed methodology is applied to an electric power network, whose sensors are under attack. Simulation results illustrate the efficacy of the proposed observers.

2. System Dynamics

Consider a completely observable and controllable LTI system (Fig. 1), whose measured output is corrupted by an attack.

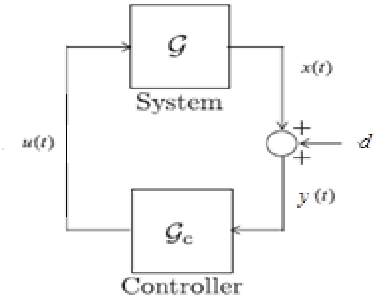

Fig. 1. The closed-loop dynamical system in the presence of sensor attacks

The system dynamics in Fig. 1 are given by

$$\begin{aligned} \dot{x} &= \bar{A}x + \bar{B}u \\ y &= Cx + Dd \end{aligned} \tag{1}$$

where the triplet $\left(\bar{A},\bar{B},C\right)$ is completely controllable and observable, $x \in \mathbb{R}^n$ denotes the states, $u \in \mathbb{R}^q$ is a control signal, and $y \in \mathbb{R}^p$ represents the measured output. The $d(t) \in \mathbb{R}^m$ is the smooth norm-bounded sensor attack signal that is to be reconstructed on-line.

The following assumption is made:

(A1): the Kimura-Davidson condition [24]

$$l + p + 1 \geq n \tag{2}$$

holds.

Assuming assumption A1 holds then there exists a static output feedback control

$$u = -Ky \tag{3}$$

where $K \in \mathbb{R}^{l \times p}$ is a gain matrix, that stabilizes the system (1).

Substituting (3) into (1) results in the closed loop system

$$\begin{aligned} \dot{x} &= \left(\bar{A} - \bar{B}KC\right)x - \bar{B}KDd \\ y &= Cx + Dd \end{aligned} \tag{4}$$

Therefore, the closed-loop system (4) can be rewritten as

$$\begin{aligned} \dot{x} &= Ax + Bd \\ y &= Cx + Dd \end{aligned} \tag{5}$$

where $A = \bar{A} - \bar{B}KC$ is Hurwitz, and $B = -\bar{B}KD$.

Again, note that $d = d(t)$ is a bounded, smooth sensor cyber-attack signal.

Discussion: Assume that the sensor attacks are reconstructed, i.e. $\hat{d}(t) \rightarrow d(t)$ as time increases. Then the polluted measurement $y = Cx + Dd$ can be "cleaned" as $y_{clean} = y - D\hat{d}$, and $y_{clean} - \tilde{y} \rightarrow 0$ as time increases, where $\tilde{y} = Cx$ is a measured output in the absence of attack. Therefore, substituting $\tilde{y} = Cx$ for $y = Cx + Dd$ in (3) we obtain

$$\begin{cases} \dot{x} = Ax + Bd \\ y = Cx + Dd \end{cases} \rightarrow \begin{cases} \dot{x} = Ax \\ \tilde{y} = Cx \end{cases} \tag{6}$$

In the other words, the compensated dynamics of the output feedback control system (1), whose sensors are under attack, will converge to the stable system (6) with the desired asymptotic dynamics that are not affected by the sensor attack signals. The problem of the output feedback controller (3) design is out of the scope of this work.

Therefore, the main problem addressed in this paper is on-line reconstruction of the sensor attack signal $d(t)$ in system (5) with application to an electrical power system network shown in Fig. 2.

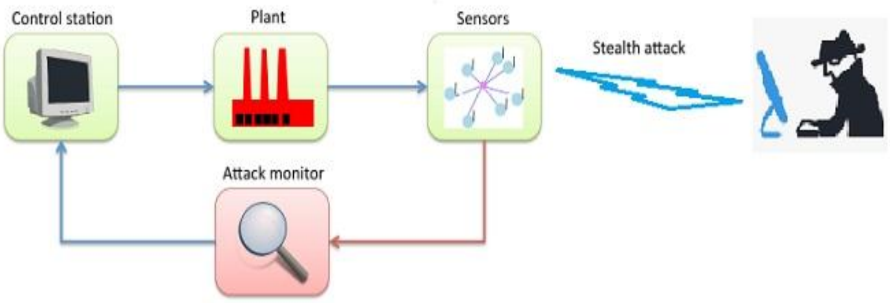

Fig. 2. Sensor Attack Analyzer

3. Problem formulation

The problem is to protect system (1) from the norm-bounded smooth sensor attack $d(t) \in \mathbb{R}^m$, with $\|d(t)\| \le L_1,\ \|\dot{d}(t)\| \le L_2,\ L_1, L_2 > 0$ by

(a) reconstructing on-line the sensor attack $d(t)$, so that the estimate

$$\hat{d}(t) \to d(t) \tag{7}$$

as time increases, and

(b) "cleaning up" the measurement: $y_{clean} = y - D\hat{d}$ to compensate the effects of the sensor attack on the system closed-loop performance

4. Sensor attack On-line Reconstruction

Firstly, a case when the number of sensors and the number of attacks is the same ($p = m$) is studied, and two different scenarios are investigated:

(a) all sensors can be attacked,

(b) k sensors ($k < p$ **)** are protected from the attacks.

Then, the case when the number of sensors is greater than the number of attacks ($p > m$) is studied:

(c) all sensors can be attacked.

4.1. All sensors can be attacked

Since the system (5) is completely controllable and observable, it can be partitioned as

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} d \\ y = C_1 x_1 + C_2 x_2 + Dd \tag{8}$$

where $p = m,\ x_1 \in \mathbb{R}^{n-p},\ x_2 \in \mathbb{R}^p,\ C_1 \in \mathbb{R}^{p\times(n-p)},\ C_2 \in \mathbb{R}^{p\times p},\ D \in \mathbb{R}^{p\times m}, \det(C_2) \neq 0$. Firstly, the closed-loop system (5) is transformed to a form convenient for the observer design. Specifically, the state variable $x_2 \in \mathbb{R}^p$ is replaced by the output variable $y \in \mathbb{R}^p$. This is

$$\begin{aligned} \dot{x}_1 &= G_{21}x_1 + G_{22}y + G_{23}d \\ \dot{y} &= G_{11}x_1 + G_{12}y + G_{13}d + D\dot{d} \end{aligned} \tag{9}$$

where

$$\begin{aligned} G_{11} &= C_1A_{11} - C_1A_{12}C_2^{-1}C_1 + C_2A_{21} - C_2A_{22}C_2^{-1}C_1 \\ G_{12} &= C_1A_{12}C_2^{-1} + C_2A_{22}C_2^{-1} \\ G_{13} &= -C_1A_{12}C_2^{-1}D + C_1B_1 - C_2A_{22}C_2^{-1}D + C_2B_2 \\ G_{21} &= A_{11} - A_{12}C_2^{-1}C_1 \\ G_{22} &= A_{12}C_2^{-1} \\ G_{23} &= -A_{12}C_2^{-1}D + B_1 \end{aligned} \tag{10}$$

An observer is designed mimicking system (9)

$$\begin{aligned} \dot{\hat{x}}_1 &= G_{21}\hat{x}_1 + G_{22}\hat{y} \\ \dot{\hat{y}} &= G_{11}\hat{x}_1 + G_{12}\hat{y} + \upsilon \end{aligned} \tag{11}$$

where $\upsilon \in \mathbb{R}^p$ is the injection term. The estimation errors are introduced

$$e_y = y - \hat{y},\ e_{x_1} = x_1 - \hat{x}_1 \tag{12}$$

The following assumptions are made concerning matrices in (9) and (10):

(A2): The matrix G_{21} is Hurwitz.

(A3): The entries of the matrix transfer function $G_{11}\left(sI - G_{21}\right)^{-1}G_{23} + G_{13} + Ds$ have numerators with the roots located in the left hand side of the complex plane (a minimum phase case). Here s is the Laplace variable.

(A4): For the term $\varphi = G_{11}e_{x_1} + G_{12}e_y + G_{13}d + D\dot{d}$ the following inequality holds at least locally:

$$\|\varphi\| \le L_{G_{11}} L_{e_{x_1}} + L_{G_{12}} L_{e_y} + L_{G_{13}} L_1 + L_D L_2 \le L_3 \tag{13}$$

where $\|G_{11}\|_\infty \le L_{G_{11}}, \|G_{12}\|_\infty \le L_{G_{12}}, \|G_{13}\|_\infty \le L_{G_{13}}, \|D\|_\infty \le L_D, \|e_{x_1}\| \le L_{e_{x_1}}, \|e_y\| \le L_{e_y}$, $L_{G_{11}}, L_{e_{x_1}}, L_{G_{12}}, L_{e_y}, L_{G_{13}}, L_1, L_D, L_2, L_3 > 0$.

4.1.1 Sensor attack reconstruction via fixed-gain SMO

The first main result is formulated in the following Theorem:

Theorem 1: Consider the system in (9) and (10) with the observer (11), whose injection term υ is designed in a unit vector format

$$\upsilon = (\rho + L_3)\frac{e_y}{\|e_y\|}, \ \rho, L_3 > 0 \tag{14}$$

that makes the observer (11) and (14) the SMO. Assume that the assumptions A1-A4 hold.

Then the sensor attack signal $d(t)$ is exactly reconstructed as

$$\hat{d} = \left(G_{11}(sI - G_{21})^{-1} G_{23} + G_{13} + Ds\right)^{-1} \upsilon_{eq} \tag{15}$$

where υ_{eq} is the *equivalent* injection function, $e_y \to 0$ in finite time, and $\hat{d}(t) \to d(t)$ as time increases in the sliding mode.

The proof of the Theorem 1 is presented in Appendix.

Remark 1: Given *equivalent* control υ_{eq}, the attack estimate in (15), where the dynamic filter appears naturally, is exact.

Remark 2: Although the equivalent control υ_{eq} was conceived as an abstraction to allow the analysis of the reduced order sliding motion, a close approximation can be obtained in real-time by low-pass filtering of the switching signal (14) [25]. Therefore, if $\bar{\upsilon}_{eq}$ satisfies

$$\tau\dot{\bar{\upsilon}}_{eq} = (\rho + L_1)\frac{e_y}{\|e_y\|} - \bar{\upsilon}_{eq} \tag{16}$$

where $\tau > 0$ is a (small) time constant, then

$$\|\bar{\upsilon}_{eq} - \upsilon_{eq}\| \sim \mathrm{O}(\tau) \tag{17}$$

Therefore, the υ_{eq} estimation error in (17) is small, for a small enough choice of τ [26].

Replacing υ_{eq} by $\bar{\upsilon}_{eq}$ in (15) we obtain

$$\bar{d} = \left(G_{11}\left(sI - G_{21}\right)^{-1} G_{23} + G_{13} + Ds\right)^{-1} \bar{\upsilon}_{eq} \tag{18}$$

and the attack estimation error after a transient is over can be computed

$$\left\|\bar{d} - d\right\| \le \Lambda \left\|\bar{\upsilon}_{eq} - \upsilon_{eq}\right\| \sim \mathrm{O}(\tau) \tag{19}$$

where

$$\left\|\left(G_{11}\left(sI - G_{21}\right)^{-1} G_{23} + G_{13} + Ds\right)^{-1}\right\|_{\infty} = \Lambda,\ \ \Lambda > 0 \tag{20}$$

Note that the low pass filter in (16) is the simplest choice, but other higher order systems with low-pass characteristics can be employed.

Remark 3: In many practical cases the entries of the transfer function $\left(G_{11}\left(sI - G_{21}\right)^{-1} G_{23} + G_{13} + Ds\right)^{-1}$ of the estimator (15) are the regular ones. This fact is demonstrated in the case study. It means that the SMC injection term υ in (14) can be used in (15) instead of υ_{eq}, bearing in mind that υ_{eq} is recovered/estimated approximately via the low pass filtering of υ that takes place while υ is processed by the proper transfer function $\left(G_{11}\left(sI - G_{21}\right)^{-1} G_{23} + G_{13} + Ds\right)^{-1}$.

4.1.2 Sensor attack reconstruction via adaptive-gain SMO

In subsection 4.1.1 it was assumed that the perturbations term φ is locally norm-bounded as in (13), and the boundary $L_3 > 0$ is known. In many practical cases this bound is unknown, and the gain of the sliding mode injection term (14) in the fixed gain SMO in (15) can be overestimated. This gain overestimation could increase chattering that is difficult to attenuate.

In this section an adaptive-gain SMO is considered for the cyber-attack on-line reconstruction. The following assumption is made:

Assumption (A5): The disturbance term φ satisfies the conditions

$$\|\varphi\| \le L_3, \|\dot{\varphi}\| \le L_4, \tag{21}$$

where $L_3, L_4 > 0$ exist but are *unknown*.

The *dual layer nested adaptive SMO* [27] is used for designing the injection term υ in (13). In accordance with the *dual layer nested adaptive sliding mode observation algorithm* [27] the constant gain L_3 in the injection term (14) is to be replaced by the adaptive gain $L(t)$ (without $L(t)$ overestimation). This is

$$\upsilon = \left(\rho + L(t)\right)\frac{e_y}{\left\|e_y\right\|},\ \ \rho > 0 \tag{22}$$

Following the *dual layer nested sliding mode observation adaptive algorithm* in [27] applied to the unit-vector injection term in (14), an error signal is defined as

$$\sigma(t) = L(t) - \frac{1}{\alpha}\left\|\bar{\upsilon}_{eq}(t)\right\| - \varepsilon \tag{23}$$

where the scalars $0 < \alpha < 1$, $\varepsilon > 0$, and $\bar{\upsilon}_{eq}$ represents a low-pass filtered estimate of υ_{eq} obtained as

$$\tau\dot{\bar{\upsilon}}_{eq} = \upsilon - \bar{\upsilon}_{eq} \tag{24}$$

The task of selecting $\tau > 0$ is discussed in Remark 2.

The adaptation dynamics of $L(t)$ in (22) are defined as [27]

$$\dot{L}(t) = -r(t)sign(\sigma(t)) \tag{25}$$

where $r(t) > 0$ is a time-varying scalar that is supposed to supersede the upper-bound of the rate of change of the generalized attack, $\left\|\dot{\varphi}\right\| \le L_4$, by some finite time. In this paper it is assumed that $r(t)$ has the structure

$$r(t) = \ell_0 + \ell(t) \tag{26}$$

where ℓ_0 is a fixed positive scalar. The evolution of $\ell(t)$ is chosen to satisfy an adaptive law [27]

$$\dot{\ell}(t) = \begin{cases} \gamma\left|\sigma(t)\right| & if\ \left|\sigma(t)\right| > \sigma_0 \\ 0 & otherwise \end{cases} \tag{27}$$

where $\gamma, \sigma_0 > 0$ are design scalars.

The second main result is summarized in the following proposition.

Proposition 1: Consider the system in (9) and (10), and assume that the assumptions A1 – A5, hold. A SMO is designed as in (11) with the *adaptive* injection term in (22)-(27). If $\varepsilon > 0$ in (23) is chosen to satisfy

$$\frac{1}{4}\varepsilon^2 > \sigma_0^2 + \frac{1}{\gamma}\left(\frac{qL_4}{\alpha}\right)^2 \tag{28}$$

for any given σ_0 in (27), L_4 in (21), $q > 1$, and, $0 < \alpha < 1$, then

- the injection term (22) exploiting the *dual layer adaptive* scheme given by (23)-(27), drives $\sigma(t)$ to a domain $\left|\sigma(t)\right| < \varepsilon/2$ in finite time and consequently ensures a sliding motion $e_y = 0$ can be reached in finite time and sustained thereafter. Furthermore, the gains $r(t)$ and $L(t)$ remain bounded;
- the sensor attack signal $d(t)$ is reconstructed in the sliding mode as time increases as in (15) with the equivalent adaptive injection term υ_{eq} or $\bar{\upsilon}_{eq}$.

The proof of the Proposition 1 is presented in the Appendix.

Remark 4: The proposed unit vector injection gain-adaptation algorithm in (22)-(27) does not require the knowledge of the boundaries $L_3, L_4 > 0$.

4.2 k sensors ($k < p,\ \ p = m$) are protected from the attacks

Again consider the system (5), whose sensors are under the attacks.

Assumption (A6): k out of p sensors are protected, and the remaining $p-k$ sensors might be attacked/corrupted.

Separating the protected and unprotected measurements, system (5) can be partitioned as

$$\begin{gathered}
\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} d \\
\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ D_1 \end{bmatrix} d
\end{gathered} \tag{29}$$

where

$$\begin{aligned}
&\bar{y} = \left[y_1^T, y_2^T \right]^T, y_1 \in \mathbb{R}^k, y_2, \in \mathbb{R}^{p-k} \\
&x = \left[x_1^T, x_2^T, x_3^T \right]^T, x_1 \in \mathbb{R}^{n-p}, x_2 \in \mathbb{R}^k, x_3 \in \mathbb{R}^{p-k} \\
&B = \left[B_1^T, B_2^T, B_3^T \right]^T, B_1 \in \mathbb{R}^{(n-p)\times p}, B_2 \in \mathbb{R}^{k\times p}, B_3 \in \mathbb{R}^{(p-k)\times p} \\
&D_1 \in \mathbb{R}^{(p-k)\times p}
\end{aligned} \tag{30}$$

It is assumed that

Assumption (A7): The square matrices $C_{12} \in \mathbb{R}^{k \times k}$ and $C_{23} \in \mathbb{R}^{(p-k)\times(p-k)}$ are non-singular.

Next, the partitioned system (29), (30) is transformed to a convenient for the Lyapunov analysis form. Specifically, the state variables $x_2 \in \mathbb{R}^k$ and $x_3 \in \mathbb{R}^{p-k}$ are replaced by the output variables $y_1 \in \mathbb{R}^k$ and $y_2 \in \mathbb{R}^{p-k}$. This is:

$$\begin{aligned} \dot{x}_1 &= Q_{11}x_1 + Q_{12}y_1 + Q_{13}y_2 + Q_{14}d \\ \dot{y}_1 &= Q_{21}x_1 + Q_{22}y_1 + Q_{23}y_2 + Q_{24}d \\ \dot{y}_2 &= Q_{31}x_1 + Q_{32}y_1 + Q_{33}y_2 + Q_{34}d + D_1\dot{d} \end{aligned} \tag{31}$$

where

$$\begin{aligned} Q_{11} &= A_{11} + A_{12}h_{11} + A_{13}h_{21} \\ Q_{12} &= A_{12}h_{12} + A_{13}h_{22} \\ Q_{13} &= A_{12}h_{13} + A_{13}h_{23} \\ Q_{14} &= A_{12}h_{14} + A_{13}h_{24} + B_1 \\ Q_{21} &= C_{11}A_{11} + C_{12}A_{21} + C_{13}A_{31} \\ &\quad + \left(C_{11}A_{12} + C_{12}A_{22} + C_{13}A_{32}\right)h_{11} + \left(C_{11}A_{13} + C_{12}A_{23} + C_{13}A_{33}\right)h_{21} \\ Q_{22} &= \left(C_{11}A_{12} + C_{12}A_{22} + C_{13}A_{32}\right)h_{12} + \left(C_{11}A_{13} + C_{12}A_{23} + C_{13}A_{33}\right)h_{22}, \\ Q_{23} &= \left(C_{11}A_{12} + C_{12}A_{22} + C_{13}A_{32}\right)h_{13} + \left(C_{11}A_{13} + C_{12}A_{23} + C_{13}A_{33}\right)h_{23}, \\ Q_{24} &= C_{11}B_1 + C_{12}B_2 + C_{13}B_3 \\ Q_{31} &= C_{21}A_{11} + C_{22}A_{21} + C_{23}A_{31} \\ &\quad + \left(C_{21}A_{12} + C_{22}A_{22} + C_{23}A_{32}\right)h_{11} + \left(C_{21}A_{13} + C_{22}A_{23} + C_{23}A_{33}\right)h_{21} \\ Q_{32} &= \left(C_{21}A_{12} + C_{22}A_{22} + C_{23}A_{32}\right)h_{12} + \left(C_{21}A_{13} + C_{22}A_{23} + C_{23}A_{33}\right)h_{22}, \\ Q_{33} &= \left(C_{21}A_{12} + C_{22}A_{22} + C_{23}A_{32}\right)h_{13} + \left(C_{21}A_{13} + C_{22}A_{23} + C_{23}A_{33}\right)h_{23}, \\ Q_{34} &= C_{21}B_1 + C_{22}B_2 + C_{23}B_3 \end{aligned} \tag{32}$$

with

$$
\begin{aligned}
h_{11} &= \left(I - C_{12}{}^{-1}C_{13}C_{23}{}^{-1}C_{22}\right)^{-1} C_{12}{}^{-1}\left(-C_{11} + C_{13}C_{23}{}^{-1}C_{21}\right)\\
h_{12} &= \left(I - C_{12}{}^{-1}C_{13}C_{23}{}^{-1}C_{22}\right)^{-1} C_{12}{}^{-1}\\
h_{13} &= -\left(I - C_{12}{}^{-1}C_{13}C_{23}{}^{-1}C_{22}\right)^{-1} C_{12}{}^{-1}C_{13}C_{23}{}^{-1}\\
h_{14} &= -\left(I - C_{12}{}^{-1}C_{13}C_{23}{}^{-1}C_{22}\right)^{-1} C_{12}{}^{-1}C_{13}C_{23}{}^{-1}D_1\\
h_{21} &= \left(I - C_{23}{}^{-1}C_{22}C_{12}{}^{-1}C_{13}\right)^{-1} C_{23}{}^{-1}\left(-C_{21} + C_{22}C_{12}{}^{-1}C_{11}\right)\\
h_{22} &= -\left(I - C_{23}{}^{-1}C_{22}C_{12}{}^{-1}C_{13}\right)^{-1} C_{23}{}^{-1}C_{22}C_{12}{}^{-1}\\
h_{23} &= \left(I - C_{23}{}^{-1}C_{22}C_{12}{}^{-1}C_{13}\right)^{-1} C_{23}{}^{-1}\\
h_{24} &= -\left(I - C_{23}{}^{-1}C_{22}C_{12}{}^{-1}C_{13}\right)^{-1} C_{23}{}^{-1}D_1
\end{aligned}
\qquad (33)
$$

The main results relating to sensor attack reconstruction in systems with $k < p,\ \ p = m$ sensors protected from the attacks are presented in the following Theorem.

Theorem 2: Consider system in (31)-(33), and assume that assumption A7 holds. The proposed SMO is given by

$$
\begin{aligned}
\dot{\hat{x}}_1 &= Q_{11}\hat{x}_1 + Q_{12}\hat{y}_1 + Q_{13}\hat{y}_2\\
\dot{\hat{y}}_1 &= Q_{21}\hat{x}_1 + Q_{22}\hat{y}_1 + Q_{23}\hat{y}_2 + \upsilon_1\\
\dot{\hat{y}}_2 &= Q_{31}\hat{x}_1 + Q_{32}\hat{y}_1 + Q_{33}\hat{y}_2 + \upsilon_2
\end{aligned}
\qquad (34)
$$

where $\upsilon_1 \in \mathbb{R}^k$ and $\upsilon_2 \in \mathbb{R}^{p-k}$ are sliding mode injection terms that are defined as

$$
\begin{aligned}
\upsilon_1 &= \left(\rho_1 + L_{11}\right)\frac{e_{y_1}}{\left\|e_{y_1}\right\|}\\
\upsilon_2 &= \left(\rho_2 + L_{12}\right)\frac{e_{y_2}}{\left\|e_{y_2}\right\|}
\end{aligned}
\qquad (35)
$$

with $\rho_1, \rho_2, L_{11}, L_{12} > 0$ and

$$
e_{x_1} = x_1 - \hat{x}_1,\quad e_{y_1} = y_1 - \hat{y}_1,\quad e_{y_2} = y_2 - \hat{y}_2 \qquad (36)
$$

then the sensor cyber-attack is exactly estimated as

$$
\hat{d} = \begin{bmatrix} \hat{d}_1 \\ \hat{d}_2 \end{bmatrix} = \begin{bmatrix} H_{11}{}^{-1}\left(\upsilon_{1eq} - H_{12}\hat{d}_2\right) \\ \left(-H_{21}H_{11}{}^{-1}H_{12} + H_{22}\right)^{-1}\left(\upsilon_{2eq} - H_{21}H_{11}{}^{-1}\upsilon_{1eq}\right) \end{bmatrix} \qquad (37)
$$

in the sliding mode $\forall t \geq t_r$ where $t = t_r$ is sliding mode reaching time, and $\hat{d}_1 \in \mathbb{R}^k$, $\hat{d}_2 \in \mathbb{R}^{p-k}$, where the matrix $\begin{bmatrix} Q_{21}\left(sI - Q_{11}\right)^{-1} Q_{14} + Q_{24} \\ Q_{31}\left(sI - Q_{11}\right)^{-1} Q_{14} + Q_{34} + D_1 sI \end{bmatrix} \in \mathbb{R}^{p \times p}$ is partitioned as

$$\begin{bmatrix} Q_{21}\left(sI - Q_{11}\right)^{-1} Q_{14} + Q_{24} \\ Q_{31}\left(sI - Q_{11}\right)^{-1} Q_{14} + Q_{34} + D_1 sI \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \tag{38}$$

while the matrices $H_{11} \in \mathbb{R}^{k \times k}$, $H_{12} \in \mathbb{R}^{k \times (p-k)}$, $H_{21} \in \mathbb{R}^{(p-k) \times k}$, $H_{22} \in \mathbb{R}^{(p-k) \times (p-k)}$. The proof of the Theorem 2 is presented in Appendix.

Remark 5: The injection terms υ_1, υ_2 in (35) can be also designed in the dual layer adaptive form as shown in section IV. Note that the problem of estimating $\upsilon_{1eq}, \upsilon_{2eq}$ used in (37) is discussed in Remark 2.

Remark 6: Estimating the attack vector $d = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$ in (37) requires inversion of the matrices (that represent the dynamic filters) of smaller dimensions rather than in (15) due the fact that $k < p$ sensors are protected from the cyber-attacks.

4.3 All sensors $p > m$ can be attacked.

Since $p > m$, the system (5) can be partitioned using a nonsingular transformation $M \in \mathbb{R}^{m \times m}$

$$y = M\bar{y} \tag{39}$$

that is selected so that

$$M^{-1}D = \begin{bmatrix} \mathbf{0}_{(p-m) \times m} \\ \bar{D}_{m \times m} \end{bmatrix} \tag{40}$$

Using (40), system (5) is reduced to

$$\begin{aligned} \dot{x} &= Ax + Bd \\ \bar{y}_1 &= \bar{C}_1 x \\ \bar{y}_2 &= \bar{C}_2 x + \bar{D}d \end{aligned} \tag{41}$$

where $\bar{y}_1 \in \mathbb{R}^{p-m}$ and $\bar{y}_2 \in \mathbb{R}^{m}$ are parts of $\bar{y}$ as $\bar{y} = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \end{bmatrix}$, and $\bar{C}_1 \in \mathbb{R}^{p-m}, \bar{C}_2 \in \mathbb{R}^{m}$ are partitions of $\bar{C} = M^{-1}C$.

Remark 7. In (41) the virtual measurements $\bar{y}_1, \bar{y}_2$ are available, and $\bar{y}_1$ can be interpreted as a *virtual sensor* measurement protected from cyber-attacks.

After further state vector partitioning, system (41) may be presented in a (29) format. Then the algorithms presented in Sub-section 4.2 can be employed for reconstructing the cyber-attack $d(t)$ in system (5) in the case $p > m$ **.**

5. Case study: Sensor attacks On-line Reconstruction in Electrical Power Network

5.1. Mathematical model of electrical power network

The descriptor (Differential Algebraic Equations (DAE)) swing mathematical model is used to describe the electromechanical behavior of an electric power network [28, 29].

The linearized DAE swing mathematical model for an electrical power network stabilized by a linear output feedback controller is given by [29]:

$$\begin{bmatrix} I & 0 & 0 \\ 0 & M_g & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{\delta} \\ \dot{\omega} \\ \dot{\theta} \end{bmatrix} = -\underbrace{\begin{bmatrix} 0 & -I & 0 \\ L^{\theta}_{g,g} & E_g & L^{\theta}_{g,l} \\ L^{\theta}_{l,g} & 0 & L^{\theta}_{l,l} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \delta \\ \omega \\ \theta \end{bmatrix}}_{x} + \underbrace{\begin{bmatrix} 0 \\ B_{\omega} \\ B_{\theta} \end{bmatrix}}_{B} d(t) + \begin{bmatrix} 0 \\ P_{\omega} \\ P_{\theta} \end{bmatrix} \qquad (42)$$

$$y = Cx + Dd(t)$$

where $x = \begin{bmatrix} \delta^T & \omega^T & \theta^T \end{bmatrix}^T$ is the vector of states of the system, $\delta \in \mathbb{R}^a$, $\omega \in \mathbb{R}^a$, $\theta \in \mathbb{R}^b$ are vectors of the phase angles of the source measured in *rad*, generator speed *deviations* from synchronous measured in rad/s , and of bus angles measured in *rad* respectively. The index a is the number of generators, and b is the number of buses in the electrical system. The vector $y \in \mathbb{R}^p$ is the measurement vector, the vector $d \in \mathbb{R}^m$ is the sensor attack vector, and $B \in \mathbb{R}^{(2a+b)\times m}$, $D \in \mathbb{R}^{p\times m}$ are the attack distribution matrices; P_{ω}, P_{θ} are known inputs, corresponding to mechanical torque and power demand. The matrices $E_g, M_g \in \mathbb{R}^{a\times a}$ are diagonal matrices whose nonzero entries consist of the damping coefficients and the normalized inertias of the generators respectively. Finally, the matrices $L_{g,g}^{\theta}$, $L_{g,l}^{\theta}$, $L_{l,g}^{\theta}$, $L_{l,l}^{\theta}$ form the following symmetric susceptance matrix

$$L^{\theta} = \begin{bmatrix} L_{g,g}^{\theta} & L_{g,l}^{\theta} \\ L_{l,g}^{\theta} & L_{l,l}^{\theta} \end{bmatrix} \tag{43}$$

that is the Laplacian associated with the susceptance-weighted graph.

Assumption (A8) The matrix $L_{l,l}^{\theta}$ is nonsingular (such an assumption usually holds in practical electric power systems).

Note that the terms that appear in the electric power network model (42)

$$\begin{bmatrix} 0 \\ B_{\omega} \\ B_{\theta} \end{bmatrix} d(t) + \begin{bmatrix} 0 \\ P_{\omega} \\ P_{\theta} \end{bmatrix} \tag{44}$$

are due to the output feedback control that processes the output $y = Cx + Dd(t)$ corrupted by the cyber-attack signal $d(t)$.

5.2. *Transformation of DAE (42) to ODE*

Assuming (A8) holds, the variable θ can be expressed as

$$\theta = \left(R_{l,l}^{\theta}\right)^{-1}\left(-R_{l,g}^{\theta}\delta + P_{\theta} + B_{\theta}d\right) \tag{45}$$

Substituting (45) into (42) we obtain

$$\begin{gathered} \begin{bmatrix} \dot{\delta} \\ \dot{\omega} \end{bmatrix} = \begin{bmatrix} \varphi_{\delta}(\delta,\omega) \\ \varphi_{\omega}(\delta,\omega) \end{bmatrix} + \begin{bmatrix} 0 \\ P_{\theta\omega} \end{bmatrix} + \begin{bmatrix} 0 \\ B_{\theta\omega} \end{bmatrix} d(t) \\ y = C\begin{bmatrix} \delta \\ \omega \end{bmatrix} + Dd(t) \end{gathered} \tag{46}$$

where

$$\begin{gathered} \begin{bmatrix} \varphi_{\delta}(\delta,\omega) \\ \varphi_{\omega}(\delta,\omega) \end{bmatrix} = \begin{bmatrix} 0 & I_{p\times p} \\ M_g^{-1}\left(-R_{g,g}^{\theta} + R_{g,l}^{\theta}\left(R_{l,l}^{\theta}\right)^{-1}R_{l,g}^{\theta}\right) & -M_g^{-1}E_g \end{bmatrix} \begin{bmatrix} \delta \\ \omega \end{bmatrix} \\ P_{\theta\omega} = M_g^{-1}\left(P_{\omega} - R_{g,l}^{\theta}\left(R_{l,l}^{\theta}\right)^{-1}P_{\theta}\right), \quad B_{\theta\omega} = M_g^{-1}\left(B_{\omega} - R_{g,l}^{\theta}\left(R_{l,l}^{\theta}\right)^{-1}B_{\theta}\right) \end{gathered} \tag{47}$$

and

$$C = \begin{bmatrix} C_{\delta} & 0 \\ 0 & C_{\omega} \end{bmatrix}, D = \begin{bmatrix} D_{\delta} \\ D_{\omega} \end{bmatrix}$$

5.3. System (46), (47) Parameterization

The electrical power system considered in this paper is a classical nine-bus configuration adopted from [29]. It consists of 3 generators $\{g_1, g_2, g_3\}$ and *6* load buses $\{b_1, ..., b_6\}$ as presented in Fig. 3 [29].

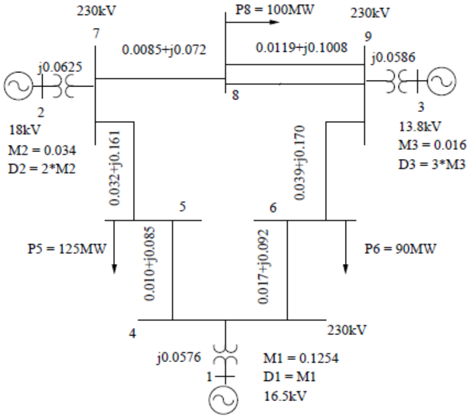

Fig. 3. The Western Electricity Coordinating Council power system [29]

Consider that the measurements of the system (46) are $\omega \in \mathbb{R}^3$, i.e.

$$C_\delta = D_\delta = 0, \quad C_\omega = D_\omega = I_{3\times 3} \tag{48}$$

and

$$B_\omega = I_3, \; B_\theta = 0_{6\times 3} \tag{49}$$

Therefore, system (46) is rewritten for this parameterized case as

$$\begin{bmatrix} I & 0 & 0 \\ 0 & M_g & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{\delta} \\ \dot{\omega} \\ \dot{\theta} \end{bmatrix} = - \begin{bmatrix} 0 & -I & 0 \\ R^\theta_{g,g} & E_g & R^\theta_{g,l} \\ R^\theta_{l,g} & 0 & R^\theta_{l,l} \end{bmatrix} \begin{bmatrix} \delta \\ \omega \\ \theta \end{bmatrix} + \begin{bmatrix} 0 \\ I \\ 0 \end{bmatrix} d(t) + \begin{bmatrix} 0 \\ P_\omega \\ P_\theta \end{bmatrix} \tag{50}$$
$$y = \omega + d(t)$$

where $\omega = [\omega_1 \quad \omega_2 \quad \omega_3]^T \in \mathbb{R}^3$, $\delta = [\delta_1 \quad \delta_2 \quad \delta_3]^T \in \mathbb{R}^3$, $\theta \in \mathbb{R}^6$, $y = [y_1 \quad y_2 \quad y_3]^T \in \mathbb{R}^3$, and $d = [d_1 \quad d_2 \quad d_3]^T \in \mathbb{R}^3$ is the sensor cyber-attack signal. Matrices M_g and E_g are the diagonal matrices of the generator inertial and damping coefficients

$$M_g = \begin{bmatrix} 0.125 & 0 & 0 \\ 0 & 0.034 & 0 \\ 0 & 0 & 0.016 \end{bmatrix}, \; E_g = \begin{bmatrix} 0.125 & 0 & 0 \\ 0 & 0.068 & 0 \\ 0 & 0 & 0.048 \end{bmatrix} \tag{51}$$

The Laplacian matrix associated with the susceptibility-weighted graph is the symmetric susceptibility matrix $L^{\theta} \in \mathbb{R}^{9\times 9}$ given by

$$L^{\theta} = \begin{bmatrix} 0.058 & 0 & 0 & -0.058 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.063 & 0 & 0 & -0.063 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.059 & 0 & 0 & -0.059 & 0 & 0 & 0 \\ -0.058 & 0 & 0 & 0.235 & 0 & 0 & -0.085 & -0.092 & 0 \\ 0 & -0.063 & 0 & 0 & 0.296 & 0 & -0.161 & 0 & -0.072 \\ 0 & 0 & -0.059 & 0 & 0 & 0.330 & 0 & -0.170 & -0.101 \\ 0 & 0 & 0 & -0.085 & -0.161 & 0 & 0.246 & 0 & 0 \\ 0 & 0 & 0 & -0.092 & 0 & -0.170 & 0 & 0.262 & 0 \\ 0 & 0 & 0 & 0 & -0.072 & -0.101 & 0 & 0 & 0.173 \end{bmatrix} \tag{52}$$

The system (50) is reduced to

$$\begin{cases} \dot{\delta} = \omega \\ \dot{\omega} = \varphi_{\omega}(\delta, \omega) + P_{\theta\omega} + M_g^{-1} d(t) \\ y = \omega + d(t) \end{cases} \tag{53}$$

and can be presented in a numerical format

$$\begin{bmatrix} \dot{\delta}_1 \\ \dot{\delta}_2 \\ \dot{\delta}_3 \\ \dot{\omega}_1 \\ \dot{\omega}_2 \\ \dot{\omega}_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ -0.3145 & 0.1187 & 0.1158 & -1 & 0 & 0 \\ 0.4363 & -0.8474 & 0.4111 & 0 & -2 & 0 \\ 0.9046 & 0.8736 & -1.7782 & 0 & 0 & -3 \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \omega_1 \\ \omega_2 \\ \omega_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1.0559 \\ 14.5564 \\ -19.8079 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 8 & 0 & 0 \\ 0 & 29.4118 & 0 \\ 0 & 0 & 62.5 \end{bmatrix} d(t)$$

$$\tag{54}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{bmatrix} + \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}$$

In order to apply the sensor attack reconstruction algorithms proposed in sub-sections 4.1 the system (53) is presented in a form of (9) as

$$\begin{cases} \dot{\delta} = y - d(t) \\ \dot{y} = \varphi_{\omega 1}\delta + \varphi_{\omega 2} y - \varphi_{\omega 2} d(t) + P_{\theta\omega} + M_g^{-1} d(t) + \dot{d}(t) \end{cases} \tag{55}$$

where $\varphi_{\omega}(\delta, \omega)$ is presented in a linearized form as

$$\varphi_{\omega}(\delta, \omega) = \varphi_{\omega 1}\delta + \varphi_{\omega 2} y - \varphi_{\omega 2} d(t) \tag{56}$$

5.4. The observer design: all sensors may be under attack ($p = m = 3$)

Note that in this case, no sensors have been protected, and all sensors might be attacked, therefore the number of sensors under attack could be zero, one, two, or three. It is not known ahead of time if any particular sensor is attacked.

The observer for system (50) is designed in the format of (11)

$$\begin{cases} \dot{\hat{\delta}} = \hat{y} \\ \dot{\hat{y}} = \varphi_{\omega 1}\hat{\delta} + \varphi_{\omega 2}\hat{y} + P_{\theta\omega} + \upsilon \end{cases} \tag{57}$$

where υ is the injection term designed in a format of (14).

Finally, in accordance with (15), the sensor cyber-attack is exactly reconstructed

$$\hat{d} = \left(\frac{-\varphi_{\omega 1}}{s} - \varphi_{\omega 2} + M_g^{-1} + sI \right)^{-1} \upsilon_{eq} \tag{58}$$

Remark 8: The matrix $\left(\frac{-\varphi_{\omega 1}}{s} - \varphi_{\omega 2} + M_g^{-1} + sI_{3\times 3} \right) \in \mathbb{R}^{3\times 3}$ is invertible.

Note that the problem of estimating υ_{eq} used in (58) is discussed in Remarks 2 and 3.

5.5. The SMO design: the first sensor is protected ($k = 1$)

Consider the case when the first sensor, is *protected* from the attack, in other words

$$Dd(t) = \begin{bmatrix} 0 & d_2 & d_3 \end{bmatrix}^T \tag{59}$$

and the output/sensed equations in (54) become

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} \tag{60}$$

To apply the sensor attack reconstruction algorithms proposed in the sub-section 4.2 the system (54) with one protected sensor/measurement is presented in a form of (55) as

$$\begin{aligned} \dot{\delta} &= \begin{bmatrix} y_1 \\ y_2 - d_2 \\ y_3 - d_3 \end{bmatrix} \\ \dot{y} &= \varphi_{\omega 1}\delta + \varphi_{\omega 2} \begin{bmatrix} y_1 \\ y_2 - d_2 \\ y_3 - d_3 \end{bmatrix} + P_{\theta\omega} + M_g^{-1} d(t) + \dot{d}(t) \end{aligned} \tag{61}$$

And the state/cyber-attack observer is designed

$$\begin{aligned} &\dot{\hat{\delta}} = \hat{y} \\ &\dot{\hat{y}} = \varphi_{\omega 1}\hat{\delta} + \varphi_{\omega 2}\hat{y} + P_{\theta\omega} + \upsilon \end{aligned} \tag{62}$$

where υ is the injection term.

Finally, the attack signals d_2, d_3 are exactly estimated

$$\begin{bmatrix} \hat{d}_2 \\ \hat{d}_3 \end{bmatrix} = \left(\frac{-\varphi'_{\omega 1}}{s} - \varphi'_{\omega 2} + \left(M'_g\right)^{-1} + sI_{2\times 2} \right)^{-1} \upsilon_{eq} \tag{63}$$

where, in accordance with (37)

$$\begin{aligned} &\varphi'_{\omega 1} = \begin{bmatrix} -0.8474 & 0.4111 \\ 0.8736 & -1.7782 \end{bmatrix}, \quad \varphi'_{\omega 2} = \begin{bmatrix} -2 & 0 \\ 0 & -3 \end{bmatrix} \\ &\left(M'_g\right)^{-1} = \begin{bmatrix} 29.4118 & 0 \\ 0 & 62.5 \end{bmatrix}, \quad \upsilon_{eq} = \begin{bmatrix} \upsilon_{2eq} \\ \upsilon_{3eq} \end{bmatrix} \end{aligned} \tag{64}$$

Remark 9: The matrix $\left(\frac{-\varphi'_{\omega 1}}{s} \delta - \varphi'_{\omega 2} + \left(M'_g\right)^{-1} + sI_{2\times 2} \right) \in \mathbb{R}^{2\times 2}$ is invertible. Apparently, the invertibility condition presented here is easier to verify than the one in the Remark 7 due to the reduced order of the matrix to be inverted.

Note that the problem of estimating $\upsilon_{2eq}, \upsilon_{3eq}$ used in (63) and (64) is discussed in Remarks 1 and 2.

5.6. Cleaning up the measurements corrupted by the sensor cyber-attacks

As soon as the sensor attacks are exactly reconstructed in (58) or (63), the measurement vector $y = \omega + d(t)$ is to be "cleaned up" from the attack signal as $y_c = y - \hat{d}(t)$. The "cleaned" system (53) becomes

$$\begin{cases} \dot{\delta}_c = \omega_c \\ \dot{\omega}_c = \varphi_\omega(\delta_c, \omega_c) + P_{\theta\omega} + M_g^{-1}\left(d(t) - \hat{d}(t)\right) \\ y_c = \omega_c + \left(d(t) - \hat{d}(t)\right) \end{cases} \tag{65}$$

where δ_c, ω_c, y_c are the states of the system and the output of the system after "cleaning" the measurements respectively. Note that the system (65) converges to

$$\begin{cases} \dot{\delta}_c = \omega_c \\ \dot{\omega}_c = \varphi_\omega(\delta_c, \omega_c) + P_{\theta\omega} \\ y_c = \omega_c \end{cases} \tag{66}$$

as soon as $\hat{d}(t) \to d(t)$.

5.7. Simulation Results

Simulation set-up: The simulation results have been obtained via MATLAB.

Three simulation experiments have been performed using the Electrical Power Network model in (50).

Experiment 1 No sensor attacks are assumed, i. e. $d(t) \equiv 0$.

Experiment 2 It is assumed that the attacker has access to the actual measurement vector $y^T = (\omega_1, \omega_2, \omega_3)^T$, then the cyber-attack named *stealth attack* [14] that completely corrupts the measurement vector is generated as

$$\begin{aligned} d_1 &= -\omega_1 + 2\sin(t) \\ d_2 &= -\omega_2 + \cos(0.5t) \\ d_3 &= -\omega_3 + \sin(t) \end{aligned} \tag{67}$$

Experiment 3 The sensor *stealth attacks* are reconstructed on-line and the measurements are “cleaned up.”

The results of the simulations are presented in Figs. 4-14. In this case study, the attack observations are done by both the fixed-gain sliding mode observer and the adaptive sliding mode observer presented in Section 4.

Experiments 1: The plot presented in Fig. 4 demonstrates the stabilization of the outputs (generator speed *deviations* from synchronous measured in rad/s) at zero as expected.

Experiments 2: Figure 5 demonstrates the dynamics of the measured outputs while the sensors are under *stealth attacks* (67). The effects of the *stealth attack* are observed.

Experiment3: Figure 6 shows the compensated outputs, when the attacks are reconstructed and cleaned from measurements. The cleaned output dynamics practically coincide with the outputs of the systems without attack after a short transient. In Figs. 7-9, the outputs of system in the three scenario which are without attack, corrupted by attack, and compensated after being attacked are compared.

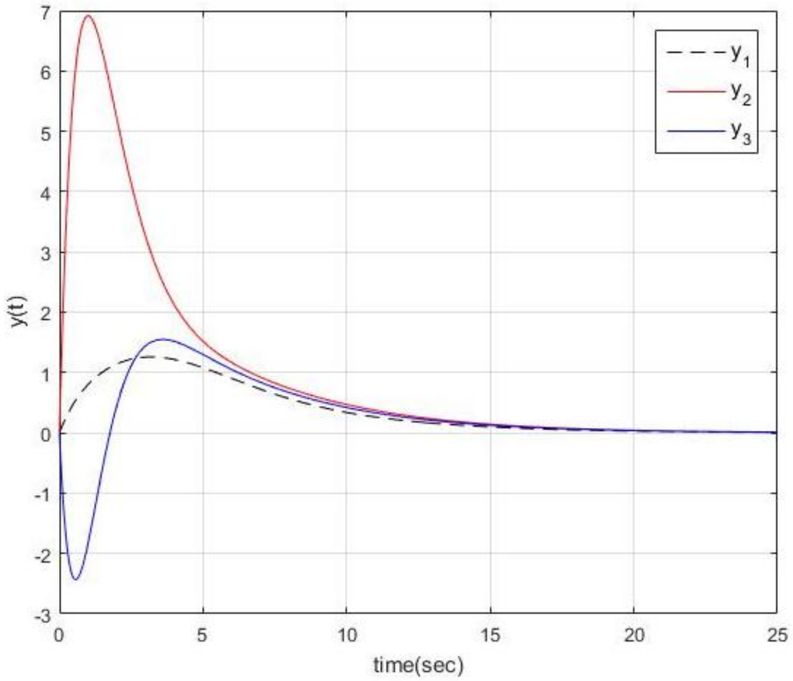

Fig. 4. Outputs of the system (50) without attacks

It is shown in Figs. 10-12 that sensor attacks d_1,d_2,d_3 are accurately estimated by $\hat{d}_1,\hat{d}_2,\hat{d}_3$. The attack observation is done by both the sliding mode observer and the adaptive sliding mode observer presented in Section 4. Figs. 13, 14 show the sliding mode injection terms used in both observers. The output stabilization plots under both observers look the same. However, the main advantage of the adaptive sliding mode observer is in self-tuning.

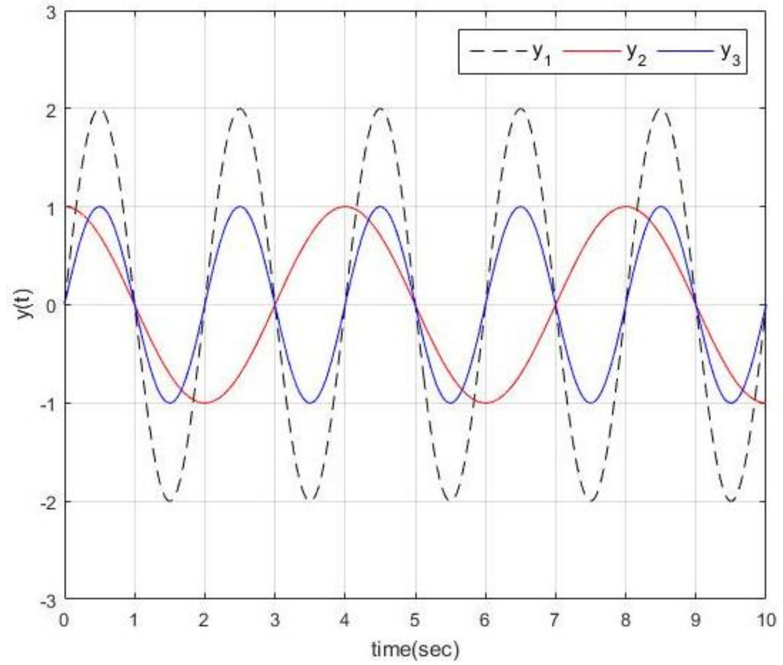

Fig. 5. Outputs of the system (50) under attack

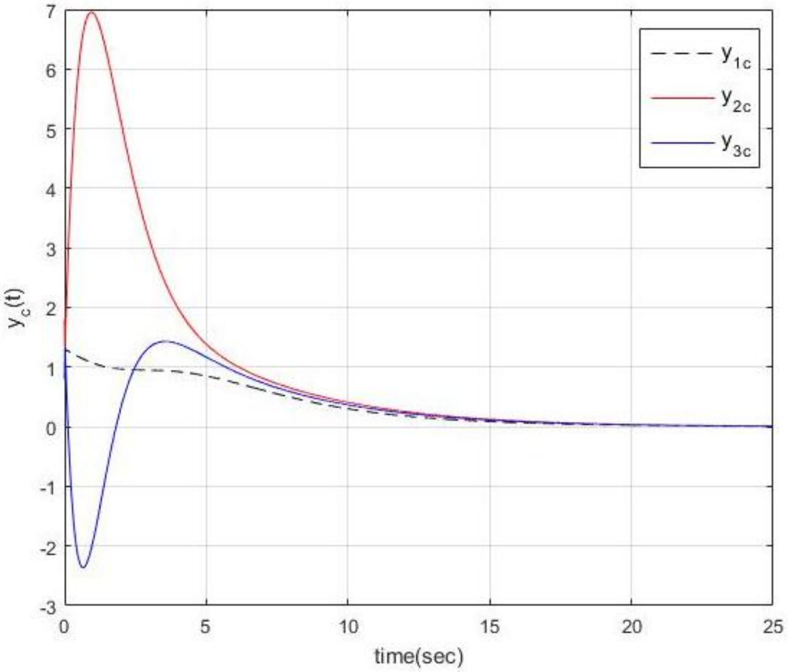

Fig. 6. Compensated outputs of the system (50) as obtained in (65) after the corrupted measurements are cleaned

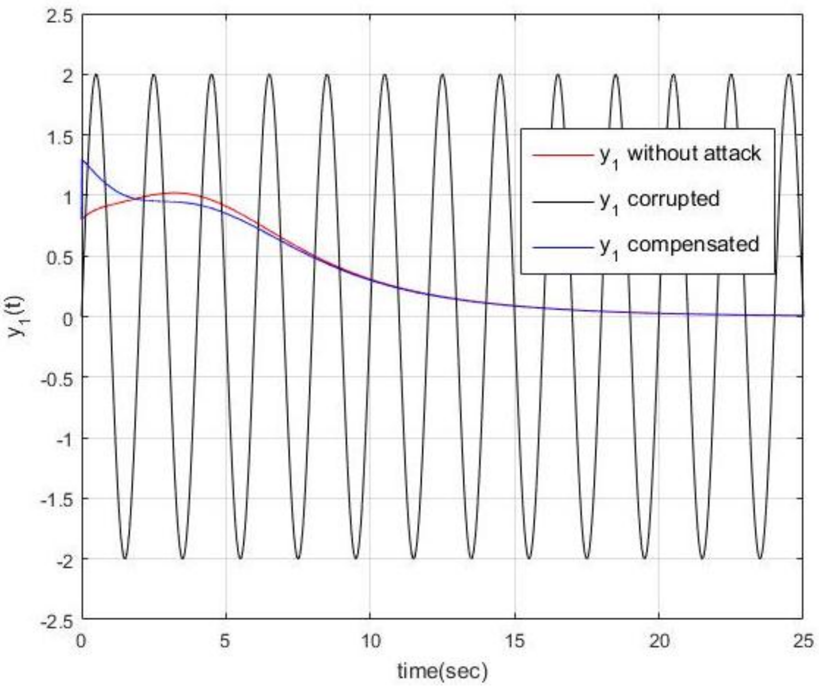

Fig. 7. Comparing y_1 without attack, corrupted y_1, and compensated y_1 after being attacked

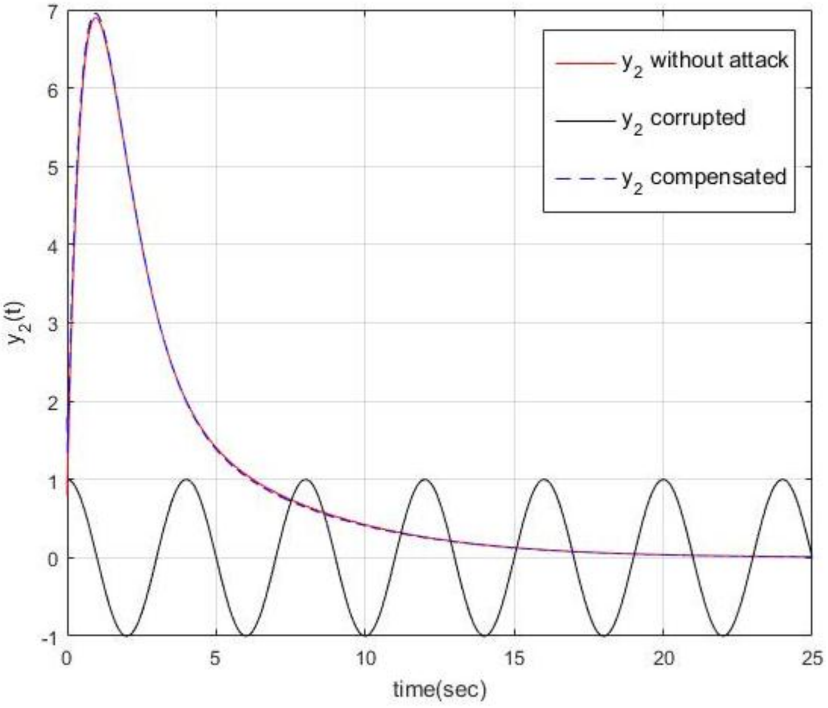

Fig. 8. Comparing y_2 without attack, corrupted y_2, and compensated y_2 after being attacked

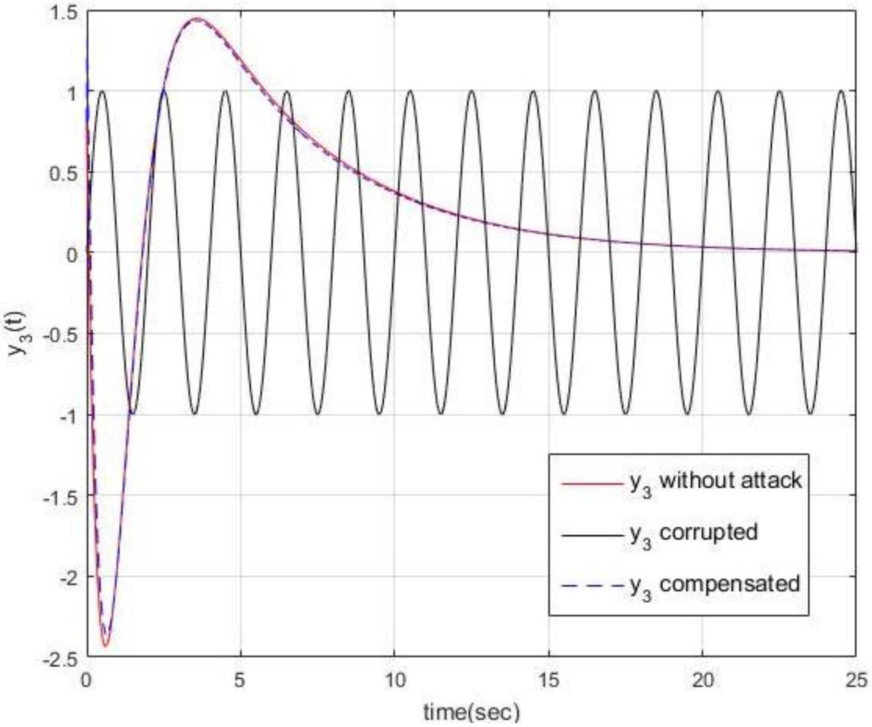

Fig. 9. Comparing y_2 without attack, corrupted y_2, and compensated y_2 after being attacked

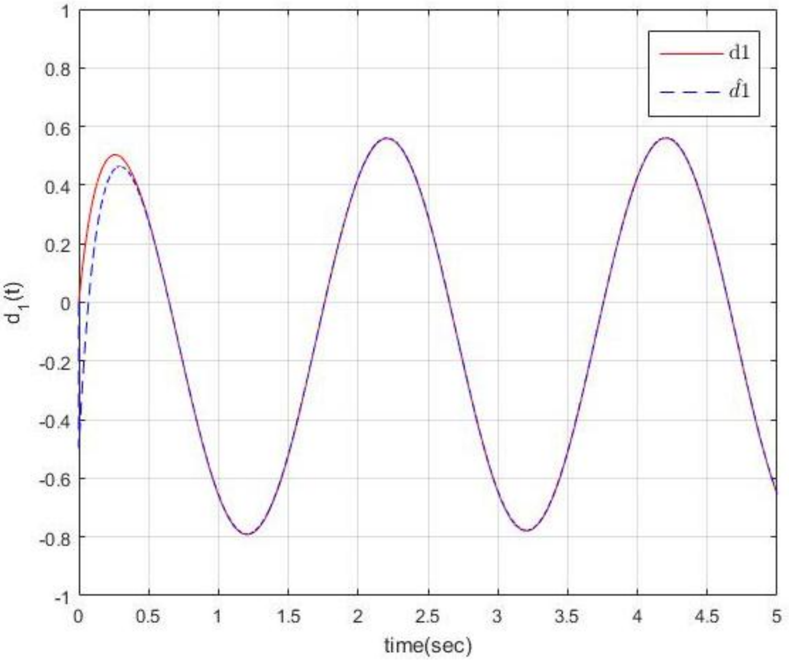

Fig. 10. Comparing sensor attacks d_1 and it's reconstruction $\hat{d}_1$

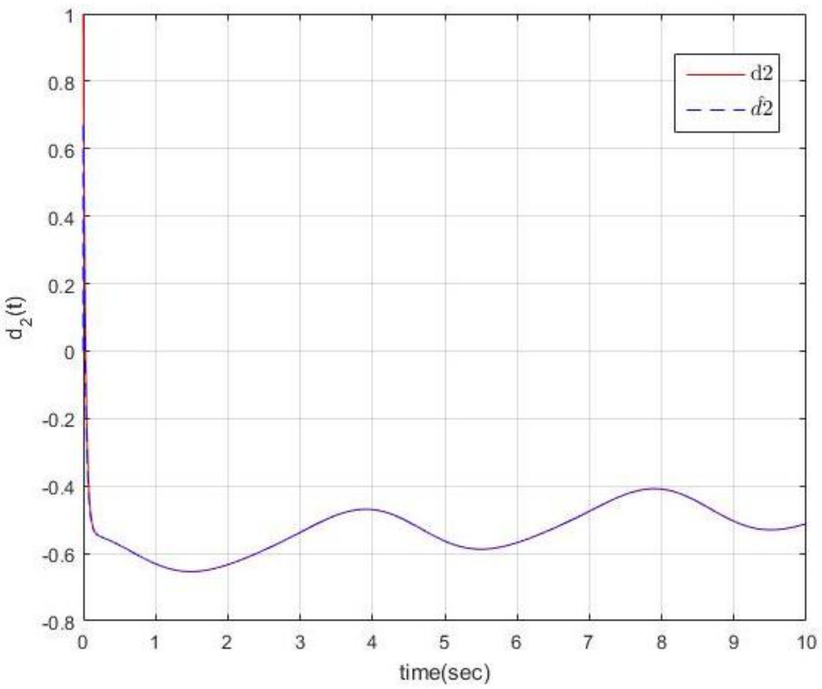

Fig. 11. Comparing sensor attacks d_2 and it's reconstruction $\hat{d}_2$

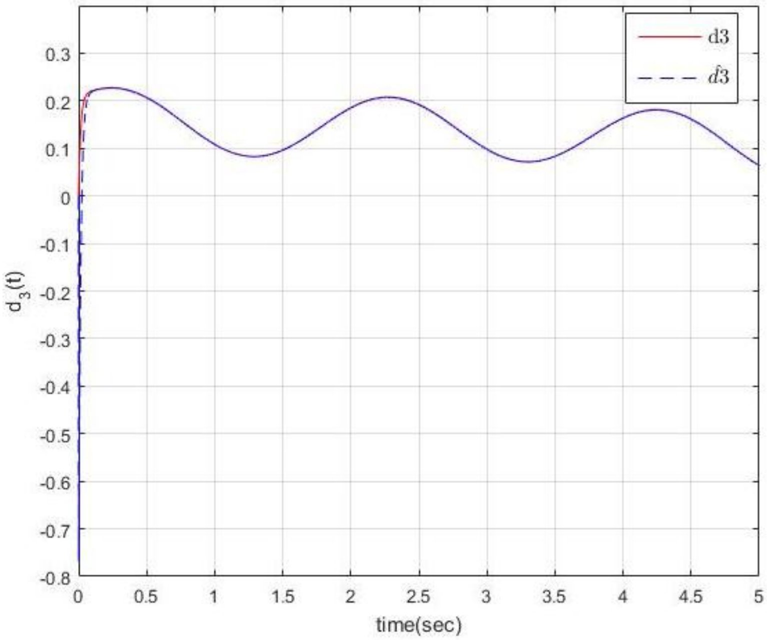

Fig. 12. Comparing sensor attacks d_3 and it's reconstruction $\hat{d}_3$

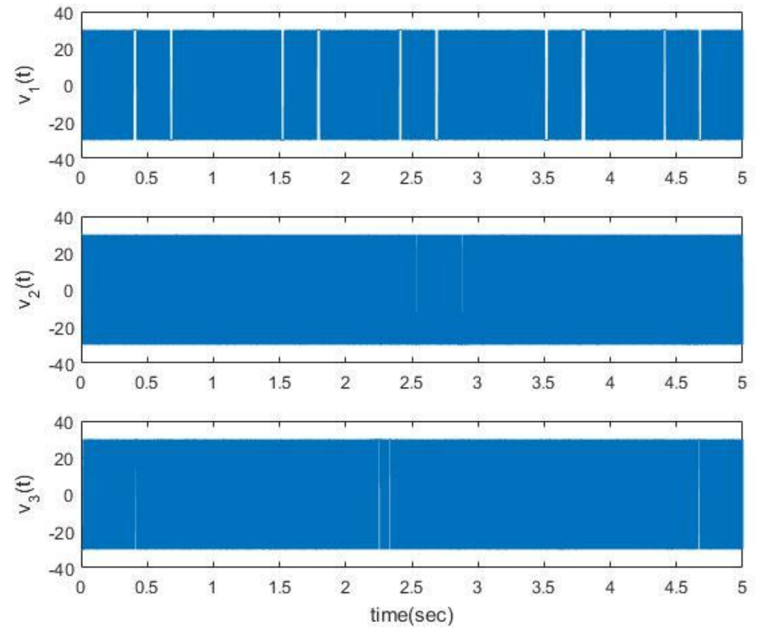

Fig. 13. Sliding mode injection terms $\upsilon_1, \upsilon_2, \upsilon_3$

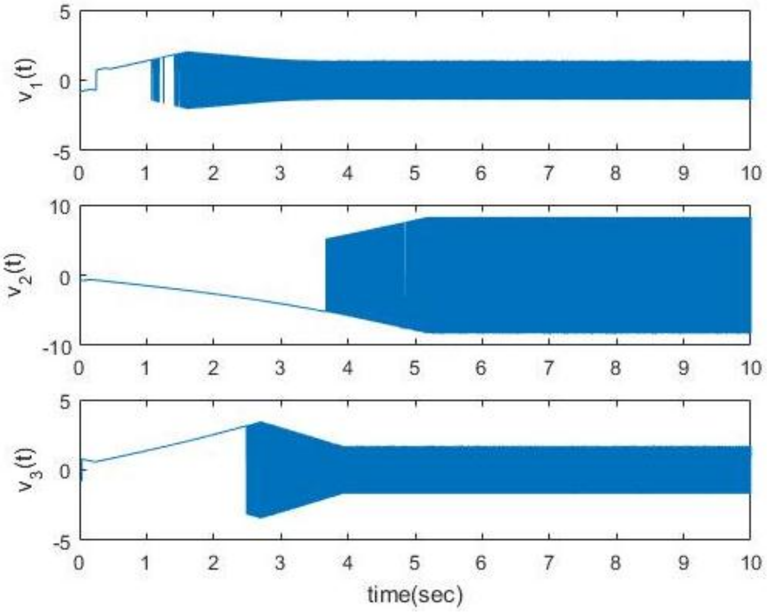

Fig. 14. Adaptive sliding mode injection terms $\upsilon_1, \upsilon_2, \upsilon_3$

6. Conclusion

In this work, linear (linearized) cyber-physical systems under output feedback control, whose sensors are experiencing faults or are under cyber-attack, are investigated. Specifically, the sensor attacks are reconstructed on-line. Two different conditions are considered. Firstly, all of the sensors are prone to get attacked.

Secondly, some of sensors are protected from attacks. Finite time convergent sliding mode observers, including observers with gain adaptation, are proposed for on-line reconstruction of the sensor attacks. The dynamic filters that address the attack propagation dynamics are proposed and employed for the attack reconstruction for the first time. As soon as the attacks are reconstructed, the corrupted measurements are cleaned from attacks, and the feedback control that uses the cleaned measurements/outputs provides the cyber-physical system performance close to the one without attack. Simulation results of a real electrical power network with sensors under stealth attack show the effectiveness of the proposed approach.

Appendix

Proof of Theorem 1: The observation error dynamics are obtained as

$$\begin{aligned} \dot{e}_{x_1} &= G_{21}e_{x_1} + G_{22}e_y + G_{23}d \\ \dot{e}_y &= G_{11}e_{x_1} + G_{12}e_y + G_{13}d + D\dot{d} - \upsilon \end{aligned} \tag{A.1}$$

For the second equation of (A.1) consider a following Lyapunov function candidate

$$V = \frac{1}{2}e_y^T e_y = \frac{1}{2}\left\|e_y\right\|^2 \tag{A.2}$$

Denoting

$$\varphi = G_{11}e_{x_1} + G_{12}e_y + G_{13}d + D\dot{d} \tag{A.3}$$

and, taking into account the assumption (A4), the derivative of the Lyapunov function candidate (A.2) is estimated

$$\begin{aligned} &\dot{V} = e_y^T \dot{e}_y = e_y^T\left(G_{11}e_{x_1} + G_{12}e_y + G_{13}d + D\dot{d} - \upsilon\right) = \\ &e_y^T(\varphi - \upsilon) = e_y^T\left(\varphi - (\rho + L_3)\frac{e_y}{\left\|e_y\right\|}\right) = e_y^T\varphi - (\rho + L_3)\left\|e_y\right\| \le \\ &\left\|e_y\right\|\left(\left\|\varphi\right\| - (\rho + L_3)\right) \le -\rho\left\|e_y\right\| = -\rho\sqrt{2}V^{1/2} \end{aligned} \tag{A.4}$$

Therefore, $e_y \to 0$ in finite time at least locally. The estimation error dynamics (A.1) in the sliding mode $e_y = 0$ (that is achieved in finite time $t = t_r$ due to (A.4)) are obtained

$$\begin{aligned} &\dot{e}_{x_1} = G_{21}e_{x_1} + G_{23}d \\ &G_{11}e_{x_1} + G_{13}d + D\dot{d} = \upsilon_{eq} \end{aligned} \tag{A.5}$$

Transforming (A.5) by taking Laplace transform and solving for d , we obtain the estimate $\hat{d}$ given by (15) The theorem is proven.

Proof of Proposition 1: Consider the e_y dynamics from second equation of (A.1)

$$\dot{e}_y = G_{11}e_{x_1} + G_{12}e_y + G_{13}d + D\dot{d} - \upsilon \tag{A.6}$$

with bounded perturbation term

$$\left\| G_{11}e_{x_1} + G_{12}e_y + G_{13}d + D\dot{d} \right\| \le L_3 \tag{A.7}$$

at least locally with unknown L_3 . Firstly, we need to prove that the adaptive injection term υ in (22)-(27) drives $e_y \to 0$ in finite time. The proof of the finite time convergence $e_y \to 0$ by the adaptive injection term υ in (22)-(27) follows the one of the Proposition 2 in [27, pp. 185-186]. Convergence $e_y \to 0$ in finite time yields (A.5) and then (15). Therefore, $d(t)$ is reconstructed as in (15) with the adaptive injection term υ_{eq} or $\bar{\upsilon}_{eq}$. The proposition is proven.

Proof of Theorem 2: Taking into account (31), (34), and (36), the estimation error dynamics are derived as

$$\begin{aligned} \dot{e}_{x_1} &= Q_{11}e_{x_1} + Q_{12}e_{y_1} + Q_{13}e_{y_2} + Q_{14}d \\ \dot{e}_{y_1} &= Q_{21}e_{x_1} + Q_{22}e_{y_1} + Q_{23}e_{y_2} + Q_{24}d - \upsilon_1 \\ \dot{e}_{y_2} &= Q_{31}e_{x_1} + Q_{32}e_{y_1} + Q_{33}e_{y_2} + Q_{34}d + D_1\dot{d} - \upsilon_2 \end{aligned} \tag{A.8}$$

Introduce a Lyapunov function candidate

$$V_1 = \frac{1}{2}e_{y_1}^T e_{y_1} = \frac{1}{2}\left\| e_{y_1} \right\|^2 \tag{A.9}$$

that is applied to the second equation in (A.8) in order to prove the convergence $e_{y_1} \to 0$ in finite time.

The derivative of the Lyapunov function (A.9) can be computed as

$$\dot{V}_1 = e_{y_1}^T \dot{e}_{y_1} = e_{y_1}^T \left(Q_{21}e_{x_1} + Q_{22}e_{y_1} + Q_{23}e_{y_2} + Q_{24}d - \upsilon_1 \right) \tag{A.10}$$

Denoting

$$\varphi_1 = Q_{21}e_{x_1} + Q_{22}e_{y_1} + Q_{23}e_{y_2} + Q_{24}d \tag{A.11}$$

and assuming $\left\| \varphi_1 \right\| \le L_{11}$ at least locally, where $L_{11} > 0$ is known, we obtain

$$\begin{aligned}&\dot{V}_1 = e_{y_1}^T\left(\varphi_1 - \upsilon_1\right) = e_{y_1}^T\left(\varphi_1 - \left(\rho_1 + L_{11}\right)\frac{e_{y_1}}{\left\|e_{y_1}\right\|}\right) = e_{y_1}^T\varphi_1 - \left(\rho_1 + L_{11}\right)\left\|e_{y_1}\right\| \le \\ &\left\|e_{y_1}\right\|\left(\left\|\varphi_1\right\| - \left(\rho_1 + L_{11}\right)\right) \le -\rho_1\left\|e_{y_1}\right\| = -\rho\sqrt{2}V_1^{1/2}\end{aligned} \tag{A.12}$$

Therefore, $e_{y_1} \to 0$ in finite-time $\tilde{t}_{r_1} > 0$ at least locally.

Next, introduce a Lyapunov function candidate

$$V_2 = \frac{1}{2}e_{y_2}{}^T e_{y_2} = \frac{1}{2}\left\|e_{y_2}\right\|^2 \tag{A.13}$$

that is applied to the third equation in (A.8) in order to prove the convergence $e_{y_2} \to 0$ in finite time. Denoting

$$\varphi_2 = Q_{31}e_{x_1} + Q_{32}e_{y_1} + Q_{33}e_{y_2} + Q_{34}d + D_1\dot{d} \tag{A.14}$$

and assuming $\left\|\varphi_2\right\| \le L_{12}$ at least locally it is easy to show that $e_{y_2} \to 0$ in finite time $\tilde{t}_{r_2} > 0$ at least locally by means of the unit vector injection term υ_2 in (35). The proof is similar to the one that proves $e_{y_1} \to 0$ in finite-time by means of the unit-vector injection term υ_1 in (35).

Then, the estimation error dynamics (A.8) in the sliding mode $e_{y_1} = e_{y_2} = 0$ (that is achieved in finite time $\tilde{t}_r = \max\left(\tilde{t}_{r_1}, \tilde{t}_{r_2}\right) > 0$ at least locally) are reduced to

$$\begin{aligned}&\dot{e}_{x_1} = Q_{11}e_{x_1} + Q_{14}d \\ &0 = Q_{21}e_{x_1} + Q_{24}d - \upsilon_{1eq} \\ &0 = Q_{31}e_{x_1} + Q_{34}d + D_1\dot{d} - \upsilon_{2eq}\end{aligned} \tag{A.15}$$

where υ_{1eq} and υ_{2eq} are the equivalent injection signals.

Transforming (A.15) using Laplace, and excluding e_{x_1}, we obtain

$$\begin{aligned}&\left[Q_{21}\left(sI - Q_{11}\right)^{-1}Q_{14} + Q_{24}\right]d = \upsilon_{1eq} \\ &\left[Q_{31}\left(sI - Q_{11}\right)^{-1}Q_{14} + Q_{34} + D_1 s\right]d = \upsilon_{2eq}\end{aligned} \tag{A.16}$$

Finally, after algebraic transformations, the attack $d = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$, $d_1 \in \mathbb{R}^k$, $d_2 \in \mathbb{R}^{p-k}$ that satisfies (A.16) is estimated by $\hat{d} = \begin{bmatrix} \hat{d}_1 \\ \hat{d}_2 \end{bmatrix}$ as in (37). Theorem 2 is proven.

References

[1] P. Antsaklis, Goals and challenges in cyber-physical systems research, IEEE Transactions on Automatic Control 59 (12) (2014) 3117–3119.

[2] R. Baheti, H. Gill, Cyber-physical systems, The impact of control technology 12 (1) (2011) 161-166.

[3] E. A. Lee, Cyber Physical Systems: Design Challenges, in: 11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC), 2008, pp. 363-369.

[4] J. P. Conti, The day the samba stopped, Engineering & Technology 5(4) (2010) 46-47.

[5] S. Karnouskos, Stuxnet worm impact on industrial cyber-physical system security, in: 37th Annual Conference of the IEEE Industrial Electronics Society, 2011, pp. 4490–4494.

[6] J. Slay, M. Miller, Lessons Learned from the Maroochy Water Breach. In: Goetz E., Shenoi S. (eds) Critical Infrastructure Protection. IFIP International Federation for Information Processing 253, Springer, Boston, MA, 2007, pp. 73-82.

[7] F. Pasqualetti, F. Dörfler, and F. Bullo. Control-Theoretic Methods for Cyber-physical Security: Geometric Principle for Optimal Cross-Layer Resilient Control Systems, IEEE Control Systems Magazine 35(1) (2015) 110-127.

[8] A. Cardenas, S. Amin, and S. Sastry, Secure Control: Towards Survivable Cyber-Physical Systems, in: The 28th International Conference on Distributed Computing Systems Workshops, 2008, pp. 495–500.

[9] M. Zhu and S. Martinez, On the performance analysis of resilient networked control systems under replay attacks, IEEE Transactions on Automatic Control 59(3) (2014) 804–808.

[10] Y. Mo and B. Sinopoli, Secure control against replay attacks, in: Proc. Allerton Conf. Communications, Control Computing, 2010, pp. 911–918.

[11] R. Smith, A decoupled feedback structure for covertly appropriating network control systems, in: Proc. Int. Federation Automatic Control World Congress, 2011, pp. 90–95.

[12] Y. Mo and B. Sinopoli, False data injection attacks in control systems, Preprints of the 1st workshop on Secure Control Systems (2010) 1-6.

[13] V. D. Gligor, A note on denial-of-service in operating systems, IEEE Transactions on Software Engineering (1984) 320–324.

[14] G. Dan and H. Sandberg, Stealth attacks and protection schemes for state estimators in power systems, in: Proc. IEEE Int. Conf. Smart Grid Communications, 2010, pp. 214–219.

[15] F. Pasqualetti, F. Dorfler, and F. Bullo, Attack detection and identification in cyber-physical systems, IEEE Transactions on Automatic Control 58(11) (2013) 2715–2729.

[16] W. Wang, Z. Lu, Cyber security in the Smart Grid: Survey and challenges, Computer Networks 57 (2013) 1344-1371.

[17] Y. Yan, Y. Qian, H. Sharif and D. Tipper, A Survey on Cyber Security for Smart Grid Communications, IEEE Communications Surveys & Tutorials 14(4) (2012) 998-1010.

[18] X. Jin, W. M. Haddad, and T. Yucelen, An Adaptive Control Architecture for Mitigating Sensor and Actuator Attacks in Cyber- Physical Systems, IEEE Transactions on Automatic Control, 62(11) (2017) 6058–6064.

[19] S. Nateghi, Y. Shtessel, J-P Barbot, and C. Edwards, Cyber Attack Reconstruction of Nonlinear Systems via Higher-Order Sliding-Mode Observation and Sparse Recovery Algorithm, in: Conference on Decision and Control, 2018, pp. 5963-5968.

[20] S. Nateghi, Y. Shtessel, J-P Barbot, G. Zheng, and L. Yu, Cyber- Attack Reconstruction via Sliding Mode Differentiation and Sparse Recovery Algorithm: Electrical Power Networks Application, in 15th international workshop on Variable Structure System, 2018, pp. 285-290.

[21] M. L. Corradini and A. Cristofaro, Robust detection and reconstruction of state and sensor attacks for cyber-physical systems using sliding modes, IET Control Theory & Applications 11(11) (2017) 1756–1766.

[22] S. Mousavian, J. Valenzuela, and J. Wang, A probabilistic risk mitigation model for cyber-attacks to PMU networks, IEEE Trans. Power Syst. 30(1) (2015) 156–165.

[23] A. Taha, J. Qi, J. Wang, and J. Panchal, Risk mitigation for dynamic state estimation against cyber-attacks and unknown inputs, IEEE Transactions on Smart Grid 9(2) (2018) 886-899.

[24] E.J. Davision and S.H. Wang, On pole assignment in linear multivariable systems using output feedback, IEEE Trans. Autom. Control 20(4) (1975) 516-518.

[25] V. I. Utkin. Sliding Modes in Control Optimization. Springer-Verlag, Berlin,1992.

[26] Y. Shtessel, C. Edwards, L. Fridman, and A. Levant. Sliding Mode Control and Observation. Birkhauser, 2014.

[27] C. Edwards, Y. B. Shtessel, Adaptive continuous higher order sliding mode control, Automatica, 65 (2016) 183–190.

[28] P. W. Sauer and M. A. Pai, Power System Dynamics and Stability. Prentice Hall Inc., 1998.

[29] E. Scholtz, Observer-based monitors and distributed wave controllers for electromechanical disturbances in power systems, Ph.D. dissertation, Dept. Electr. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, 2004.