2019

Attack reconstruction and secure state estimation in cyber-physical systems using sliding mode observers

Shamila Nateghiboroujeni

ATTACK RECONSTRUCTION AND SECURE STATE ESTIMATION IN CYBER-PHYSICAL SYSTEMS USING SLIDING MODE OBSERVERS

by

SHAMILA NATEGHIBOROUJENI

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in
The Department of Electrical and Computer Engineering
to
The School of Graduate Studies
of
The University of Alabama in Huntsville

HUNTSVILLE, ALABAMA
2019

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree from The University of Alabama in Huntsville, I agree that the Library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by my advisor or, in his/her absence, by the Chair of the Department or the Dean of the School of Graduate Studies. It is also understood that due recognition shall be given to me and to The University of Alabama in Huntsville in any scholarly use which may be made of any material in this dissertation.

Shamila Nateghiboroujeni 11/22/2019
SHAMILA NATEGHIBOROUJENI (date)

DISSERTATION APPROVAL FORM

Submitted by SHAMILA NATEGHIBOROUJENI in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical Engineering and accepted on behalf of the Faculty of the School of Graduate Studies by the dissertation committee.

We, the undersigned members of the Graduate Faculty of The University of Alabama in Huntsville, certify that we have advised and/or supervised the candidate of the work described in this dissertation. We further certify that we have reviewed the dissertation manuscript and approve it in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical Engineering.

Yuri Shtessel 11/22/2019 Committee Chair
Dr. Yuri Shtessel (Date)

Shangbing Ai 12/4/2019
Dr. Shangbing Ai (Date)

Farbod Fa 12/4/2019
Dr. Farbod Fahimi (Date)

Laurie J Joiner 12/4/19
Dr. Laurie Joiner (Date)

Mark Tillman 11/22/19
Dr. Mark Tillman (Date)

12/4/19 Department Chair
Dr. Ravi Gorur (Date)

Mahalingam 12/06/19 College Dean
Dr. Shankar Mahalingam (Date)

12/9/19 Graduate Dean
Dr. David Berkowitz (Date)

ABSTRACT

School of Graduate Studies
The University of Alabama in Huntsville

Degree Doctor of Philosophy College/Dept. Engineering/Electrical and Computer Engineering

Name of Candidate SHAMILA NATEGHIBOROUJENI

Title Attack Reconstruction And Secure State Estimation in Cyber-Physical Systems

A Cyber-Physical System (CPS) represents a tight coupling of computational resources, network communication and physical processes. CPSs are composed of a set of networked components including sensors, actuators, control processing units, and communication agents that instrument the physical world to make it smarter. However, cyber components are also the source of new and unprecedented vulnerabilities to malicious attacks. Cyber security of CPS should provide three main security goals: *availability*, *confidentiality*, and *integrity*. 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. To protect a CPS from attacks, three security levels are considered: I) protection of the system from being attacked, II) detecting whether any attack happened, and III) resilient control of the system after being attacked. In this dissertation, we focused on attack reconstruction and secure state estimation in CPSs under sensor and state attacks in order to facilitate the resilient control of attacked CPSs. Numerous methods that study the resilient control of CPSs are presented in the literature. Applications of these approaches are limited to the special formats of CPSs, attacks mathematical models, and a variety of restrictive assumptions. To avoid these limitations, sliding mode differentiators and observers, as a robust observation approach, are used in this dissertation for online reconstruction of the sensor and state attacks as well as state estimation in nonlinear and linear CPSs under attacks. Next, the corrupted measurements and states are to be cleaned up on-line to stop the attack propagation into the CPS via the feedback control signal. A variety of attack scenarios are considered including (a) different combinations of the number of potential attacks and the number of sensors (b) linearized and nonlinear mathematical models of CPSs under attack. Corresponding observation algorithms were proposed and studied for on-line attack reconstruction and state estimation. The proposed observation algorithms and methodologies are applied to the US Western Electricity Coordinating Council power network, whose states and sensors are under attacks. Simulation results illustrate the efficacy of proposed observers.

Abstract Approval: Committee Chair Yuri Shtessel 11/22/2019
Dr. Yuri Shtessel

Department Chair 12/5/19
Dr. Ravi Gorur

Graduate Dean 12/9/19
Dr. David Berkowitz

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor Dr. Yuri B. Shtessel for his endless support during my Ph.D. program. His enthusiasm for his research is unbelievable and admirable which motivated me to work diligently every time after I met him.

I would like to thank my other advisory committee members: Dr. Shangbing Ai, Dr. Farbod Fahimi, Dr. Laurie L. Joiner, and Dr. Mark Tillman for their detailed review, constructive criticism and excellent advices during the preparation of the dissertation.

I appreciate the Electrical and Computer department chair, Dr. Ravi Gorur, for giving me the opportunity to pursue my Ph.D. program by offering me a Teaching Assistantship.

I would like to thank my beloved parents, Hojatallah and Niloofar, my dear siblings, Nima, Shima, Shakiba, and Mohammad Amin. I wish to have them beside myself during these years in my Ph.D. program, however, I have been receiving their endless support and love continuously which made this work possible.

I also appreciate my friends who provided motivation and support for me.

TABLE OF CONTENT

Page

IST OF FIGURES

gure Page

LIST OF TABLES

NOMENCLATURE

CPS	Cyber Physical System
DAE	Differential Algebraic Equation
DCS	Distributed Control System
HOSM	Higher Order Sliding Mode
PLC	Programmable Logic Controller
PMU	Phasor Measurement Unit
RIP	Restricted Isometry Property
SCADA	Supervisory Control and Data Acquisition
SMC	Sliding Mode Control
SMO	Sliding Mode Observer
STW	Super Twisting
SR	Sparse Recovery
UAV	Unmanned Arial Vehicles
WECC	Western Electricity Coordinating Council

CHAPTER 1

INTRODUCTION

1.1 State of Art and Literature Review

Cyber-Physical Systems (CPS) represent 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 [1-3], for example transportation networks, traffic control and safety, electric power networks, water networks, integrated biological systems, advanced automotive and industrial automation systems, and economic systems [4-8].

More specifically, using computer networks and related Internet technologies in industrial control systems to transfer information from the plant floor to supervisory computer systems has increased significantly in the last decade. For example, most industrial plants now use networked process servers to allow users to access real-time data from the Distributed Control Systems (DCS) and Programmable Logic Controllers (PLC) [9].

Talking specifically about another CPS, a platoon-based vehicular networked control system is an advanced automated method of driving a group of vehicles with some common interests on a road. In order to do that, each vehicle must be equipped with on-board sensors including radar, camera, lidar and also with a shared wireless communication network where inter-vehicle data in each vehicle are exchanged so that the platoon can reach its common interest. Since the control commands and sensor measurements are transmitted through wireless communication channels between vehicles, the content of these signals can be modified by cyber-attacks [10]. Identification and modeling process as [11, 12] which are based on data can be seriously affected by corrupted data.

Recent real-world cyber-attacks, including multiple power blackouts in Brazil [13], the StuxNet attack [14] in 2010, Maroochy attack to the water services in Queensland, Australia in 2000 [15], and a cyber-attack to Ukrainian power distribution networks [16] illustrate the importance of providing security to CPSs.

In 2000, the Maroochy water services in Australia, were attacked by an employee who attacked by infiltrating the Supervisory Control and Data Acquisition (SCADA) network of water services and modified the control signals. The result of his attack was the evacuation of one million liters of untreated sewage, over a three-month period, into storm water drains and on to local waterways [15].

A car hacking attack experiment on a Jeep which was driving in 70 mph on a highway in St. Louis, USA showed that cyber-attacks can cause very serious problem for modern automotive systems. Various electronic control units, from wiper to brake and engine systems, can be manipulated remotely by cyber attackers through the cellular connection inside the vehicle [17].

A number of studies have shown cyber-attack on the Unmanned Arial Vehicles (UAV) [18]. As an example, in 2011, US operators lost control of an RQ-170 UAV which was landed in Iran. The reason could be that Iranian forces jammed GPS communications followed by a spoof of GPS signals. As a result, the drone was landed in the Iranian's desired location [19].

Cyber security of CPS must provide three main security goals of *availability*, *confidentiality*, and *integrity* [20]. 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 or disruption, disclosure, and deception respectively (see Figure 1.1).

A deception attack happens when an authorized party receives false data and believes it to be true [21]. A specific kind of deception attack called a *Replay attack* is carried out by "hijacking" the sensors, recording the readings for a certain time, and repeating such readings while injecting an exogenous signal into the system's sensors. In *Replay attacks,* the system model is unknown to the attackers but they have access to the all sensors. In [22-25], it is shown that these attacks can be detected by injecting a random signal unknown to the attacker into the system. In the case when the system's dynamic model is known to the attacker, another kind of deception attack, *covert attack*, has been studied in [26, 27], 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 [28, 30]. 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 [31, 34].

A *Denial of service attack* assaults data availability through blocking information flows between different components of CPS. The attackers can jam the communication channels, modify devices and prevent them from sending data, violate routing protocols, etc. [35-37].

A *Disclosure attack* refers to any intrusions to the privacy of the agents of a CPS which include eavesdropping [38]. Most of the techniques which aim to provide a confidentiality service use randomization of data [39, 40].

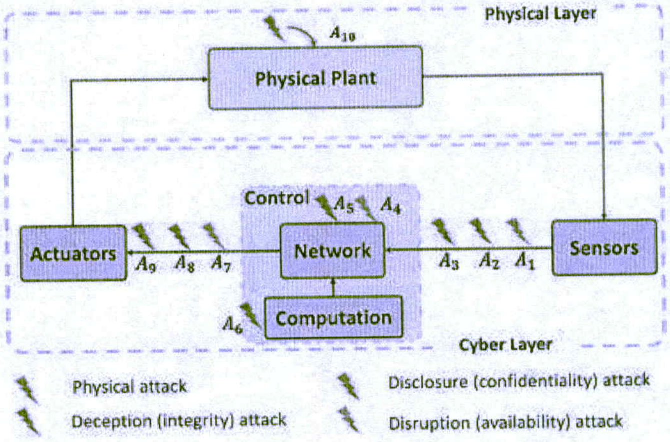


Figure 1.1 Attacks that May Happen to a Cyber-Physical System [41]

Cyber-physical system security including information security, protection of CPS from being attacked and detection in adversarial environments have been considered in the literature. A majority of the methods and tools for protecting CPSs from cyber-attacks are based on the development of special resilient software [42-55]. *Cryptography* and *Randomization* are two main approaches to protect a CPS against disclosure attacks: Cryptography is an approach to prevent third parties or the public from reading private messages by defining some protocols [56, 57]. Randomization is a defensive strategy to confuse the potential attacker about deterministic rules and information of the system [58].

However, how to ensure the CPS can continue functioning properly if a cyber-attack has happened is another serious problem that should be investigated. If the defense strategy just relies on detection, then system's performance still degrades and the threat of the same attack recurring is not diminished. In addition, in the interval between the onset of the attack and detection, the system could experience significant damage [41]. A good example of such a scenario is the Stuxnet [59]. The Maroochy attack happened because of the lack of detection and resilience mechanisms as well [15]. In RQ-170, the absence of resilience control caused the system to be unable to defend itself against the spoofing attack [19].

It is suggested in [20] that information security mechanisms have to be complemented by specially designed resilient control systems until the system is restored to normal operation. The focus of this dissertation is on reconstruction of the cyber-attack as a step to provide the resilient control for a CPS.

The control/observation algorithms are proposed in the literature for recovering CPS performance on-line if an attacker penetrates the information security mechanisms.

A *game-theoretic* approach that provides resilience consists of trying to minimize the damage that an attacker can apply to the system or maximize the price of attacking a system. For example, a zero-sum stochastic differential game between a defender and an attacker is used to find an optimal control design to provide system security in [60].

Event-triggered control schemes instead of time-triggered schemes, which are based on how frequent the attacks occur, are an appropriate strategy to increase the resilience of CPS [61]. Event-triggered control is especially used to mitigate the effect of a disruption attack [62]. *Mean Subsequence Reduced* as a resilient control approach ignores suspicious values and computes the control input at every moment [63, 64].

In *Trust-based approaches,* a function of trust value between the nodes of system is defined since some of nodes of system may be untrustworthy [65]. In [66], authors found the number of attacks that can be tolerated so that the state of the system can still be exactly recovered. They designed a secure local control loop to improve the resilience of the system. In [67], deception attacks are analyzed in stochastic systems, and the number of sensors to secure the system using a Kalman filter approach is proposed. In [68], new adaptive control architectures that can foil malicious sensors and actuator attacks are developed for linear CPS without reconstructing the attacks, by means of feedback control only.

The mentioned approaches suffer some disadvantages, limitations, and challenges, including:

I. It is assumed that the maximum number of malicious sensors in the network is known and bounded. Once the number of attacked sensors exceeds the upper bound, the proposed secure estimation or resilient control schemes fail to work.

II. Only specific types of malicious actions acting on the cyber layer are considered.

III. Only special structures of the cyber-physical system are considered.

On the other hand, the Sliding Mode Control (SMC) and Higher Order Sliding Mode (HOSM) control and observation techniques can handle systems of arbitrary relative degree perturbed by bounded perturbations/attacks of arbitrary shape. The Sliding Mode Observers (SMO) are capable to estimate the system states and reconstruct the bounded perturbations/attacks asymptotically or in finite time [69-73] while addressing the outlined challenges.

Detection and observation of a scalar attack by a SMO has been accomplished for a linearized differential-algebraic model of an electric power network when plant and sensor attacks do not occur simultaneously [74]. A SMO has been designed to simultaneously reconstruct states, attacks, and unknown input of a linear discrete-time state-space model when malicious attacks are sparse vector [75]. An adaptive SMO is designed coupled with a parameter estimator and a robust differentiator for detection and reconstruction of attacks in linear cyber-physical systems in [76] when state and sensor attacks do not happen simultaneously. Cyber-attacks against Phasor Measurement Unit (PMU) networks are considered in [77], where a risk-mitigation technique determines whether a certain PMU should be kept connected to network or removed. In [78] 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 the mentioned studies above which use a Sliding Mode approach for resilient control of CPSs, they all consider linear CPS and have their specific limitations.

1.2 Motivating Examples

In order to demonstrate the importance of reconstructing the cyber-attacks on cyber physical systems, motivating examples are presented in this section. At first, two tutorial examples of a CPS under state attack and a CPS under sensor attack are provided. Then, the model of US Western Electricity Coordinating Council (WECC) power system under stealth attack is investigated and it is shown how attacks degrade the performance of the power network.

1.2.1 Tutorial examples

Example 1.

Consider a CPS under bounded attack, whose dynamics are described by

$$\begin{aligned} \dot{x}(t) &= -2x(t) + v(t) + u_1(t) \\ y(t) &= x(t) \end{aligned} \tag{1.1}$$

where $y(t)$ is the measured output, $v(t)$ is the control signal, and $u_1(t)$ is the plant attack signal. The goal is to design the control law $v(t)$ that drives $x(t) \to 0$ as time increases in the presence of the bounded attack signal $u_1(t)$.

The problem can be addressed via

- feedback state controller design .
- reconstruction of the state attack $u_1(t)$.
- using the reconstructed state attack $\hat{u}_1(t)$ in the state feedback controller for compensating the attack.

The controller that is robust to the attack signal is designed as

$$v(t) = -3x(t) - \hat{u}_1(t) \tag{1.2}$$

where $\hat{u}_1(t)$ is an reconstructed attack signal base on the measurement $y(t)$. In order to get the $\hat{u}_1(t)$, find the derivation of output of CPS (1.1) as

$$\dot{y}(t) = \dot{x}(t) = -2x(t) + v(t) + u_1(t) \tag{1.3}$$

Substituting $x(t)$ with $y(t)$ in (1.3) gives the on-line reconstruction of attack as

$$\hat{u}_1(t) = \dot{y}(t) + 2y(t) - v(t) \tag{1.4}$$

The simulation results of applying the control signal (1.2) including the attack estimation (1.4) are presented in Figures 1.2 – 1.4. As it is clear in Figure 1.2, the output of CPS (1.1) under state attack is deviated from $x = 0$ while its output after applying control signal (1.2), shown in Figure 1.4, converges to $x = 0$ very well.

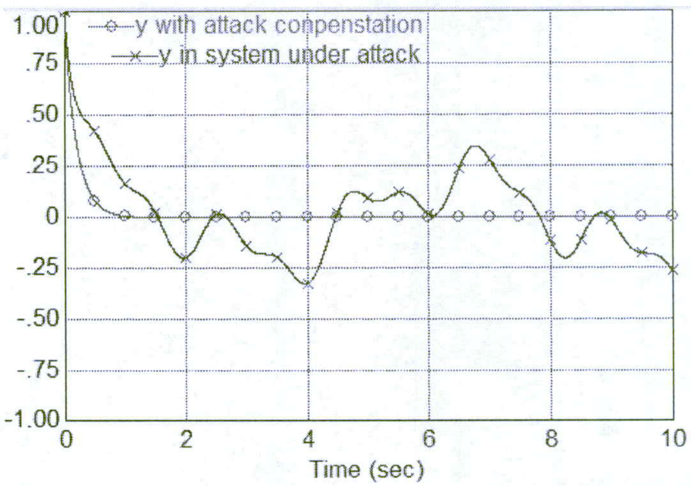


Figure 1.2 Output Dynamics with and without Attack Compensation

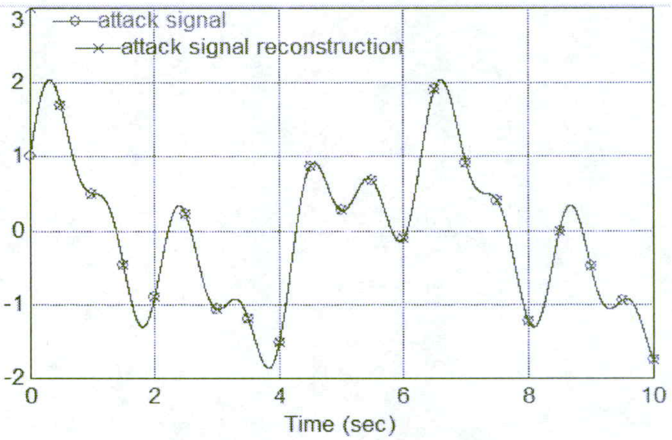


Figure 1.3 Attack Signal Reconstruction

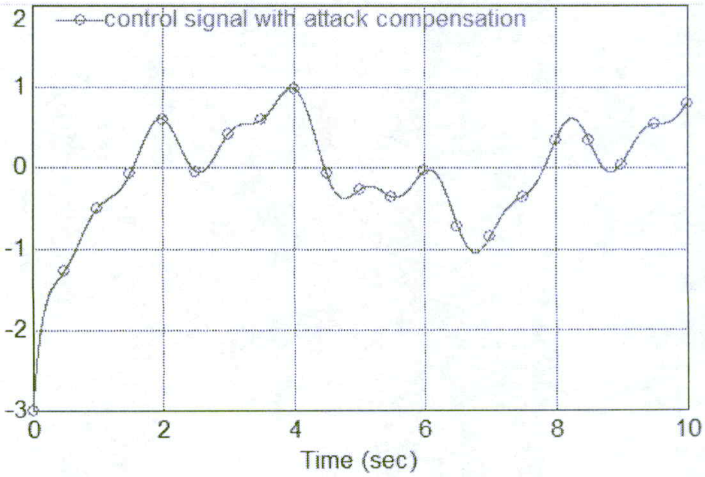


Figure 1.4 Feedback Control

Example 2

Consider the cyber physical control system whose sensor/measurement is corrupted by a bounded attack signal

$$\begin{aligned} \dot{x}(t) &= -2x(t) + v(t) \\ y(t) &= x(t) + u_2(t) \end{aligned} \tag{1.5}$$

where $y(t)$ is the measured output, $v(t)$ is control, and $u_2(t)$ is a sensor/measurement corruption attack signal.

The goal is to design the output tracking control law $v(t)$ that drives $x(t) \to x_c(t)$ as time increases in the presence of bounded measurement corruption attack signal $u_2(t)$.

The problem can be addressed via

- reconstruction of the sensor attack $u_2(t)$.
- cleaning up the sensor measurement using the reconstructed sensor attack $\hat{u}_2(t)$.
- feedback controller design using the cleaned measurement.

Find the derivation of output of CPS (1.5) to compute the on-line reconstruction of sensor attack $u_2(t)$ as follows

$$\dot{y}(t) = \dot{x}(t) + \dot{u}_2(t) = -2x(t) + v(t) + \dot{u}_2(t) \tag{1.6}$$

Replacing $x(t)$ with the second equation of eq. (1.5) gives

$$\dot{y}(t) = -2\left(y(t) - u_2(t)\right) + v(t) + \dot{u}_2(t) \tag{1.7}$$

Laplace transform of (1.7) is written as

$$Y(s)\left(s+2\right) = U_2(s)\left(s+2\right) + V(s) \tag{1.8}$$

where $Y(s)$, $U_2(s)$ and $V(s)$ are the Laplace transform of $y(t)$, $u_2(t)$ and $v(t)$ respectively. It is assumed that initial conditions $y(0)=0$, $u_2(0)=0$. Therefore, the estimation of cyber sensor attack $u_2(t)$ is obtained as

$$\hat{u}_2(t)=L^{-1}\left(Y(s)-\frac{1}{s+2}V(s)\right) \tag{1.9}$$

where $L^{-1}(.)$ shows the Inverse Laplace transform.

The “cleaned” measurement is computed as

$$y_{clean}(t)=y(t)-\hat{u}_2(t) \tag{1.10}$$

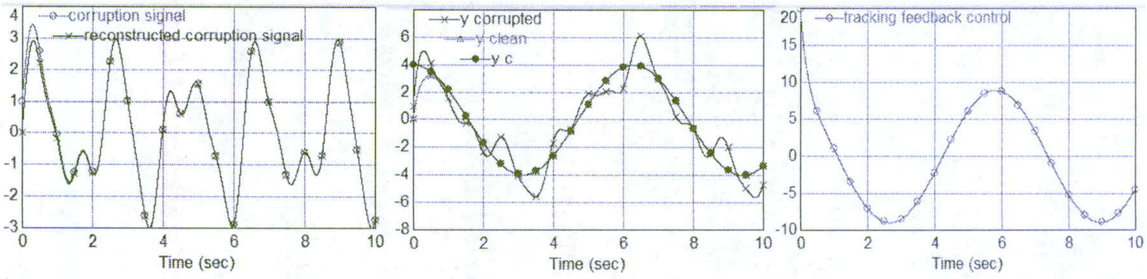
Then, it is used in the output tracking controller design

$$v(t)=\dot{x}_c(t)+2x_c(t)+3e(t) \tag{1.11}$$

where $e(t)=x_c(t)-y_{clean}(t)$, which converges to $e(t)=x_c(t)-x(t)$ as time increases. Replacing $\dot{x}(t)$, eq. (1.5), where $v(t)$ is equal to eq. (1.11) in $\dot{e}(t)=\dot{x}_c(t)-\dot{x}(t)$ gives

$$\dot{e}(t)=-e(t) \tag{1.12}$$

Equation (1.11) proves that the control signal eq. (1.11) makes $e\rightarrow 0$ by increasing time and provides the tracking goal for CPS eq. (1.5). The results of the simulation for CPS eq. (1.5) controlled by eq. (1.11) are presented in Figure 1.5.



(a) (b) (c)

Figure 1.5 (a). Attack Signal Reconstruction, (b). Output Tracking, (c). Tracking Feedback Control

Observations based on Examples 1 and 2:

- The attacks on the plant and on the sensors lead to significant degradation of the system's performance.
- The on-line reconstruction of attacks with a consecutive compensation by means of feedback control recovers the system's performance.

1.2.2 Electrical Power Network Example

In a real-world electrical power network only a small groups of generator rotor angles and rates are directly measured, and typical attacks aim at injecting disturbance signals that mainly affect the sensor-less generators [74]. The CPS that motivates the results presented in this section is the US WECC power system [83, 84] under attack with three generators and six buses, whose electrical schematic is presented in Figure 1.6.

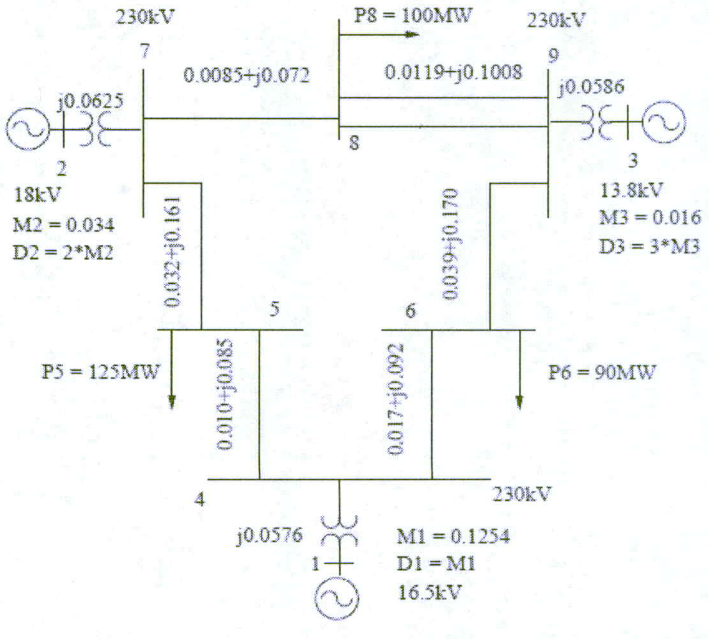


Figure 1.6 The Western Electricity Coordinating Council Power System [83]

The mathematical model of the power network in Figure 1.6 under sensor stealth attack can be represented as follows [83].

$$\begin{bmatrix} I & 0 & 0 \\ 0 & M_g & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} \dot{\delta} \\ \dot{\omega} \\ \dot{\theta} \end{bmatrix} = -\begin{bmatrix} 0 & -I & 0 \\ L^{\theta}_{g,g} & E_g & L^{\theta}_{g,l} \\ L^{\theta}_{l,g} & 0 & L^{\theta}_{l,l} \end{bmatrix}\underbrace{\begin{bmatrix} \delta \\ \omega \\ \theta \end{bmatrix}}_{x} + \underbrace{\begin{bmatrix} 0 \\ B_{\omega} \\ B_{\theta} \end{bmatrix}}_{B} d + \begin{bmatrix} 0 \\ P_{\omega} \\ P_{\theta} \end{bmatrix} \tag{1.13}$$

$$y = Cx + Dd$$

where the state vector $x = \begin{bmatrix} \delta^T & \omega^T & \theta^T \end{bmatrix}^T$ includes a vector of rotor angles $\delta \in \mathbb{R}^3$, vectors of the generator speed deviations from synchronicity $\omega \in \mathbb{R}^3$, as well as vector of voltage angles at the buses $\theta \in \mathbb{R}^6$.

The matrices $E_g, M_g \in \mathbb{R}^{3\times3}$ are diagonal matrices whose nonzero entries consist of the damping coefficients and the normalized inertias of the generators respectively and given by

$$M_g = \begin{bmatrix} 0.125 & 0 & 0 \\ 0 & 0.034 & 0 \\ 0 & 0 & 0.016 \end{bmatrix}, \; E_g = \begin{bmatrix} 0.125 & 0 & 0 \\ 0 & 0.068 & 0 \\ 0 & 0 & 0.048 \end{bmatrix} \tag{1.14}$$

The inputs P_{ω} and P_{θ} are due to *known* changes in the mechanical input power to the generators and real power demands at the loads. The $L^{\theta} \in \mathbb{R}^{9\times9}$ is an edge-weighted Laplacian of the graph associated with the lossless power network which is partitioned as

$$L^{\theta} = \begin{bmatrix} L^{\theta}_{g,g} & L^{\theta}_{g,l} \\ L^{\theta}_{l,g} & L^{\theta}_{l,l} \end{bmatrix} \tag{1.15}$$

where $L^{\theta}_{g,g} \in \mathbb{R}^{3\times3}$, $L^{\theta}_{g,l} \in \mathbb{R}^{3\times6}$, $L^{\theta}_{l,g} \in \mathbb{R}^{6\times3}$, $L^{\theta}_{l,l} \in \mathbb{R}^{6\times6}$ and L^{θ} is equal to

$$L^{\theta}=\begin{bmatrix} 0.058 & 0 & 0 & -0.058 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.063 & 0 & 0 & -0.063 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.059 & 0 & 0 & -0.059 & 0 & 0 & 0 \\ -0.058 & 0 & 0 & 0.265 & 0 & 0 & -0.085 & -0.092 & 0 \\ 0 & -0.063 & 0 & 0 & 0.296 & 0 & -0.161 & 0 & -0.072 \\ 0 & 0 & -0.059 & 0 & 0 & 0.330 & 0 & -0.170 & -0.101 \\ 0 & 0 & 0 & -0.085 & -0.161 & 0 & 0.246 & 0 & 0 \\ 0 & 0 & 0 & -0.092 & 0 & -0.170 & 0 & 0.262 & 0 \\ 0 & 0 & 0 & 0 & -0.072 & -0.101 & 0 & 0 & 0.173 \end{bmatrix} \quad (1.16)$$

The vector $y \in \mathbb{R}^6$, $y = \begin{bmatrix} \delta & \omega \end{bmatrix}^T$, is the measurement vector, the vector $d \in \mathbb{R}^3$ is the stealth attack vector corrupting the measurements and consequently affects the states of the plant through the feedback control, $B \in \mathbb{R}^{12\times3}$ and $D \in \mathbb{R}^{6\times3}$ are the attack distribution matrices and $C \in \mathbb{R}^{6\times12}$ is output gain matrix.

Note that generator rotor angles $\delta_i \quad i = 1,2,3$ are supposed to converge to the constant values, while the generator speed deviations from synchronicity $\omega_i \to 0 \quad i = 1,2,3$ in a case of nominal performance (without attack) of the studied network.

Consider the case when the sensors which measure the generator speed deviations from synchronicity, $\omega_1, \omega_2, \omega_3$, are corrupted by the following *stealth* attacks

$$d_1 = -1.1\omega_1 + 2\sin(t), \quad d_2 = -0.9\omega_2 + \cos(0.5t), \quad d_3 = -0.8\omega_3 + \sin(t) \qquad (1.17)$$

In order to show the effects of the stealth attacks eq. (1.17) on the performance of the electrical power system eq. (1.13), the system was simulated with and without attacks. The results of the simulations are shown in Figures 1.7 and 1.8. In Figure 1.7, corrupted sensor measurements ($\omega_1, \omega_2, \omega_3$ under attack) and sensor measurements when there is no attack are compared. In Figure 1.8, the states of system, vector of rotor angles $\delta_1, \delta_2, \delta_3$, when the power system is under attack and when there is no attack are compared.

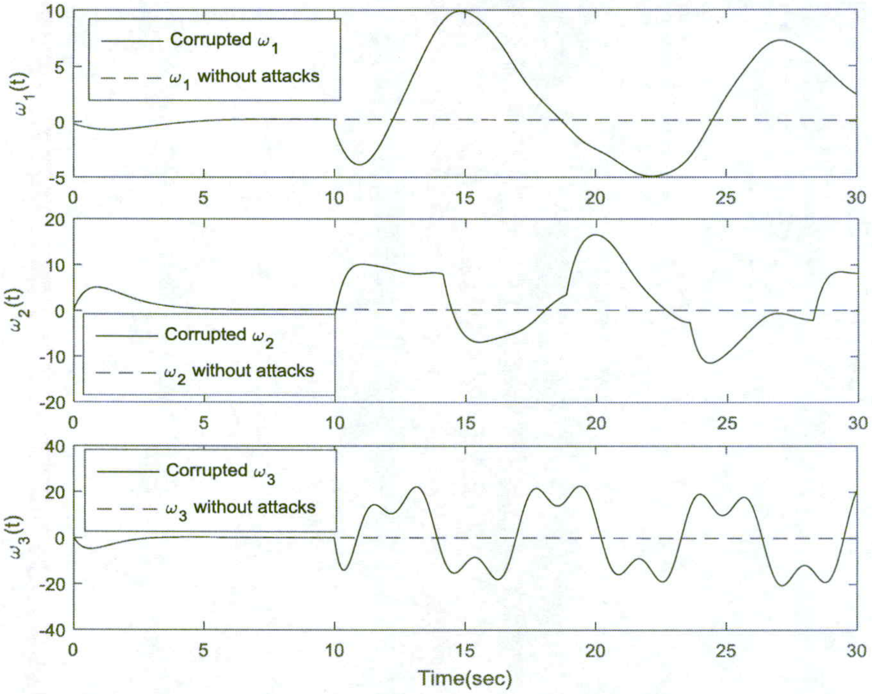


Figure 1.7 Corrupted Sensor Measurements $\omega_1, \omega_2, \omega_3$ Compared with Sensor Measurements $\omega_1, \omega_2, \omega_3$ when There is no Attack

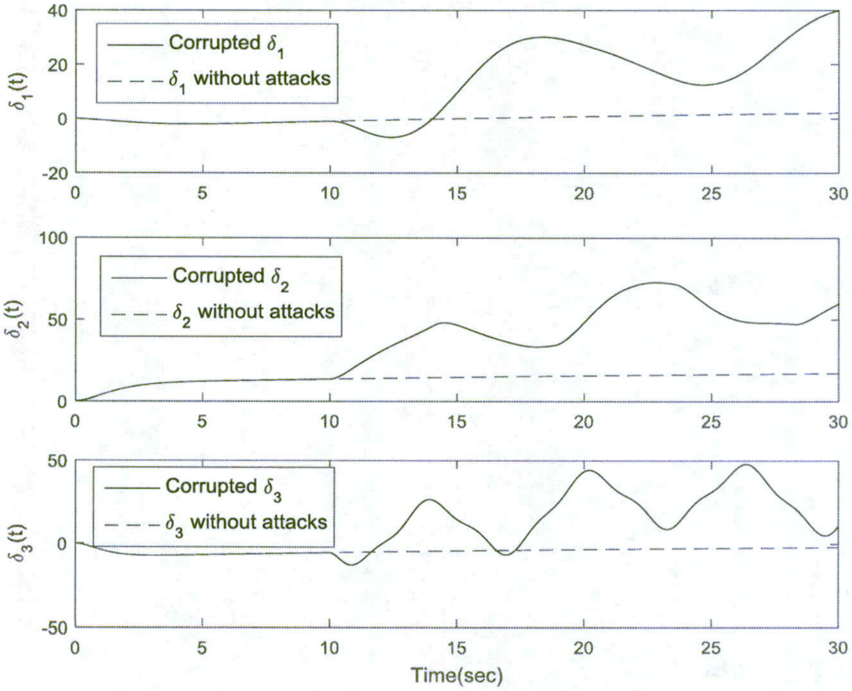


Figure 1.8 Corrupted States $\delta_1, \delta_2, \delta_3$ Compared with Stats $\delta_1, \delta_2, \delta_3$ when There is no Attack

Observations based on Electrical Power Network Example:

The stealth attacks lead to un-appropriate degradation of the power network performance. Estimation of the attacks and cleaning up the measurements before using them in feedback control is very important for retaining the performance of the electrical power system.

1.3 Research Contribution

Nonlinear and linearized CPS under sensor and state attacks are considered

I. A novel observation algorithm based on a sparse recovery (SR) technique along with a sliding mode differentiator is proposed for reconstructing on-line the sparse cyber-attacks on nonlinear CPS when there are more potential attacks than sensors. The novel result of this work is presented in [79].

II. A finite time convergent Higher Order Sliding Mode (HOSM) observer based on a HOSM differentiator is proposed to reconstruct on-line the plant attacks and estimate the states of a nonlinear CPS when the number of sensors is greater than the number of potential sensor attacks. The result of this work is published in [80].

III. Fixed-gain and adaptive-gain SMOs that include a newly proposed dynamic extension of the injection term is developed for the on-line reconstruction of attacks in a linearized CPS when the number of sensors is greater or equal to the number of potential attacks. Specifically, a novel adaptive sliding mode observation algorithm that reconstructs the smooth bounded CPS attacks with unknown boundaries on their amplitude and rates is proposed. Dynamic filters that address the attack propagation dynamics are proposed and employed for attack reconstruction for the first time. The results of this novel approach is presented in [81, 82].

1.4 Organization and Content

This dissertation consists of nine chapters. Chapter 1 presents literature review and research objectives as well as motivating examples. Chapter 2 describes the problem formulation. Chapter 3 provides background material that is used in this dissertation. Chapter 4 discusses the attack reconstruction and state estimation in a linearized CPS when the number of sensors is equal to the number of potential attacks. Chapter 5 studies attack reconstruction and state estimation in a linearized CPS when the number of sensors is greater than the number of potential attacks. Chapter 6 explores attack reconstruction in a nonlinear CPS when the number of potential attacks is greater than the number of sensors. Chapter 7, presents state estimation and attack reconstruction in a nonlinear CPS when the number of sensors is greater than the number of potential sensor attacks. In Chapter 8, proposed approaches in chapters 4 to 7 are tested in a real case study, the WECC power network system, and the simulation results are illustrated to show the effectiveness of developed approaches. Finally, chapter 9 declares conclusions and future works of this research.

1.5 Summary

A Cyber Physical System contains components that are accessible wirelessly through a network [1]. Many CPSs, including transportation networks, electric power networks, integrated biological systems, industrial automation systems, and economic systems [1-3], benefit from networked communication between different parts of a CPS. The downside of the CPS remote access is that an adversary has the capability to attack the system states and sensors remotely, and cause damage to the CPS and degrade CPS performance [3]. Protecting CPS from being attacked is of paramount importance. At the same time, CPSs should have the ability to continue working and recover their performance after being attacked.

Two tutorial examples of systems under cyber plant attack and sensor attack respectively, and model of US WECC power system under stealth attack are discussed in this chapter to show how cyber-attacks to plant and/or sensors can degrade the performance of a CPS and why the on-line reconstruction of the cyber-attacks with a consecutive cleaning up the measurements prior to using them in feedback control is of a prime importance for retaining the performance of CPS.

There exist a variety of studies to find resilience-increasing mechanism for CPSs. They mostly are based on Game theory, Event- triggered Control, Mean Subsequence Reduced algorithms, and Trust-based approaches. These approaches suffer disadvantages including having information of the maximum number of malicious sensors in the network, considering specific type of malicious action acting on the cyber layer, or some especial structure of the CPS.

Sliding mode observation techniques which can handle systems of arbitrary relative degree perturbed by bounded perturbations/attacks of arbitrary shape are proposed in this dissertation to reconstruct the attacks and secure an estimate of states of a CPS under sensor and state attacks asymptotically or in finite time. The reconstructed sensor attacks can be used for cleaning up the measurements so that the sensor attacks do not affect the CPS performance through the feedback control. Also the state attacks can be compensated by the CPS feedback control that employs the reconstructed state attacks.

In the next chapters 4-8, we will focus on the task of developing the algorithms of online reconstruction of the plant's and sensors' cyber-attacks as well as state estimation that may facilitate feedback control in order to recover the system's performance demonstrated prior to attacks.

CHAPTER 2

PROBLEM FORMULATION

2.1 Mathematical Modeling

Consider the following CPS which is completely observable and asymptotically stable affected by attack

$$\begin{aligned}\dot{x} &= f_1(x) + B_1(x)\big(u + d_u(t)\big) \\ y &= C(x) + D_1 d_y(t)\end{aligned} \tag{2.1}$$

where $x \in \mathbb{R}^n$ presents the state vector of CPS, $f_1(x) \in \mathbb{R}^n$ is a smooth vector-field, $y \in \mathbb{R}^p$ denotes the sensor measurement vector, and $u \in \mathbb{R}^{q_1}$ is the control signal. The $d_u(t) \in \mathbb{R}^{q_1}$ and $d_y(t) \in \mathbb{R}^{q_2}$ are the actuator and sensor attack respectively. The vector $C(x) \in \mathbb{R}^p$ is the output smooth vector field, $B_1(x) \in \mathbb{R}^{n \times q_1}$ and $D \in \mathbb{R}^{p \times q_2}$ denote the attack/fault distribution matrices.

Since it is very difficult to distinguish cyber-attacks from other perturbations acting on the CPSs, throughout this dissertation, cyber-attacks, faults, and disturbances are referred to as *attacks*.

The output feedback control signal u is a function of sensor measurement y which can be corrupted by the sensor attacks. This is

$$u(y) = \gamma\big(C(x) + d_y\big) = \gamma\big(x, d_y\big) \tag{2.2}$$

Replacing control signal u in CPS eq. (2.1) to find the closed loop CPS model gives

$$\begin{aligned} \dot{x} &= f_1(x) + B_1(x)\left(\gamma\left(x, d_y\right) + d_u(t)\right) = f_1(x) + B_1(x)\gamma\left(x, d_y\right) + B_1(x)d_u(t) \\ y &= C(x) + D_1 d_y(t) \end{aligned} \tag{2.3}$$

Assume that u can be written as

$$\gamma\left(x, d_y\right) = \gamma_1(x) + \gamma_2\left(d_y\right) \tag{2.4}$$

then, the closed loop CPS eq. (2.3) is given as

$$\begin{aligned} \dot{x} &= f_1(x) + B_1(x)\gamma_1(x) + B_1(x)\gamma_2\left(d_y\right) + B_1(x)d_u(t) \\ y &= C(x) + D_1 d_y(t) \end{aligned} \tag{2.5}$$

Therefore, the CPS eq. (2.1) after applying control signal u is presented as

$$\begin{aligned} \dot{x} &= f(x) + B_1(x)d_x(t) \\ y &= C(x) + D_1 d_y(t) \end{aligned} \tag{2.6}$$

where

$$\begin{aligned} f(x) &= f_1(x) + B_1(x)\gamma_1(x) \\ d_x(t) &= \gamma_2\left(d_y\right) + d_u(t) \end{aligned} \tag{2.7}$$

where $d_x(t)$ presents the plant/state attack.

Define the attack/fault signal $d(t) \in \mathbb{R}^q$ where $q = q_1 + q_2$ as

$$d = \begin{bmatrix} d_x \\ d_y \end{bmatrix} \tag{2.8}$$

where $d_x \in \mathbb{R}^{q_1}$ and $d_y \in \mathbb{R}^{q_2}$, and

$$B(x) = \left[B_1(x) \quad \mathbf{0}_1\right], D = \left[\mathbf{0}_2 \quad D_1\right] \tag{2.9}$$

where $B_1(x) \in \mathbb{R}^{n \times q_1}$, $D_1 \in \mathbb{R}^{p \times (q - q_1)}$, $\mathbf{0}_1 \in \mathbb{R}^{n \times (q - q_1)}$, $\mathbf{0}_2 \in \mathbb{R}^{p \times q_1}$. Then, the closed loop CPS eq. (2.6) is rewritten as

$$\begin{aligned} \dot{x} &= f(x) + B(x)d(t) \\ y &= C(x) + Dd(t) \end{aligned} \tag{2.10}$$

2.2 **Problem Statement**

The problem is two-fold

1. Develop an observation algorithm that reconstructs on-line the state $x \in \mathbb{R}^n$ and attack signal $d(t) \in \mathbb{R}^q$ in CPS eq. (2.10) so that

$$\hat{x}(t) \rightarrow x(t), \quad \hat{d}(t) \rightarrow d(t) \tag{2.11}$$

2. Develop an observation algorithm that reconstructs on-line the state $x \in \mathbb{R}^n$, the plant attack signal $d_x(t) \in \mathbb{R}^{q_1}$, and sensor attack signal $d_y(t) \in \mathbb{R}^{q_2}$ in CPS eq. (2.6) so that

$$\hat{x}(t) \rightarrow x(t), \quad \hat{d}_x(t) \rightarrow d_x(t), \quad \hat{d}_y(t) \rightarrow d_y(t) \tag{2.12}$$

as time increases.

Remark 2.1 The attack strategies are presented in Table 1.1 and discussed in section 1.1.

Table 1.1 Attack Strategies

Attack plan	$d_x(t) \neq 0$	$d_y(t) \neq 0$	Access to all sensors	Need to know the system model
Stealth attack		√		
Deception attack	√			
Reply attack	√	√	√	
Covert attack	√	√		√
False data injection attack		√		√

Remark 2.2 As soon as the sensor attack $d_y(t)$ is reconstructed the measurement $y = C(x) + D_1 d_y(t)$ could be cleaned as

$$y_{clean} = y - D_1 \hat{d}_y(t) = C(\hat{x}) + D_1\left(d_y(t) - \hat{d}_y(t)\right) \quad \rightarrow \quad y_{clean} = C(\hat{x}) \tag{2.13}$$

as time increases. Next, the clean measurement y_{clean} can be used in the feedback control of CPS. This allows blocking the propagation of the sensor attack to the dynamics of CPS through the feedback control.

CHAPTER 3

BACKGROUND

The algorithms which are used to reconstruct the attacks in this dissertation are reviewed below.

3.1 Sparse Recovering Algorithm

The problem of recovering an unknown input signal from measurements is well known, as a left invertibility problem, as seen in [85, 86] (in the nonlinear case see for example [87]), but this problem was only treated in the case where the number of measurements is equal or greater than the number of unknown inputs. The left invertibility problem in the case of fewer measurements than unknown inputs has no solution or more exactly has an infinity of solutions.

In this section, the problem is to find the exact recovery under sparse assumption denoted for the sake of simplicity as "Sparse Recovery", i.e. finding a concise representation of a signal s which is described as

$$\xi = \Phi s + \varepsilon \tag{3.1}$$

where $s \in \mathbb{R}^N$ are the unknown inputs with no more than j nonzero entries, $\xi \in \mathbb{R}^M$ are the measurements, ε is a measurement noise, and $\Phi \in \mathbb{R}^{M \times N}$ is a matrix where $M < N$.

Definition 3.1 The RIP (Restricted Isometry Property) condition of j -order with constant $\varsigma_j \in (0,1)$ (ς_j is as small as possible for computational reasons) of the matrix Φ yields

$$(1-\varsigma_s)\|s\|_2^2 \le \|\Phi s\|_2^2 \le (1+\varsigma_s)\|s\|_2^2 \tag{3.2}$$

for any j sparse of signal s .

Consider Φ_Γ as the index set of nonzero elements of Φ, then eq. (3.2) is equivalent to [88]

$$1-\varsigma_s \le eig\left(\Phi_\Gamma^T \Phi_\Gamma\right) \le 1+\varsigma_s \tag{3.3}$$

where Φ_Γ is the sub-matrix of Φ with active nodes.

The problem of SR is often cast as an optimization problem that minimizes a cost function constructed by leveraging the observation error term and the sparsity inducing term [88] i.e.

$$s^* = \arg \min_{s \in \mathbb{R}^N} \frac{1}{2}\|\xi - \Phi s\|_2^2 + \lambda \Theta(s) \tag{3.4}$$

where the sparsity term $\Theta(s)$ can be replaced by $\Theta(s) = \|s\|_1 \triangleq \sum_i |s_i|$ as long as the RIP conditions hold. The $\lambda > 0$ in eq. (3.4) is the balancing parameter and s^* is the *critical point*, i.e., the solution of eq. (3.1).

For sparse vectors s with j-sparsity, where j must be equal or smaller than $\frac{M-1}{2}$ [88], solution to the SR problem is unique and coincides with the critical point of eq. (3.1) when the RIP condition for Φ with order $2j$ is verified [88].

Under the sparse assumption of s and fulfilling j-RIP condition of matrix Φ, the estimate of s proposed in [88] is

$$\begin{aligned} \mu\dot{v}(t) &= -\left\lceil v(t)+\left(\Phi^T\Phi - I_{N\times N}\right)a(t) - \Phi^T\gamma \right\rfloor^{\beta} \\ \hat{s}(t) &= a(t) \end{aligned} \tag{3.5}$$

where $v \in \mathbb{R}^N$ is the state vector, $\hat{s}(t)$ represents the estimate of the sparse signal s of eq. (3.1), and $\mu > 0$ is a time-constant determined by the physical properties of the implementing system. Note that $\lceil . \rfloor^{\beta} = |.|^{\beta}\, sign(.)$ and $a(t) = H_{\lambda}(v)$ where $H_{\lambda}(.)$ is a continuous soft thresholding function and defined as

$$H_{\lambda}(v) = \max\left(|v| - \lambda, 0\right)\mathrm{sgn}(v) \tag{3.6}$$

where $\lambda > 0$ is chosen with respect to the noise and the minimum absolute value of the nonzero terms.

Under Definition 3.1, the state v of eq. (3.5) converges in finite time to its equilibrium point v^*, and $\hat{s}(t)$ in eq. (3.5) converges in finite-time to $s*$ of eq. (3.4).

3.2 Line-by-Line Super-Twisting Sliding Mode Observer for Linear Systems

Consider the following linear system

$$\begin{aligned} \dot{x} &= Ax + Bd(t) \\ y &= \begin{bmatrix} y_1 & y_2 & ,..., & y_q \end{bmatrix}^T = Cx\,, \quad y_i = C_i x \end{aligned} \tag{3.7}$$

where $x \in \mathbb{R}^n$ presents the system states, $y \in \mathbb{R}^p$ is the output of system, and $d(t) \in \mathbb{R}^q$ denotes the unknown input to the system, while $p = q$. The $y_i \in \mathbb{R}$ and C_i is the i^{th} row of matrix C for $i = 1,...,q$.

Assumption (A 3.1): The system in eq. (3.7) is assumed to have an input-output vector relative degree $r=\left\{r_1,r_2,...,r_q\right\}$, i.e.

$$\begin{aligned} &C_j A^k B=0 \quad for\ all \quad k<r_j-1 \\ &C_j A^{r_j-1}B\neq 0, \quad j=1,2,.q \end{aligned} \tag{3.8}$$

Without loss of generality, it is assumed that $r_1\leq ...\leq r_q$, where the integers $1\leq r_{\alpha_i}\leq r_i$ are such that $rank\left(C_a B\right)=rank\left(B\right)$. Furthermore, the r_{α_i}, $i=1,2,...,q$ are chosen such that $r_s=\sum_{i=1}^{q} r_{\alpha_i}$ is minimal.

The problem is the state estimation and unknown inputs reconstruction in the linear system eq. (3.7) subject to unknown inputs $d(t)$.

According to (A 1.3), expression $y_i=C_i x\in\mathbb{R}$ for $i=1,...,q$ are as follows

$$\begin{aligned} &\dot{y}_i=y_i^1=C_i A x \\ &\dot{y}_i^1=y_i^2=C_i A^2 x \\ &\vdots \\ &\dot{y}_i^{r_i-1}=y_i^{r_i}=C_i A^{r_i-1}x \\ &\dot{y}_i^{r_i}=C_i A^{r_i}x+C_i A^{r_i-1}Bd(t) \end{aligned} \tag{3.9}$$

Consider the following observer [89] to estimate y_i^j for $j=1,...,i-1$ as follows

$$\begin{aligned} &\dot{\hat{y}}_i=v\left(y_i-\hat{y}_i\right)=v\left(s_i^1\right) \\ &\dot{\hat{y}}_i^1=v\left(\tilde{y}_i^1-\hat{y}_i^1\right)=v\left(s_i^2\right) \\ &\vdots \\ &\dot{\hat{y}}_i^{r_i-1}=v\left(\tilde{y}_i^{r_i-1}-\hat{y}_i^{r_i-1}\right)=v\left(s_i^{r_i}\right) \\ &\dot{\hat{y}}_i^{r_i}=v\left(\tilde{y}_i^{r_i}-\hat{y}_i^{r_i}\right)=v\left(s_i^{r_i+1}\right) \end{aligned} \tag{3.10}$$

where

$$\begin{aligned} &\tilde{y}_{1i}^1=y_i \\ &\tilde{y}_i^j=v\left(\tilde{y}_i^{j-1}-y_i^{j-1}\right), \quad 2\leq j\leq r_{\alpha_i}-1 \end{aligned} \tag{3.11}$$

Denoting $e_{y_i} = y_i - \hat{y}_i$, the error dynamics are given by

$$\begin{aligned}
\dot{e}_{y_i} &= \dot{y}_i - \dot{\hat{y}}_i \\
&= C_i A x - v\left(y_i - \hat{y}_i\right) = y_i^{1} - v\left(y_i - \hat{y}_i\right) \\
\dot{e}_{y_i^1} &= \dot{y}_i^{1} - \dot{\hat{y}}_i^{1} \\
&= C_i A^2 x - v\left(\tilde{y}_i^{1} - \hat{y}_i^{1}\right) = y_i^{2} - v\left(\tilde{y}_i^{1} - \hat{y}_i^{1}\right) \\
&\vdots \\
\dot{e}_{y_i^{r_i - 1}} &= \dot{y}_i^{r_i - 1} - \dot{\hat{y}}_i^{r_i - 1} \\
&= C_i A^{r_i - 1} x - v\left(\tilde{y}_i^{r_i - 1} - \hat{y}_i^{r_i - 1}\right) = y_i^{r_i} - v\left(\tilde{y}_i^{r_i - 1} - \hat{y}_i^{r_i - 1}\right) \\
\dot{e}_{y_i^{r_i}} &= \dot{y}_i^{r_i} - \dot{\hat{y}}_i^{r_i} = C_i A^{r_i} x + C_i A^{r_i - 1} B d(t) - v\left(\tilde{y}_i^{r_i} - \hat{y}_i^{r_i}\right)
\end{aligned} \tag{3.12}$$

where in each case the continuous injection term $v(.)$ is given by the Super Twisting (STW) algorithm [90]

$$\begin{aligned}
&v(s_i^j) = \varphi(s_i^j) + \lambda_i^j \left|s_i^j\right|^{1/2} sign(s_i^j) \\
&\dot{\varphi}(s_i^j) = \beta_i^j sign(s_i^j), \quad \lambda_i^j, \beta_i^j > 0
\end{aligned} \tag{3.13}$$

where $\lambda_i^j \in \mathbb{R}$ and $\beta_i^j \in \mathbb{R}$ are suitably chosen gains and the $s_i^j \in \mathbb{R}$ for $i = 1,\ldots,q$ and $j = 1,\ldots,r_{i+1}$ are the sliding variables where

$$\begin{aligned}
&s_i^1 = y_i - \hat{y}_i \\
&s_i^j = \tilde{y}_i^{j-1} - \hat{y}_i^{j-1}, \quad for \quad j = 2,\ldots,r_{i+1}
\end{aligned} \tag{3.14}$$

It is assumed that $\left|y_i^{j+1}\right| \le L_i^j$ for $j = 1,\ldots,r_{i-1}$, and $\left|C_i A^{r_i} x + C_i A^{r_i - 1} B d(t)\right| \le L_i^{r_i}$ where L_i^j's are fixed and known.

It is shown [89] that with λ_i^j and β_i^j chosen as [91]

$$\lambda_i^j = 1.5\sqrt{L_i^j}, \quad \beta_i^j = 1.1 L_i^j \tag{3.15}$$

for $i = 1,\ldots,q$ and $j = 1,\ldots,r_i$, a second order sliding mode emerges in finite time on

$e_{y_i} = y_i - \hat{y}_i = 0$.

As a result, $y_i^{1} \to v\left(y_i - \hat{y}_i\right)$ in finite time, therefore, the estimation of y_i^{1} which is shown as $\tilde{y}_i^{1}$, is obtained as

$$\tilde{y}_i^{1} = v\left(y_i - \hat{y}_i\right) = C_i A x \tag{3.16}$$

By replacing eq. (3.16) in eq. (3.12), y_i^{2} is given in the same way. Continue the *Line-by-Line* observer eqs. (3.10)-(3.15), then it is given for $1 \le i \le q$ that

$$\begin{aligned} &\tilde{y}_i^{1} = y_i \\ &\tilde{y}_i^{j} = v\left(\tilde{y}_i^{j-1} - y_i^{j-1}\right) = C_i A^j x, \quad 2 \le j \le r_{\alpha_i} - 1 \end{aligned} \tag{3.17}$$

Remark 3.1 The values $L_i^j > 0$ are difficult to predict.

Overestimating $L_i^j > 0$ may lead to the gains λ_i^j and β_i^j being overestimated, and, therefore, to increase chattering. The adaptive version of the unknown input estimation algorithm, eqs. (3.10) – (3.15), with non-overestimated gains is discussed in Chapter 5.

Consider the SMO of the form

$$\dot{\hat{x}} = A\hat{x} + G_l\left(y_a - C_a\hat{x}\right) + G_n \upsilon_c (y_a - C_a\hat{x}) \tag{3.18}$$

where the matrices $G_l \in \mathbb{R}^{n \times r_s}$ and $G_n \in \mathbb{R}^{n \times r_s}$ are of appropriate dimension and are to be designed. The auxiliary output y_a which contains both real and synthetic measurements and the matrix C_a are defined as follows

$$y_a = \begin{bmatrix} y_1 \\ \nu\left(y_1 - y_1^1\right) \\ \vdots \\ \nu\left(\tilde{y}_1^{r_{\alpha_1}-1} - y_1^{r_{\alpha_1}-1}\right) \\ \vdots \\ \bar{y}_q \\ \vdots \\ \nu\left(\tilde{y}_q^{r_{\alpha_1}-1} - y_q^{r_{\alpha_1}-1}\right) \end{bmatrix}, \quad C_a = \begin{bmatrix} C \\ \vdots \\ CA^{r_{\alpha 1}-1} \\ \vdots \\ C_q \\ \vdots \\ C_q\bar{A}^{r_{\alpha q}-1} \end{bmatrix} \tag{3.19}$$

and $\upsilon_c(.)$ is the injection vector

$$\upsilon_c(y_a - C_a\hat{x}) = \begin{cases} -(\kappa + \eta_0)\dfrac{P(y_a - C_a\hat{x})}{\left\| P(y_a - C_a\hat{x}) \right\|} & if\ (y_a - C_a\hat{x}) \neq 0 \\ 0 & otherwise \end{cases} \tag{3.20}$$

where η_0 is a small positive constant and κ is a positive constant suitably larger than the upper bound of the unknown input d [89]. The positive definite matrix P can be found by solving a corresponding Lyapunov equation [92].

The presented results can be summarized as

Proposition 3.1 [89]. The states x are estimated asymptotically in eq. (3.18) using the SMO in eqs. (3.10) – (3.15) and the STW injection terms in eqs. (3.13) – (3.15), while the unknown input d in the CPS eq. (3.7) is estimated asymptotically as

$$\hat{d} = \left((C_aB)^T C_aB\right)^{-1} (C_aB)^T C_aG_n(\upsilon_c)_{eq} \tag{3.21}$$

3.3 Higher Order Sliding Mode Observer for Nonlinear Systems

Consider the following locally stable system

$$\begin{aligned} \dot{x} &= f(x) + B(x)d(t) \\ y &= C(x) \end{aligned} \tag{3.22}$$

where $x \in \mathbb{R}^n$ denotes the system states, $y \in \mathbb{R}^p$ is the output vector, $d(t) \in \mathbb{R}^q$ represents the unknown input to the system, where $p = q$.

The vector y is $y = \begin{bmatrix} y_1 & y_2 & ,..., & y_q \end{bmatrix}^T$ and matrix B is $B = \begin{bmatrix} b_1, b_2, ..., b_q \end{bmatrix} \in \mathbb{R}^{n \times q}$ where $b_i \in \mathbb{R}^n, \forall i = 1,...,m$ are smooth vector-fields defined on an open $\Omega \subset \mathbb{R}^n$.

The problem is: considering a nonlinear system with unknown input in eq. (3. 22), find an observer to reconstruct the unknown input vector and estimate the states of system.

System Transformation: The following properties introduced by Isidori in [93] are assumed at a neighborhood of any point $x \in \Omega$:

Assumption (A 3.2): The system in (3.22) is assumed to have vector relative degree $r = \{r_1, r_2, ..., r_q\}$, i.e.

$$\begin{aligned} &L_{bj} L_f^k y_i(x) = 0 \quad \forall j = 1,...,q, \quad \forall k < r_i - 1, \quad \forall i = 1,...,q \\ &L_{bj} L_f^{r_i - 1} y_i(x) \neq 0 \text{ for at least one } 1 \le j \le q \end{aligned} \tag{3.23}$$

Assumption (A 3.3): The matrix

$$L(x) = \begin{bmatrix} L_{b_1}\left(L_f^{r_1-1} y_1\right) & L_{b_2}\left(L_f^{r_1-1} y_1\right) & \cdots & L_{b_q}\left(L_f^{r_1-1} y_1\right) \\ L_{b_1}\left(L_f^{r_2-1} y_2\right) & L_{b_2}\left(L_f^{r_2-1} y_2\right) & \cdots & L_{b_q}\left(L_f^{r_2-1} y_2\right) \\ \vdots & \vdots & \vdots & \vdots \\ L_{b_1}\left(L_f^{r_q-1} y_q\right) & L_{b_2}\left(L_f^{r_q-1} y_q\right) & \cdots & L_{b_q}\left(L_f^{r_q-1} y_q\right) \end{bmatrix} \tag{3.24}$$

is full rank.

Assumption (A 3.4): The distribution $\Gamma = span\{b_1, b_2, ..., b_q\}$ is involutive [93, 94].

which means that no new direction is generated by the Lie bracket of the distribution vector fields. This ensures that the zero dynamics (when they exists) can be rewritten independently of the unknown input.

The system given by eq. (3.22) with the involutive distribution $\Gamma = span\left\{b_1, b_2, ..., b_q\right\}$ and total relative degree $r = \sum_{i=1}^{q} r_i \leq n$ can be rewritten as

$$\dot{\delta}_i = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{r_i \times r_i} \delta_i + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ L_f^{\,r_i} y_i(x) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \sum_{j=1}^{m} L_{b_j} L_f^{\,r_i - 1} y_i(x) d(t) \end{bmatrix}, \forall i = 1, ..., q \tag{3.25}$$

$$\dot{\gamma} = g(\delta, \gamma)$$

where

$$\delta = \begin{bmatrix} \delta_1 & \delta_2 & \cdots & \delta_q \end{bmatrix}^T, \qquad \delta_i = \begin{bmatrix} \delta_{i1} \\ \delta_{i2} \\ \vdots \\ \delta_{ir_i} \end{bmatrix} = \begin{bmatrix} \eta_{i1}(x) \\ \eta_{i2}(x) \\ \vdots \\ \eta_{ir_i}(x) \end{bmatrix} = \begin{bmatrix} y_i(x) \\ L_f y_i(x) \\ \vdots \\ L_f^{\,r_i - 1} y_i(x) \end{bmatrix} \in \mathbb{R}^{r_i} \qquad \forall i = 1, ..., q$$

$$\gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_{n-r} \end{bmatrix} = \begin{bmatrix} \eta_{r+1}(x) \\ \eta_{r+2}(x) \\ \vdots \\ \eta_n(x) \end{bmatrix} \tag{3.26}$$

Assumption (A 3.5): The norm-bounded solution of the internal dynamics $\dot{\gamma} = g(\delta, \gamma)$ is assumed to be locally asymptotically stable [95].

The variables $\eta_{r+1}(x), ..., \eta_n(x)$ are defined to satisfy

$$L_{b_j} \eta_i(x) = 0 \;\; \forall i = r+1, ..., n, \; \forall j = 1, ..., q \tag{3.27}$$

if assumption (A 3.4) is satisfied then it is always possible to find $n - r$ functions $\eta_{r+1}(x), ..., \eta_n(x)$ such that

$$\Psi(x) = col\left\{\eta_{11}(x),\ldots,\eta_{1r_1}(x),\ldots,\eta_{q1}(x),\ldots,\eta_{qr_q}(x),\eta_{r+1}(x),\ldots,\eta_n(x)\right\} \in \mathbb{R}^n \tag{3.28}$$

is a local diffeomorphism in a neighborhood of any point $x \in \bar{\Omega} \subset \Omega \subset \mathbb{R}^n$ which means

$$x = \Psi^{-1}\left(\delta,\gamma\right) \tag{3.29}$$

In order to estimate the derivatives $\delta_{ij}(t)$ $\forall i = 1,\ldots,q$, $\forall j = 1,\ldots,r_i$ of the outputs y_i in finite time, higher-order sliding-mode differentiators [96] are used.

$$\begin{aligned}
&\dot{z}_0^i = v_0^i, \quad v_0^i = -\lambda_0^i \left|z_0^i - y_i(t)\right|^{(r_i/(r_i+1))} sign\left(z_0^i - y_i(t)\right) + z_1^i, \\
&\dot{z}_1^i = v_1^i, \quad v_1^i = -\lambda_1^i \left|z_1^i - v_0^i\right|^{((r_i-1)/r_i)} sign\left(z_1^i - v_0^i\right) + z_2^i, \\
&\vdots \\
&\dot{z}_{r_i-1}^i = v_{r_i-1}^i, \quad v_{r_i-1}^i = -\lambda_{r_i-1}^i \left|z_{r_i-1}^i - v_{r_i-2}^i\right|^{(1/2)} sign\left(z_{r_i-1}^i - v_{r_i-2}^i\right) + z_{r_i}^i, \\
&\dot{z}_{r_i}^i = -\lambda_{r_i}^i sign\left(z_{r_i}^i - v_{r_i-1}^i\right)
\end{aligned} \tag{3.30}$$

for $i = 1,\ldots,q$.

By construction,

$$\begin{aligned}
&\hat{\delta}_1^1 = \hat{\eta}_1^1(x) = z_0^1 \ ,\ldots, \ \hat{\delta}_{r_1}^1 = \hat{\eta}_{r_1}^1(x) = z_{r_1-1}^1, \ \dot{\hat{\delta}}_{r_1}^1 = \dot{\hat{\eta}}_{r_1}^1(x) = z_{r_1}^1 \\
&\vdots \\
&\hat{\delta}_1^q = \hat{\eta}_1^q(x) = z_0^q \ ,\ldots, \ \hat{\delta}_{r_q}^q = \hat{\eta}_{r_q}^q(x) = z_{r_q-1}^q, \ \dot{\hat{\delta}}_{r_1}^q = \dot{\hat{\eta}}_{r_q}^q(x) = z_{r_q}^1
\end{aligned} \tag{3.31}$$

Therefore, the following exact estimates are available in finite time:

$$\hat{\delta}_i = \begin{bmatrix} \hat{\delta}_{i1} \\ \hat{\delta}_{i2} \\ \vdots \\ \hat{\delta}_{ir_1} \end{bmatrix} = \begin{bmatrix} \hat{\eta}_{i1}(\hat{x}) \\ \hat{\eta}_{i2}(\hat{x}) \\ \vdots \\ \hat{\eta}_{ir_1}(\hat{x}) \end{bmatrix} \in \mathbb{R}^{r_i} \quad \forall i = 1,\ldots,q \qquad \hat{\delta} = \begin{bmatrix} \hat{\delta}^1 \\ \hat{\delta}^2 \\ \vdots \\ \hat{\delta}^q \end{bmatrix} \in \mathbb{R}^{r_i} \tag{3.32}$$

Next, integrating eq. (3.32) and replacing δ by $\hat{\delta}$, the internal dynamics is

$$\dot{\hat{\gamma}} = g\left(\hat{\delta},\hat{\gamma}\right) \tag{3.33}$$

and with some initial condition from the stability domain of the internal dynamics, a asymptotic estimate $\hat{\gamma}$ can be obtained locally

$$\hat{\gamma} = \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \vdots \\ \hat{\gamma}_{n-r} \end{pmatrix} = \begin{pmatrix} \hat{\eta}_{r+1}(\hat{x}) \\ \hat{\eta}_{r+2}(\hat{x}) \\ \vdots \\ \hat{\eta}_n(\hat{x}) \end{pmatrix} \tag{3.34}$$

Therefore, the asymptotic estimate for the mapping eq. (3.29) is identified as

$$\Psi(\hat{x}) = col\left\{\hat{\eta}_{11}(\hat{x}),...,\hat{\eta}_{1r_1}(\hat{x}),...,\hat{\eta}_{q1}(\hat{x}),...,\hat{\eta}_{qr_q}(\hat{x}),\hat{\eta}_{r+1}(\hat{x}),...,\hat{\eta}_n(\hat{x})\right\} \tag{3.35}$$

The asymptotic estimate $\hat{x}$ of the state vector x can be easily identified via eqs. (3.29) and (3.35) as

$$\hat{x} = \Psi^{-1}\left(\hat{\delta},\hat{\gamma}\right) \tag{3.36}$$

Since the finite-time exact estimates $\hat{\dot{\delta}}_{ir_i}$ of $\dot{\delta}_{ir_i}$, $\forall i = 1,...,q$ are available via the higher-order sliding-mode differentiator, and using the estimates $\hat{\delta}$, $\hat{\gamma}$ for δ, γ, an asymptotic estimate $\hat{d}(t)$ of the unknown input $d(t)$ in eq. (3.22) can be identified as

$$\hat{d}(t) = L^{-1}\left(\Psi^{-1}\left(\hat{\delta},\hat{\gamma}\right)\right)\left[\begin{pmatrix} \hat{\dot{\delta}}_{1r_1} \\ \hat{\dot{\delta}}_{2r_2} \\ \vdots \\ \hat{\dot{\delta}}_{qr_q} \end{pmatrix} - \begin{pmatrix} L_f^{r_1} y_1\left(\Psi^{-1}\left(\hat{\delta},\hat{\gamma}\right)\right) \\ L_f^{r_2} y_2\left(\Psi^{-1}\left(\hat{\delta},\hat{\gamma}\right)\right) \\ \vdots \\ L_f^{r_q} y_q\left(\Psi^{-1}\left(\hat{\delta},\hat{\gamma}\right)\right) \end{pmatrix}\right] \tag{3.37}$$

where $L\left(\Psi^{-1}\left(\hat{\delta},\hat{\gamma}\right)\right) = \sum_{j=1}^{q} L_{b_j} L_f^{\,r_i-1} y_{1i}\left(x\right)$.

Remark 3.2: The convergence $\hat{d} \rightarrow d$ can be achieved only locally and as time increases due to the local asymptotic stability of the norm-bounded solution of the internal dynamics $\dot{\gamma} = g\left(\delta,\gamma\right)$. However convergence will be achieved *in finite time* if the total relative degree $r = n$ and no internal dynamics exist.

CHAPTER 4

Attacks Reconstruction in Linearized Cyber Physical Systems: The Number of Sensors Is Equal to the Number of Potential Attacks

4.1 Introduction

Consider the linearized format of a CPS eq. (2.10) as

$$\begin{aligned} \dot{x} &= Ax + Bd \\ y &= Cx + Dd \end{aligned} \tag{4.1}$$

The problem is to reconstruct the norm-bounded smooth sensor attack $d(t) \in \mathbb{R}^q$, where $\|d(t)\| \le L_1,\ \|\dot{d}(t)\| \le L_2,\ L_1, L_2 > 0$ and find the state estimation so that the estimate

$$\hat{d}(t) \to d(t), \quad \hat{x}(t) \to x(t) \tag{4.2}$$

as time increases.

Equation (4.1) can be the model of a linear CPS that is controlled by a feedback control which uses the corrupted measurements as it is shown in Figure 4.1.

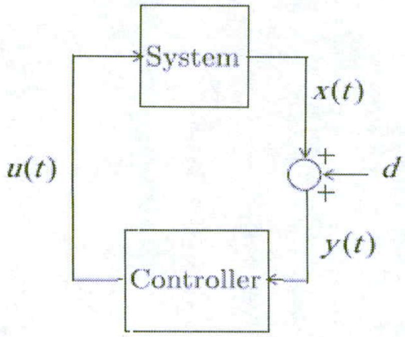


Figure 4.1 The closed-loop CPS in the Presence of Sensor Attacks

The dynamics of CPS in Figure 4.1 is given by

$$\begin{aligned} \dot{x} &= \bar{A}x + \bar{B}u \\ y &= Cx + Dd \end{aligned} \tag{4.3}$$

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

The following assumption is made:

(A 4.1): the Kimura-Davison condition [97]

$$m + p + 1 \geq n \tag{4.4}$$

holds.

Assuming assumption (A 4.1) holds then there exists a static output feedback control

$$u = -Ky \tag{4.5}$$

where $K \in \mathbb{R}^{m \times q}$ is a gain matrix, that stabilizes the CPS in eq. (4.1).

Substituting control input eq. (4.5) into CPS eq. (4.1) results in the following closed loop CPS

$$\begin{aligned} \dot{x} &= \left(\bar{A} - \bar{B}KC\right)x - \bar{B}KDd \\ y &= Cx + Dd \end{aligned} \tag{4.6}$$

The closed-loop CPS eq. (4.6) can be rewritten as eq. (4.1) where $A = \bar{A} - \bar{B}KC$ is Hurwitz, and $B = -\bar{B}KD$.

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

$y_{clean} = y - D\hat{d}$, then $y_{clean} - \tilde{y} \to 0$ 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 eq. (4.5) gives

$$\begin{cases} \dot{x} = Ax + Bd \\ y = Cx + Dd \end{cases} \xrightarrow[\text{as time increases}]{} \begin{cases} \dot{x} = Ax \\ \tilde{y} = Cx \end{cases} \tag{4.7}$$

In the other words, the compensated dynamics of the CPS eq. (4.1), whose sensors are under attack, will converge as time increases to the stable CPS eq. (4.7) with the desired asymptotic dynamics that are not affected by the sensor attack signals. The problem of the output feedback controller eq. (4.5) design is out of the scope of this dissertation.

The main problem addressed in this chapter is on-line reconstruction of the sensor attack signal $d(t)$ and state estimation in CPS eq. (4.1) as it is shown in Figure 4.2.

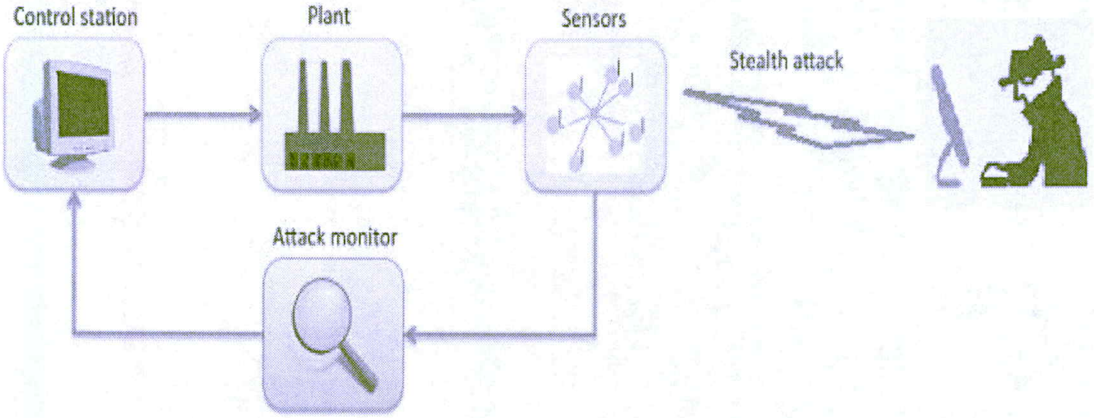


Figure 4.2 Sensor Attack Analyzer

4.2 On-line Attack Reconstruction:

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

(a) all sensors $p = q$ can be attacked,

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

4.2.1 Attack Reconstruction: All Sensors Can be Attacked

Since the CPS eq. (4.1) is completely controllable and observable, it can be partitioned as

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} d$$
$$y = C_1 x_1 + C_2 x_2 + Dd \tag{4.8}$$

where $p = q$, $x_1 \in \mathbb{R}^{n-p}$, $x_2 \in \mathbb{R}^p$, $C_1 \in \mathbb{R}^{p\times(n-p)}$, $C_2 \in \mathbb{R}^{p\times p}$, $D \in \mathbb{R}^{p\times q}$, and $\det(C_2) \neq 0$.

Firstly, the closed-loop CPS eq. (4.1) is transformed to a form convenient for the observer design. Specifically, the state variable $x_2 \in \mathbb{R}^p$ is replaced by the sensor measurements $y \in \mathbb{R}^p$. This is

$$\begin{aligned} \dot{x}_1 &= G_{21} x_1 + G_{22} y + G_{23} d \\ \dot{y} &= G_{11} x_1 + G_{12} y + G_{13} d + D\dot{d} \end{aligned} \tag{4.9}$$

where

$$\begin{aligned}
G_{11} &= C_1A_{11} - C_1A_{12}C_2^{-1}C_1 + C_2A_{21} - C_2A_{22}C_2^{-1}C_1 \\
G_{12} &= C_1A_{12}C_2^{-1} + C_2A_{22}C_2^{-1} \\
G_{13} &= -C_1A_{12}C_2^{-1}D + C_1B_1 - C_2A_{22}C_2^{-1}D + C_2B_2 \\
G_{21} &= A_{11} - A_{12}C_2^{-1}C_1 \\
G_{22} &= A_{12}C_2^{-1} \\
G_{23} &= -A_{12}C_2^{-1}D + B_1
\end{aligned} \tag{4.10}$$

An observer is designed mimicking system CPS eq. (4.9)

$$\begin{aligned}
\dot{\hat{x}}_1 &= G_{21}\hat{x}_1 + G_{22}\hat{y} \\
\dot{\hat{y}} &= G_{11}\hat{x}_1 + G_{12}\hat{y} + \upsilon
\end{aligned} \tag{4.11}$$

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

$$\begin{aligned}
e_y &= y - \hat{y} \\
e_{x_1} &= x_1 - \hat{x}_1
\end{aligned} \tag{4.12}$$

The following assumptions are made concerning matrices in eqs. (4.9) and (4.10):

(A 4.2): The matrix G_{21} is Hurwitz.

(A 4.3): 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.

(A 4.4): 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:

$$\left\|\varphi\right\| \le L_{G_{11}}L_{e_{x1}} + L_{G_{12}}L_{e_y} + L_{G_{13}}L_1 + L_DL_2 \le L_3 \tag{4.13}$$

where $\left\|G_{11}\right\|_\infty \le L_{G_{11}}$, $\left\|G_{12}\right\|_\infty \le L_{G_{12}}$, $\left\|G_{13}\right\|_\infty \le L_{G_{13}}$, $\left\|D\right\|_\infty \le L_D$, $\left\|e_{x_1}\right\| \le L_{e_{x1}}$, $\left\|e_y\right\| \le L_{e_y}$, $L_{G_{11}}, L_{e_{x1}}, L_{G_{12}}, L_{e_y}, L_{G_{13}}, L_1, L_D, L_2, L_3 > 0$.

4.2.1.1 Fixed-gain Sliding Mode Observer

The first main result is formulated in the following Theorem:

Theorem 4.1 [81]: Consider the CPS in eqs. (4.9) and (4.10) with the observer eq. (4.11), whose injection term υ is designed in a unit vector format

$$\upsilon = \left(\rho + L_3\right)\frac{e_y}{\left\|e_y\right\|},\ \rho, L_3 > 0 \tag{4.14}$$

that makes the observer eq. (4.11) and eq. (4.14) the SMO. Assume that the assumptions (A 4.1) – (A 4.4) hold. Then the sensor attack signal $d(t)$ is exactly reconstructed as

$$\hat{d} = \left(G_{11}\left(sI - G_{21}\right)^{-1} G_{23} + G_{13} + Ds\right)^{-1} \upsilon_{eq} \tag{4.15}$$

where υ_{eq} is the *equivalent* injection function, $e_y \rightarrow 0$ in finite time, and $\hat{d}(t) \rightarrow d(t)$ as time increases in the sliding mode. This novel result is published in [81].

The proof of the Theorem 4.1 is presented in Appendix.

Remark 4.1: Unlike in the estimation algorithms presented in [91], where the attack term $d(t)$ is assumed slow varying ($\dot{d}(t) \approx 0$), in this work it is assumed that $\dot{d}(t) \neq 0$. In order to exactly reconstruct the time varying attack $d(t)$ the *dynamic extension* of the equivalent control υ_{eq} is proposed as in (4.15). This is the major novelty of the proposed attack reconstruction algorithm in eqs. (4,14), (4.15).

Remark 4.2: Given *equivalent* control υ_{eq} the attack estimate in eq. (4.15), where the dynamic filter appears naturally, is exact.

Remark 4.3: 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 eq. (4.14) [98]. Therefore, if $\bar{\upsilon}_{eq}$ satisfies

$$\tau\dot{\bar{\upsilon}}_{eq} = (\rho + L_3)\frac{e_y}{\|e_y\|} - \bar{\upsilon}_{eq} \quad , \tag{4.16}$$

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

$$\|\bar{\upsilon}_{eq} - \upsilon_{eq}\| \sim \mathrm{O}(\tau). \tag{4.17}$$

Therefore, the υ_{eq} estimation error in eq. (4.17) is small for a small enough choice of τ [98].

Replacing υ_{eq} by $\bar{\upsilon}_{eq}$ in (4.15) we obtain

$$\bar{d} = \left(G_{11}(sI - G_{21})^{-1} G_{23} + G_{13} + Ds\right)^{-1} \bar{\upsilon}_{eq}. \tag{4.18}$$

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

$$\|\bar{d} - d\| \le \Lambda \|\bar{\upsilon}_{eq} - \upsilon_{eq}\| \sim \mathrm{O}(\tau) \quad , \tag{4.19}$$

where

$$\left\|\left(G_{11}(sI - G_{21})^{-1} G_{23} + G_{13} + Ds\right)^{-1}\right\|_{\infty} = \Lambda,\ \Lambda > 0 \ . \tag{4.20}$$

Note that the low pass filter in eq. (4.16) is the simplest choice, but other higher order systems with low-pass characteristics can be employed.

Remark 4.4: In many practical cases the entries of the transfer function $\left(G_{11}(sI - G_{21})^{-1} G_{23} + G_{13} + Ds\right)^{-1}$ of the estimator eq. (4.15) are the regular ones. This fact is demonstrated in the case study. It means that the SMC injection term υ in eq. (4.14) can be used in eq. (4.15) 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 $\left(G_{11}\left(sI-G_{21}\right)^{-1}G_{23}+G_{13}+Ds\right)^{-1}$.

4.2.1.2 Adaptive-gain Sliding Mode Observer

In subsection 4.1.2.1 it was assumed that the perturbations term φ is locally norm-bounded as in eq. (4.13), 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 eq. (4.14) in the fixed gain SMO in eq. (4.11) 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 attack on-line reconstruction. The following assumption is made:

Assumption (A 4.5): The disturbance term φ satisfies the conditions

$$\|\varphi\|\le L_3, \|\dot{\varphi}\|\le L_4, \tag{4.21}$$

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

The *dual layer nested adaptive SMO* [99] is used for designing the injection term υ in eq. (4.14). In accordance with the *dual layer nested adaptive sliding mode observation algorithm* [99], the constant gain L_3 in the injection term eq. (4.14) is to be replaced by the adaptive gain $L(t)$ (without $L(t)$ overestimation). This is

$$\upsilon=\left(\rho+L(t)\right)\frac{e_y}{\|e_y\|},\ \rho>0. \tag{4.22}$$

Following the *dual layer nested sliding mode observation adaptive algorithm* in [99] applied to the unit-vector injection term in eq. (4.14), an error signal is defined as

$$\sigma(t) = L(t) - \frac{1}{\alpha}\left\|\bar{\upsilon}_{eq}(t)\right\| - \varepsilon, \tag{4.23}$$

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

$$\tau\dot{\bar{\upsilon}}_{eq} = \upsilon - \bar{\upsilon}_{eq}. \tag{4.24}$$

The task of selecting $\tau > 0$ is discussed in Remark 4.3.

The adaptation dynamics of $L(t)$ in eq. (4.22) are defined as [99]

$$\dot{L}(t) = -r(t) sign(\sigma(t)), \tag{4.25}$$

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, $\|\dot{\varphi}\| \le L_4$, by some finite time. In this paper it is assumed that $r(t)$ has the structure

$$r(t) = \ell_0 + \ell(t) \tag{4.26}$$

where ℓ_0 is a fixed positive scalar. The evolution of $\ell(t)$ is chosen to satisfy an adaptive law [99]

$$\dot{\ell}(t) = \begin{cases} \gamma\left|\sigma(t)\right| & if\ \left|\sigma(t)\right| > \sigma_0 \\ 0 & otherwise \end{cases} \tag{4.27}$$

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

The second main result is summarized in the following proposition.

Proposition 4.1 [81]: Consider the CPS in eq. (4.9) and eq. (4.10), and assume that the assumptions (A 4.1) – (A 4.5) hold. A SMO is designed as in eq. (4.11) with the *adaptive* injection term in eqs. (4.22) - (4.27). If $\varepsilon > 0$ in eq. (4.23) is chosen to satisfy

$$\frac{1}{4}\varepsilon^2 > \sigma_0^2 + \frac{1}{\gamma}\left(\frac{\kappa L_4}{\alpha}\right)^2 \tag{4.28}$$

for any given σ_0 in eq. (4.27), L_4 in eq. (4.21), $\kappa > 1$, and, $0 < \alpha < 1$, then

- the injection term eq. (4.22) exploiting the *dual layer adaptive* scheme given by eqs. (4.23)-(4.27), drives $\sigma(t)$ to a domain $|\sigma(t)| < \varepsilon/2$ 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 eq. (4.15) with the equivalent adaptive injection term υ_{eq} or $\bar{\upsilon}_{eq}$.

The proof of the Proposition 4.1 is presented in the Appendix.

Remark 4.5: The proposed unit vector injection gain-adaptation algorithm in eqs. (4.22) - (4.27) does not require the knowledge of boundaries $L_3, L_4 > 0$.

Discussion: In accordance with Theorem 4.1, the fixed-gain injection term eq. (4.14) depends on a magnitude L_3 of the perturbation φ in eq. (4.13). Note that φ contains the attack and its derivative terms $d, \dot{d}$. Apparently, the value L_3 is difficult to estimate. Therefore, the gain $\rho + L_3$ of the fixed-gain injection term in eq. (4.14) could be significantly overestimated, and that can amplify the residual rippling in the attack estimation eq. (4.15). On the other hand, the adaptive-gain injection term in eqs. (4.22) - (4.27) has a non-overestimated gain that is automatically tuned without any prior knowledge about L_3. This provides a minimal residual rippling level in eq. (4.15) as well as convenience (self-tuning) in the observer implementation.

4.2.2 Attack Reconstruction: Some Sensors Are Protected From the Attacks

Again consider the system (4.1), whose sensors are under attacks where k sensors ($k < p,\ p = q$) are protected from attacks

Assumption (A 4.6): k out of p sensors are protected, and the remaining $p-k$ sensors might be attacked/corrupted.

Separating the protected and unprotected measurements, CPS eq. (4.1) can be partitioned as

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} d$$
$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ D_1 \end{bmatrix} d \tag{4.29}$$

where

$$\begin{aligned} \bar{y} &= \left[y_1^T, y_2^T \right]^T, y_1 \in \mathbb{R}^k, y_2, \in \mathbb{R}^{p-k} \\ x &= \left[x_1^T, x_2^T, x_3^T \right]^T, x_1 \in \mathbb{R}^{n-p}, x_2 \in \mathbb{R}^k, x_3 \in \mathbb{R}^{p-k} \\ B &= \left[B_1^T, B_2^T, B_3^T \right]^T, B_1 \in \mathbb{R}^{(n-p)\times p}, B_2 \in \mathbb{R}^{k\times p}, B_3 \in \mathbb{R}^{(p-k)\times p} \\ D_1 &\in \mathbb{R}^{(p-k)\times p} \end{aligned} \tag{4.30}$$

It is assumed that

Assumption (A 4.7): 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 CPS eqs. (4.29), (4.30) are transformed to a convenient form for the Lyapunov analysis. 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:

$$
\begin{aligned}
\dot{x}_1 &= Q_{11}x_1 + Q_{12}y_1 + Q_{13}y_2 + Q_{14}d \\
\dot{y}_1 &= Q_{21}x_1 + Q_{22}y_1 + Q_{23}y_2 + Q_{24}d \\
\dot{y}_2 &= Q_{31}x_1 + Q_{32}y_1 + Q_{33}y_2 + Q_{34}d + D_1\dot{d}
\end{aligned}
\tag{4.31}
$$

where

$$
\begin{aligned}
Q_{11} &= A_{11} + A_{12}h_{11} + A_{13}h_{21} \\
Q_{12} &= A_{12}h_{12} + A_{13}h_{22} \\
Q_{13} &= A_{12}h_{13} + A_{13}h_{23} \\
Q_{14} &= A_{12}h_{14} + A_{13}h_{24} + B_1 \\
Q_{21} &= C_{11}A_{11} + C_{12}A_{21} + C_{13}A_{31} \\
&\quad + \left(C_{11}A_{12} + C_{12}A_{22} + C_{13}A_{32}\right)h_{11} + \left(C_{11}A_{13} + C_{12}A_{23} + C_{13}A_{33}\right)h_{21} \\
Q_{22} &= \left(C_{11}A_{12} + C_{12}A_{22} + C_{13}A_{32}\right)h_{12} + \left(C_{11}A_{13} + C_{12}A_{23} + C_{13}A_{33}\right)h_{22}, \\
Q_{23} &= \left(C_{11}A_{12} + C_{12}A_{22} + C_{13}A_{32}\right)h_{13} + \left(C_{11}A_{13} + C_{12}A_{23} + C_{13}A_{33}\right)h_{23}, \\
Q_{24} &= C_{11}B_1 + C_{12}B_2 + C_{13}B_3 \\
Q_{31} &= C_{21}A_{11} + C_{22}A_{21} + C_{23}A_{31} \\
&\quad + \left(C_{21}A_{12} + C_{22}A_{22} + C_{23}A_{32}\right)h_{11} + \left(C_{21}A_{13} + C_{22}A_{23} + C_{23}A_{33}\right)h_{21} \\
Q_{32} &= \left(C_{21}A_{12} + C_{22}A_{22} + C_{23}A_{32}\right)h_{12} + \left(C_{21}A_{13} + C_{22}A_{23} + C_{23}A_{33}\right)h_{22}, \\
Q_{33} &= \left(C_{21}A_{12} + C_{22}A_{22} + C_{23}A_{32}\right)h_{13} + \left(C_{21}A_{13} + C_{22}A_{23} + C_{23}A_{33}\right)h_{23}, \\
Q_{34} &= C_{21}B_1 + C_{22}B_2 + C_{23}B_3
\end{aligned}
\tag{4.32}
$$

with

$$
\begin{aligned}
h_{11} &= \left(I - C_{12}{}^{-1}C_{13}C_{23}{}^{-1}C_{22}\right)^{-1}C_{12}{}^{-1}\left(-C_{11} + C_{13}C_{23}{}^{-1}C_{21}\right) \\
h_{12} &= \left(I - C_{12}{}^{-1}C_{13}C_{23}{}^{-1}C_{22}\right)^{-1}C_{12}{}^{-1} \\
h_{13} &= -\left(I - C_{12}{}^{-1}C_{13}C_{23}{}^{-1}C_{22}\right)^{-1}C_{12}{}^{-1}C_{13}C_{23}{}^{-1} \\
h_{14} &= -\left(I - C_{12}{}^{-1}C_{13}C_{23}{}^{-1}C_{22}\right)^{-1}C_{12}{}^{-1}C_{13}C_{23}{}^{-1}D_1 \\
h_{21} &= \left(I - C_{23}{}^{-1}C_{22}C_{12}{}^{-1}C_{13}\right)^{-1}C_{23}{}^{-1}\left(-C_{21} + C_{22}C_{12}{}^{-1}C_{11}\right) \\
h_{22} &= -\left(I - C_{23}{}^{-1}C_{22}C_{12}{}^{-1}C_{13}\right)^{-1}C_{23}{}^{-1}C_{22}C_{12}{}^{-1} \\
h_{23} &= \left(I - C_{23}{}^{-1}C_{22}C_{12}{}^{-1}C_{13}\right)^{-1}C_{23}{}^{-1} \\
h_{24} &= -\left(I - C_{23}{}^{-1}C_{22}C_{12}{}^{-1}C_{13}\right)^{-1}C_{23}{}^{-1}D_1
\end{aligned}
\tag{4.33}
$$

The main results relating to sensor attack reconstruction in systems with

$k < p,\ p = q$ sensors protected from the attacks are presented in the following Theorem that presents the novel original results which was published at [81].

Theorem 4.2 [81]: Consider CPS in eqs. (4.31) - (4.33), and assume that assumption (A 4.7) holds. The proposed SMO is given by

$$\begin{aligned} \dot{\hat{x}}_1 &= Q_{11}\hat{x}_1 + Q_{12}\hat{y}_1 + Q_{13}\hat{y}_2 \\ \dot{\hat{y}}_1 &= Q_{21}\hat{x}_1 + Q_{22}\hat{y}_1 + Q_{23}\hat{y}_2 + \upsilon_1 \\ \dot{\hat{y}}_2 &= Q_{31}\hat{x}_1 + Q_{32}\hat{y}_1 + Q_{33}\hat{y}_2 + \upsilon_2 \end{aligned} \tag{4.34}$$

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

$$\begin{aligned} \upsilon_1 &= \left(\rho_1 + L_{11}\right)\frac{e_{y_1}}{\left\|e_{y_1}\right\|} \\ \upsilon_2 &= \left(\rho_2 + L_{12}\right)\frac{e_{y_2}}{\left\|e_{y_2}\right\|} \end{aligned} \tag{4.35}$$

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

$$e_{x_1} = x_1 - \hat{x}_1,\quad e_{y_1} = y_1 - \hat{y}_1,\quad e_{y_2} = y_2 - \hat{y}_2 \tag{4.36}$$

Then the sensor attack is estimated as

$$\hat{d} = \begin{bmatrix} \hat{d}_1 \\ \hat{d}_2 \end{bmatrix} = \begin{bmatrix} H_{11}(s)^{-1}\left(\upsilon_{1eq} - H_{12}(s)\hat{d}_2\right) \\ \left(-H_{21}(s)H_{11}(s)^{-1}H_{12}(s) + H_{22}(s)\right)^{-1}\left(\upsilon_{2eq} - H_{21}(s)H_{11}(s)^{-1}\upsilon_{1eq}\right) \end{bmatrix} \tag{4.37}$$

in the sliding mode $\forall t \ge t_r$ where $t = t_r$ is the 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

$$\begin{bmatrix} Q_{21}\left(sI - Q_{11}\right)^{-1}Q_{14} + Q_{24} \\ Q_{31}\left(sI - Q_{11}\right)^{-1}Q_{14} + Q_{34} + D_1 sI \end{bmatrix} = \begin{bmatrix} H_{11}(s) & H_{12}(s) \\ H_{21}(s) & H_{22}(s) \end{bmatrix} \tag{4.38}$$

while the matrices $H_{11}(s) \in \mathbb{R}^{k \times k}$, $H_{12}(s) \in \mathbb{R}^{k \times (p-k)}$, $H_{21}(s) \in \mathbb{R}^{(p-k) \times k}$, $H_{22}(s) \in \mathbb{R}^{(p-k) \times (p-k)}$.

The proof of the Theorem 4.2 is presented in Appendix.

Remark 4.6: Unlike in the estimation algorithms presented in [91], where the attack term $d(t)$ is assume slow varying ($\dot{d}(t) \approx 0$), in this work it is assumed that $\dot{d}(t) \neq 0$. In order to exactly reconstruct the time varying attack $d(t)$ the *dynamic extension* of the equivalent control $\upsilon_{1eq}, \upsilon_{2eq}$ is proposed as in eq. (4.37), (4.38). This is the major novelty of the proposed attack reconstruction algorithm in (4.35) - (4.38).

Remark 4.7: The injection terms υ_1, υ_2 in eq. (4.35) can be also designed in the dual layer adaptive form as shown in section 4.2.1.2. Note that the problem of estimating $\upsilon_{1eq}, \upsilon_{2eq}$ used in eq. (4.37) is discussed in Remark 4.3.

Remark 4.8: Estimating the attack vector $d = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$ in eq. (4.37) requires inversion of the matrices (that represent the dynamic filters) of smaller dimensions rather than in eq. (4.15) due the fact that $k < p$ sensors are protected from the attacks.

4.3 **Summary**

In this chapter, linearized cyber-physical systems when the number of sensors is equal to the number of potential attacks are studied. Two different cases are considered. At first, all of the sensors are prone to get attacked. Then, some of sensors are protected from attacks. Novel finite time convergent SMOs, including observer with gain adaptation, which use the dynamic extension of injection term, are proposed for on-line reconstruction of the sensor attacks and estimation of states. The filters that address the attack propagation dynamics are proposed and employed for the attack reconstruction for the first time and this novel result is published in [81]. 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.

In next chapter, we will discuss attack reconstruction for linearized CPSs when the number of measurements is greater than the number of potential attacks.

CHAPTER 5

Attacks Reconstruction in Linearized Cyber Physical Systems: the Number of Sensors is Greater than Number of Potential Attacks

Two approach of attack reconstruction are discussed in this chapter. At first, a novel SMO with dynamic extension of the injection term is proposed. Next, an adaptive line-by-line super twisting SMO is developed to estimate the states of CPSs and reconstruct the plant attacks.

5.1 Sliding Mode Observer with Dynamic Extension of Injection Term

A fixed and an adaptive SMO based on filtering of injection term is designed to reconstruct the attacks in this section.

5.1.1 Introduction

Consider the linearized closed loop CPS eq. (2.10), this is

$$\begin{aligned} \dot{x} &= Ax + Bd \\ y &= Cx + Dd \end{aligned} \tag{5.1}$$

where $x \in \mathbb{R}^n$ denotes the CPS states, $y \in \mathbb{R}^p$ represents the measured output, $d(t) \in \mathbb{R}^q$ is the attack signal, and $n > p > q$.

The objective is the online reconstructing of attack signal $d(t) \in \mathbb{R}^q$ in linearized CPS in eq. (5.1) knowing that $n > p > q$.

Assume that:

(A 5.1) the pair (A,C) is observable and the output distribution matrix C has full row rank;

(A 5.2) $Rank(D)=q$.

(A 5.3) The attack $d(t)$ and its derivative are norm bounded, i.e. $\|d\|<k_d$ and $\|\dot{d}\|<l_d$ where $k_d,l_d>0$ are known.

5.1.2 System's Transformation

Assume the assumptions (A 5.1) and (A 5.2) hold, then [91] there exists a matrix $N \in R^{(n-p)\times n}$ such that the square matrix

$$T_c=\begin{bmatrix} N \\ C \end{bmatrix} \tag{5.2}$$

is nonsingular and the change of coordinates $x \mapsto T_c x$ creates, without loss of generality, a new state-space representation (A',B',C',D) where

$$A'=\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}=T_c A T_c^{-1},\quad B'=\begin{bmatrix} B_1 \\ B_2 \end{bmatrix}=T_c B,\quad C'=CT_c^{-1}=\begin{bmatrix} \mathbf{0}_{p\times(n-p)} & I_{p\times p} \end{bmatrix}, \tag{5.3}$$

After linear changing of coordinate, CPS eq. (5.1) is rewritten as

$$\begin{aligned} \dot{x}_1 &= A_{11}x_1 + A_{12}x_2 + B_1 d \\ \dot{x}_2 &= A_{21}x_1 + A_{22}x_2 + B_2 d \end{aligned} \tag{5.4}$$

and the sensor measurement is

$$y = x_2 + Dd\ , \tag{5.5}$$

where $x_1 \in R^{n-p}$, $x_2 \in R^p$, $B_1 \in R^{(n-p)\times q}$, $B_2 \in R^{p\times q}$, and $A_{11} \in R^{(n-p)\times(n-p)}$,

$A_{12} \in R^{(n-p)\times p}$, $A_{21} \in R^{p\times(n-p)}$, $A_{22} \in R^{p\times p}$.

It is well-known that (A, C) is observable if and only if (A_{11}, A_{21}) is observable [91].

Define a further change of coordinates $\bar{x}_1 = x_1 + Lx_2$ where $L \in \mathbb{R}^{(n-p)\times p}$ is the design matrix, then the system eqs. (5.4) - (5.5) can be re-written as

$$\begin{aligned} \dot{\bar{x}}_1 &= \tilde{A}_{11}\bar{x}_1 + \tilde{A}_{12}x_2 + \tilde{B}_1 d \\ \dot{x}_2 &= \tilde{A}_{21}\bar{x}_1 + \tilde{A}_{22}x_2 + \tilde{B}_2 d \ , \\ y &= x_2 + Dd \end{aligned} \tag{5.6}$$

where $\tilde{A}_{11} = A_{11} + LA_{21}$, $\tilde{A}_{12} = -A_{11}L + A_{12} - LA_{21}L + LA_{22}$, $\tilde{B}_1 = B_1 + LB_2$, $\tilde{A}_{21} = A_{21}$, $\tilde{A}_{22} = A_{22} - A_{21}L$, $\tilde{B}_2 = B_2$. Since (A_{11}, A_{21}) is observable there exist choices of the matrix L so that the matrix $\tilde{A}_{11} = A_{11} + LA_{21}$ is Hurwitz.

Since the number of sensor measurements are greater than the number of attacks, $p > q$, there is a nonsingular (output) scaling matrix $Q \in R^{p\times p}$ chosen such that

$$QD = \begin{bmatrix} \mathbf{0}_{(p-q)\times q} \\ D_2 \end{bmatrix}, \tag{5.7}$$

where $D_2 \in R^{q\times q}$ is nonsingular. Assume that the assumption (A 5.2) holds, then the matrix Q in (5.7) can be obtained by Gaussian elimination (or QR reduction).

Define $\bar{y}$ as the scaling of the measured outputs y according to

$$\bar{y} = Qy \tag{5.8}$$

Then, the output of CPS can be partitioned as unpolluted measurements $\bar{y}_1$ and polluted measurements $\bar{y}_2$ as

$$\bar{y} = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \end{bmatrix} = \begin{bmatrix} Q_1 x_2 \\ Q_2 x_2 + D_2 d \end{bmatrix} = Qx_2 + \begin{bmatrix} \mathbf{0}_{(p-q)\times q} \\ D_2 \end{bmatrix} d \tag{5.9}$$

where $\bar{y}_1 \in \mathbb{R}^{p-q}$ and $\bar{y}_2 \in \mathbb{R}^{q}$.

For convenience scale the state component x_2 and define $\bar{x}_2 = Qx_2$. Then eq. (5.6) can be rewritten as

$$\begin{aligned} \dot{\bar{x}}_1 &= \bar{A}_{11}\bar{x}_1 + \bar{A}_{12}\bar{x}_2 + \bar{B}_1 d \\ \dot{\bar{x}}_2 &= \bar{A}_{21}\bar{x}_1 + \bar{A}_{22}\bar{x}_2 + \bar{B}_2 d \end{aligned}, \tag{5.10}$$

where

$$\bar{y} = \bar{x}_2 + \begin{bmatrix} \mathbf{0} \\ D_2 \end{bmatrix} d \ , \tag{5.11}$$

with $\bar{A}_{11} = \tilde{A}_{11}$, $\bar{A}_{12} = \tilde{A}_{12}Q^{-1}$, $\bar{B}_1 = \tilde{B}_1$, $\bar{A}_{21} = Q\tilde{A}_{21}$, $\bar{A}_{22} = Q\tilde{A}_{22}Q^{-1}$, and $\bar{B}_2 = Q\tilde{B}_2$.

Define $\bar{x}_2 = col\left(\bar{x}_{21}, \bar{x}_{22}\right)$, where $\bar{x}_{21} \in \mathbb{R}^{p-q}$ and $\bar{x}_{22} \in \mathbb{R}^{q}$. Note that by definition $\bar{x}_{21}$ is known (since it is obtained from protected sensor measurement y) and it is free from the attack i.e. independent of d. Consequently the system CPS in eqs. (5.10)-(5.11) can be written in partitioned form as

$$\begin{aligned} \dot{\bar{x}}_1 &= \bar{A}_{11}\bar{x}_1 + \bar{A}_{12a}\bar{x}_{21} + \bar{A}_{12b}\bar{x}_{22} + \bar{B}_1 d \\ \dot{\bar{x}}_{21} &= \bar{A}_{21a}\bar{x}_1 + \bar{A}_{22a}\bar{x}_{21} + \bar{A}_{22b}\bar{x}_{22} + \bar{B}_{21} d \\ \dot{\bar{x}}_{22} &= \bar{A}_{21b}\bar{x}_1 + \bar{A}_{22c}\bar{x}_{21} + \bar{A}_{22d}\bar{x}_{22} + \bar{B}_{22} d \end{aligned} \tag{5.12}$$

where the scaled sensor measurements from eq. (5.11) are

$$\bar{y}_1 = \bar{x}_{21}, \quad \bar{y}_2 = \bar{x}_{22} + D_2 d \ . \tag{5.13}$$

Finally, CPS eqs. (5.12)-(5.13) are rewritten as

$$\begin{aligned} \dot{\bar{x}} &= \bar{A}\bar{x} + \bar{B}d \\ \bar{y}_1 &= \bar{C}_1\bar{x} \\ \bar{y}_2 &= \bar{C}_2\bar{x} + D_2 d \end{aligned} \tag{5.14}$$

where

$$\bar{x}=\begin{bmatrix}\bar{x}_1\\ \bar{x}_{21}\\ \bar{x}_{22}\end{bmatrix},\quad \bar{A}=\begin{bmatrix}\bar{A}_{11} & \bar{A}_{12a} & \bar{A}_{12b}\\ \bar{A}_{21a} & \bar{A}_{22a} & \bar{A}_{22b}\\ \bar{A}_{21b} & \bar{A}_{22c} & \bar{A}_{22d}\end{bmatrix},\quad \bar{B}=\begin{bmatrix}\bar{B}_1\\ \bar{B}_{21}\\ \bar{B}_{22}\end{bmatrix}$$
$$\bar{C}_1=\begin{bmatrix}\mathbf{0}_{(p-m)\times(n-p)} & I_{(p-m)\times(p-m)} & \mathbf{0}_{(p-m)\times m}\end{bmatrix},\ \bar{C}_2=\begin{bmatrix}\mathbf{0}_{m\times(n-m)} & I_{m\times m}\end{bmatrix} \tag{5.15}$$

where $\bar{A}_{11}$ is Hurwitz, and the virtual vector measurement $\bar{y}_1$ is not corrupted (we can consider it as the vector of protected sensor measurements) and $\bar{y}_2$ is the vector of attacked/corrupted measurements.

5.1.3 **Fixed-gain Sliding Mode Observer Design**

Define a (sliding mode) observer for the CPS eqs. (5.14) - (5.15) as

$$\dot{\bar{z}}=\bar{A}\bar{z}+\bar{G}_1\left(\bar{y}_1-\bar{z}_{21}\right)+\bar{G}_2\left(\bar{y}_2-\bar{z}_{22}\right)-G_n\upsilon \tag{5.16}$$

where $\bar{z}=col\left(\bar{z}_1,\bar{z}_{21},\bar{z}_{22}\right)$ and the partition of $\bar{z}$ is conformal with the partition of $\bar{x}$ of CPS in eq. (5.15). The signal υ in eq. (5.16) is a nonlinear injection signal that depends on $\left(\bar{y}_2-\bar{z}_{22}\right)$ and is used to induce a sliding motion in the estimation error space.

Define the gain matrices in observer eq. (5.16) as

$$\bar{G}_1=\begin{bmatrix}\bar{A}_{12a}\\ \bar{A}_{22a}-A_{22}^s\\ \mathbf{0}_{m\times(p-m)}\end{bmatrix},\bar{G}_2=\begin{bmatrix}\bar{A}_{12b}\\ \bar{A}_{22b}\\ \bar{A}_{22d}-A_{33}^s\end{bmatrix},G_n=\begin{bmatrix}\mathbf{0}_{(n-p)\times m}\\ \mathbf{0}_{(p-m)\times m}\\ \mathrm{I}_{m\times m}\end{bmatrix} \tag{5.17}$$

where $\bar{A}_{12a}\in\mathbb{R}^{(n-p)\times(p-m)}$, $\bar{A}_{22a}\in\mathbb{R}^{(p-m)\times(p-m)}$, $\bar{A}_{12b}\in\mathbb{R}^{(n-p)\times m}$, $\bar{A}_{22b}\in\mathbb{R}^{(p-m)\times m}$, $\bar{A}_{22d}\in\mathbb{R}^{m\times m}$ and the matrices $A_{22}^s\in\mathbb{R}^{(p-m)\times(p-m)}$, $A_{33}^s\in\mathbb{R}^{m\times m}$ are user selected Hurwitz matrices. In particular, assume that A_{33}^s is symmetric negative definite. The injection signal $\upsilon\in\mathbb{R}^m$ is defined as

$$\upsilon = -(\rho+\eta)\frac{\bar{y}_2 - \bar{z}_{22}}{\left\|\bar{y}_2 - \bar{z}_{22}\right\|},\ \ \rho,\eta > 0 \tag{5.18}$$

where the modulation scalar gain ρ will be defined in the sequel and η is a positive design scalar.

It is assumed that

(A 5.4) The matrix $\left(sI - A^*\right)^{-1}$ is Hurwitz, where

$$A^* = \bar{A} - \bar{B}D_2^{-1}\bar{C}_2 - \bar{G}_1\bar{C}_1 . \tag{5.19}$$

Define $\bar{e} = \bar{x} - \bar{z}$. Then it follows $\bar{e} = col\left(\bar{e}_1, \bar{e}_{21}, \bar{e}_{22}\right)$ where

$$\bar{e}_1 = \bar{x}_1 - \bar{z}_1,\quad \bar{e}_{21} = \bar{x}_{21} - \bar{z}_{21},\quad \bar{e}_{22} = \bar{x}_{22} - \bar{z}_{22} . \tag{5.20}$$

It follows

$$e_{y_2} = \bar{y}_2 - \bar{z}_{22} = \bar{e}_{22} + D_2 d\ , \tag{5.21}$$

and by direct substitution from eqs. (5.12) and (5.16) it is given that

$$\dot{\bar{e}} = \begin{bmatrix} \bar{A}_{11} & \mathbf{0} & \mathbf{0} \\ \bar{A}_{21a} & A_{22}^{s} & \mathbf{0} \\ \bar{A}_{21b} & \bar{A}_{22c} & A_{33}^{s} \end{bmatrix} \bar{e} - \begin{bmatrix} \bar{A}_{12b} \\ \bar{A}_{22b} \\ \bar{A}_{22d} - A_{33}^{s} \end{bmatrix} D_2 d + \begin{bmatrix} \bar{B}_1 \\ \bar{B}_{21} \\ \bar{B}_{22} \end{bmatrix} d + \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathrm{I}_{\mathrm{m}} \end{bmatrix} \upsilon . \tag{5.22}$$

The idea is to force a sliding motion on

$$e_{y_2} = \bar{y}_2 - \bar{z}_{22} = \mathbf{0} \tag{5.23}$$

The first main novel result that is based on the SMO with the fixed-gain injection term, is formulated in the following Theorem 5.1.

Theorem 5.1 [82]. Assuming the assumptions (A 5.1)-(A 5.4) hold, and $m_0 > 0$ satisfies the conditions

$$\begin{cases} \|\phi(t)\| \le m_0 k_d \\ \phi = \begin{bmatrix} \bar{A}_{21b} & \bar{A}_{22c} \end{bmatrix} \bar{e}_{11} - \left(\bar{A}_{22d} - \bar{B}_{22} D_2^{-1} \right) D_2 d \\ \bar{e}_{11} = col\left(\bar{e}_1, \bar{e}_{21} \right) \end{cases} \tag{5.24}$$

then, as soon as the sliding mode is established in finite time in eq. (5.22) in the sliding surface eq. (5.23) by means of the injection term eq. (5.18) with $\rho = m_0 k_d + \|D_2\|_\infty l_d$, the attack d is asymptotically estimated as

$$\hat{d} = G^*(s)\upsilon_{eq} \tag{5.25}$$

where

$$G^*(s) = C^*\left(sI - A^*\right)^{-1} B^* \tag{5.26}$$

$$C^* = \begin{bmatrix} \mathbf{0}_{m\times(n-m)} & -D_2^{-1} \end{bmatrix} \tag{5.27}$$

$$B^* = \begin{bmatrix} \mathbf{0}_{(n-p)\times m} \\ \mathbf{0}_{(p-m)\times m} \\ I_{m\times m} \end{bmatrix} \tag{5.28}$$

and υ_{eq} is an *equivalent* injection term.

Proof of Theorem 5.1 is given in the Appendix.

Remark 5.1: The *dynamic extension* of the equivalent control υ_{eq} is proposed as in eq. (5.25) to find the exact reconstruction of the time varying attack $d(t)$. Additionally, unlike the SMO developed in [91], where the attack term $d(t)$ is assumed slow varying ($\dot{d}(t) \approx 0$), in this work it is assumed that $\dot{d}(t) \ne 0$, This is the major novelty of the proposed attack reconstruction algorithm in eq. (5.25).

Remark 5.2: 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 eq. (5.18) [98]. Therefore, if $\bar{\upsilon}_{eq}$ satisfies

$$\tau\dot{\bar{\upsilon}}_{eq} = -(\rho+\eta)\frac{\bar{y}_2-\bar{z}_{22}}{\|\bar{y}_2-\bar{z}_{22}\|}-\bar{\upsilon}_{eq} \tag{5.29}$$

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

$$\|\bar{\upsilon}_{eq}-\upsilon_{eq}\| \sim \mathrm{O}(\tau) \tag{5.30}$$

Therefore, the υ_{eq} estimation error in eq. (5.30) is small, for a small enough choice of τ [91].

Replacing υ_{eq} by $\bar{\upsilon}_{eq}$ in eq. (5.25) gives

$$\bar{\hat{d}} = G^*(s)\bar{\upsilon}_{eq}\,, \tag{5.31}$$

Therefore, the attack estimation error after a finite-time transient can be computed as

$$\left\|\bar{\hat{d}}-\hat{d}\right\| \le \|G^*(s)\|\|\bar{\upsilon}_{eq}-\upsilon_{eq}\| \sim \mathrm{O}(\tau) \quad . \tag{5.32}$$

Remark 5.3: In a case when the filter transfer function $G^*(s)$ is a regular one and represents a low pass filter, additional filtering eq. (5.29) of υ in eq. (5.18) in order to estimate υ_{eq} in eq. (5.31) may be not needed, and $\hat{d}$ can be obtained as $\hat{d} = G^*(s)\upsilon$, since υ_{eq} will be estimated by filtering of υ by the low pass filter with the transfer function $G^*(s)$.

5.1.4 Adaptive-gain Sliding Mode Observer Design

In eq. (5.24), it was assumed that the perturbations term φ is locally norm-bounded, and ρ in the injection term eq. (5.18) is known. In many practical cases the boundary of attacks are unknown, and the gain of the sliding mode injection term eq. (5.18) in the fixed gain observer in eq. (5.16) can be overestimated. The gain overestimation could increase chattering that is difficult to attenuate. Additionally, considering known bounds for attacks is meaningless in many cases. The constant gain ρ in eq. (5.18) can be replaced by the adaptive gain $\rho(t)$ by means of applying the *dual layer nested adaptive sliding mode observation algorithm* [99] i.e.

$$\upsilon = -\left(\rho(t)+\eta\right)\frac{\bar{y}_2-\bar{z}_{22}}{\left\|\bar{y}_2-\bar{z}_{22}\right\|} \tag{5.33}$$

A sufficient condition to ensure sliding on $e_{y_2} = \mathbf{0}$ in finite time is

$$\rho(t) > \left\|A_{33}^{s}e_{y_2}+\phi+D_2\dot{d}\right\| \tag{5.34}$$

An error signal is defined as

$$\sigma(t) = \rho(t) - \frac{1}{\alpha}\left\|\bar{\upsilon}_{eq}(t)\right\| - \varepsilon \tag{5.35}$$

where the scalars $0<\alpha<1,\ \varepsilon>0$.

The adaptation dynamics of $\rho(t)$ in (5.33) is defined as [99]

$$\dot{\rho}(t) = -r(t)sign(\sigma(t)) \tag{5.36}$$

where $r(t)>0$ is a time-varying scalar which satisfies an adaptive scheme. It is assumed that $r(t)$ has the structure

$$r(t) = \ell_0 + \ell(t) \tag{5.37}$$

where ℓ_0 is a fixed positive scalar. The evolution of $\ell(t)$ is to satisfy an adaptive law [30]

$$\dot{\ell}(t) = \begin{cases} \gamma\left|\sigma(t)\right| & if \left|\sigma(t)\right| > \sigma_0 \\ 0 & otherwise \end{cases} \tag{5.38}$$

where $\gamma > 0, \sigma_0 > 0$ are design scalars. The second main results are summarized in the following Proposition 5.1.

Proposition 5.1: Consider the system in (5.22), with

$$a(t) = A_{33}^s e_{y_2} + \phi + D_2 \dot{d} \tag{5.39}$$

and assume that $\left|a(t)\right| < a_0$ and $\left|\dot{a}(t)\right| < a_1$ where a_0 and a_1 are finite but unknown. A SMO is designed as in (5.16) with the *adaptive* injection term in eqs. (5.33) - (5.38). If $\varepsilon > 0$ in eq. (5.35) is chosen to satisfy

$$\frac{1}{4}\varepsilon^2 > \sigma_0^2 + \frac{1}{\gamma}\left(\frac{qa_1}{\alpha}\right)^2 \tag{5.40}$$

for any given σ_0, $q > 1$, and, $0 < \alpha < 1$, then

- the injection term eq. (5.33) exploiting the *dual layer adaptive* scheme given by eqs. (5.34) - (5.38), 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. And, the gains $r(t)$ and $\rho(t)$ remain bounded;

- the sensor attack signal $d(t)$ is reconstructed as in eq. (5.25) with the equivalent adaptive injection term υ_{eq} or $\bar{\upsilon}_{eq}$.

The proof of the Proposition 5.1 is presented in the Appendix.

Remark 5.4: The proposed unit vector injection gain-adaptation algorithm in eq.

(5.33)-(5.38) does not require the knowledge of the boundaries $k_d, l_d > 0$ in $\|d\| < k_d$ and $\|\dot{d}\| < l_d$.

5.2 Line by Line Super Twisting Sliding Mode Observer

In this section, the problem of on-line secure state estimation and attack reconstruction in the face of an offensive that corrupts the sensor measurements and perturbs the states of cyber-physical systems, are investigated. The states of a cyber-physical system and state attacks are reconstructed on-line using a novel adaptive line-by-line super-twisting observer.

5.2.1 Introduction

Consider the linearized format of the CPS model under the state and sensor attack in eq. (2.6) when the number of sensors is greater than the number of sensor attacks. That is

$$\begin{aligned} \dot{x} &= Ax + Bd_x(t) \\ y &= Cx + Dd_y(t) \end{aligned} \quad , \quad p > q - q_1 \tag{5.41}$$

where $x \in \mathbb{R}^n$ presents the state vector of CPS, and $y \in \mathbb{R}^p$ denotes the sensor measurement vector. The $d_x(t) \in \mathbb{R}^{q_1}$ and $d_y(t) \in \mathbb{R}^{q-q_1}$ are the state and sensor attack respectively.

The objective is to estimate the states $x(t)$ and reconstruct the state/plant attack $d_x(t)$, and sensor attack $d_y(t)$ on linear CPS eq. (5.41).

Since $p > q - q_1$, there exists a nonsingular transformation $M \in \mathbb{R}^{p \times p}$ so that

$$MD = \begin{bmatrix} \mathbf{0}_1 \\ D_1 \end{bmatrix} \tag{5.42}$$

where $\mathbf{0}_1 \in \mathbb{R}^{p_1 \times (q-q_1)}$, $D_1 \in \mathbb{R}^{(p-p_1)\times(q-q_1)}$, and

$$p - p_1 \leq q - q_1 \tag{5.43}$$

Therefore, linear CPS eq. (5.41) can be presented in a partitioned format in accordance with eq. (5.42) where it is required that

$$\begin{aligned} \dot{x} &= Ax + Bd_x(t) \\ y_1 &= C_1 x \\ y_2 &= C_2 x + D_1 d_y(t) \end{aligned} \tag{5.44}$$

where $C_1 \in \mathbb{R}^{p_1 \times n}$, $C_2 \in \mathbb{R}^{(p-p_1)\times n}$ are the partitions of matrix MC.

Note that

- $y_1 \in \mathbb{R}^{p_1}$ and $y_2 \in \mathbb{R}^{p-p_1}$ are the partitions of the transformed measurement vector My. Therefore, they are called the virtual protected measurement and the virtual unprotected measurement respectfully.
- $d_x \in \mathbb{R}^{q_1}$ represents the state attack which may comprise of the plant attacks and/or sensor attacks that get propagated to the plant through feedback control.

Remark 5.5: Note that eq. (5.43) means that the number of sensor attacks is greater than or equal to the number of virtual unprotected sensor measurements.

The problem of online reconstruction of the attack signals, i.e. $\hat{d}_x(t) \to d_x(t)$, $\hat{d}_y(t) \to d_y(t)$, is addressed in two steps: first, using the protected measurement y_1 and applying the novel adaptive line-by-line STW disturbance observer, the plant states $x(t)$ and the state attacks $d_x(t)$ are reconstructed. Then, using the on-line estimated states $\hat{x}(t)$, and corrupted measurements y_2 and by applying a SR algorithm discussed in Section (3.1), sensor attacks $\hat{d}_y(t)$ are reconstructed.

Remark 5.6. Note that the reconstructed sensor attacks $\hat{d}_x(t)$ are used for cleaning up the measurements for preventing the propagation of these attacks to the attacked plant through the feedback control. The reconstructed state attacks $\hat{d}_x(t)$ are assumed to be matched to the control and can be included in the feedback controller for compensating the state attacks.

Assumption (A 5.5): The number of virtual protected measurements is equal or greater than the number of state attacks, i.e.

$$q_1 \leq p_1 \tag{5.45}$$

5.2.2 State Attack Reconstruction

Consider the linear CPS eq. (5.44) and assume that we have $p_1 = q_1$ virtual protected sensors so that

$$\begin{aligned} &\dot{x} = Ax + Bd_x(t) \\ &y_1 = \begin{bmatrix} y_{11} & y_{12} & ,..., & y_{1q_1} \end{bmatrix}^T = C_1 x, \quad y_{1i} = C_{1i} x \end{aligned} \tag{5.46}$$

where $y_{1i} \in \mathbb{R}$ and C_{1i} is the i^{th} row of matrix C_1 for $i = 1,...,q_1$.

Assume that assumptions (A 5.5) and (A 3.1) are verified for linear CPS eq. (5.47) and $r = \{r_1, r_2, ..., r_{q_1}\}$ gives the input-output vector relative degree of linear CPS (5.47) as it is defined at (A 3.1)

Consider the following SMO

$$\dot{\hat{x}} = A\hat{x} + G_l\left(y_{1a} - C_{1a}\hat{x}\right) + G_n \upsilon_c(y_{1a} - C_{1a}\hat{x}) \tag{5.47}$$

where the matrices $G_l \in \mathbb{R}^{n\times r_s}$ and $G_n \in \mathbb{R}^{n\times r_s}$ are of appropriate dimension and are to be designed. The auxiliary output y_{1a} which contains both real and synthetic measurements and the matrix C_{1a} are defined as follows

$$y_{1a} = \begin{bmatrix} y_{11} \\ \nu\left(y_{11} - y_{11}^{1}\right) \\ \vdots \\ \nu\left(\tilde{y}_{11}^{r_{\alpha_1}-1} - y_{11}^{r_{\alpha_1}-1}\right) \\ \vdots \\ y_{1q_1} \\ \vdots \\ \nu\left(\tilde{y}_{1q_1}^{r_{\alpha_1}-1} - y_{1q_1}^{r_{\alpha_1}-1}\right) \end{bmatrix}, \quad C_{1a} = \begin{bmatrix} C_{11} \\ \vdots \\ C_{11}A^{r_{\alpha 1}-1} \\ \vdots \\ C_{1q_1} \\ \vdots \\ C_{1q_1}A^{r_{\alpha q_1}-1} \end{bmatrix} \tag{5.48}$$

where $\upsilon_c(.)$ is the injection vector

$$\upsilon_c(y_{1a} - C_{1a}\hat{x}) = \begin{cases} -\left(\kappa + \eta_0\right)\dfrac{P\left(y_{1a} - C_{1a}\hat{x}\right)}{\left\|P\left(y_{1a} - C_{1a}\hat{x}\right)\right\|} & \mathit{if}\ (y_{1a} - C_{1a}\hat{x}) \neq 0 \\ 0 & \mathit{otherwise} \end{cases} \tag{5.49}$$

where η_0 is a small positive constant and κ is a positive constant suitably larger than the upper bound of attack d_x. The positive definite matrix P can be found by solving a corresponding Lyapunov equation [92].

The continuous injection term $\nu(.)$ is given by the STW algorithm [90]

$$\begin{aligned} \nu(s_i^j) &= \varphi(s_i^j) + \lambda_i^j \left|s_i^j\right|^{1/2} sign(s_i^j) \\ \dot{\varphi}(s_i^j) &= \beta_i^j sign(s_i^j), \quad \lambda_i^j, \beta_i^j > 0 \end{aligned} \tag{5.60}$$

where $\lambda_i^j \in \mathbb{R}$ and $\beta_i^j \in \mathbb{R}$ are suitably chosen gains and the $s_i^j \in \mathbb{R}$ for $i = 1,...,q_1$

and $j=1,\ldots,r_{i+1}$ are the sliding variables where

$$\begin{aligned} s_i^1 &= y_{1i} - \hat{y}_{1i} \\ s_i^j &= \tilde{y}_{1i}^{j-1} - \hat{y}_{1i}^{j-1}, \quad for \quad j=2,\ldots,r_{i+1} \end{aligned} \tag{5.61}$$

Then it is assumed $\left|\bar{y}_{1i}^{j+1}\right| \le L_i^j$ for $j=1,\ldots,r_{i-1}$, and $\left|\bar{C}_{1i}A^{r_i}x + \bar{C}_{1i}A^{r_i-1}\bar{B}d_x(t)\right| \le L_i^{r_i}$ where L_i^j's are fixed and known.

It is shown [91] that with λ_i^j and β_i^j chosen as [21]

$$\lambda_i^j = 1.5\sqrt{L_i^j}, \quad \beta_i^j = 1.1L_i^j \tag{5.62}$$

Remark 5.7. The values $L_i^j > 0$ are difficult to predict.

Overestimating $L_i^j > 0$ may lead to the gains λ_i^j and β_i^j being overestimated, and therefore to increase chattering. The adaptive version of the attack estimation algorithm with non-overestimated gains is discussed in next section.

The presented results can be summarized as

Proposition 5.2 [89]. The states x of linear CPS eq. (5.46) are estimated asymptotically using the SMO in eqs. (5.47) – (5.62), while the state attack d_x is estimated asymptotically as

$$\hat{d}_x = \left((C_{1a}B)^T C_{1a}B\right)^{-1}(C_{1a}B)^T C_{1a}G_n(\upsilon_c)_{eq} \tag{5.63}$$

5.2.3 Novel Adaptive Gain State Attack Line-by-Line Super Twisting Observer

The state attack line-by-line STW Observer discussed in section 5.2.2 and 3.2 suffer from the problem of overestimating the unknown bounds $L_i^j > 0$ as mentioned in Remark 3.1. In this section, it is proposed to augment the line-by-line STW observer in eqs. (5.47)

– (5.62) by a gain adaptation algorithm with non-overestimated gains.

In accordance with [100], the STW algorithm in (5.60) that drives $s_i^j, \dot{s}_i^j \to 0$ in finite time is augmented as

$$\begin{aligned} v(s_i^j) &= \varphi(s_i^j) + \lambda_i^j(t)\left|s_i^j\right|^{1/2} sign(s_i^j) - \phi\left(s_i^j, L_i^j\right) \\ \dot{\varphi}(s_i^j) &= \beta_i^j(t) sign(s_i^j) \end{aligned} \tag{5.64}$$

with the adaptive gains

$$\lambda_i^j(t) = \sqrt{L_i^j(t)}\lambda_0, \quad \beta_i^j(t) = L_i^j(t)\beta_0, \tag{5.65}$$

where the new term is defined as

$$\phi\left(s_i^j, L_i^j\right) = -\frac{\dot{L}_i^j(t)}{L_i^j(t)} s_i^j, \tag{5.66}$$

and λ_0, β_0 are fixed positive scalars satisfying

$$\begin{aligned} & PA_0 + A_0^T P + \varepsilon_0 P + PB_0 B_0^T P + C_0^T C_0 < 0 \\ & A_0 = \begin{bmatrix} -\frac{1}{2}\alpha_0 & \frac{1}{2} \\ -\beta_0 & 0 \end{bmatrix}, B_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, C_0 = \begin{bmatrix} 1 & 0 \end{bmatrix} \quad . \\ & P = \begin{bmatrix} p_1 & p_2 \\ p_2 & p_3 \end{bmatrix}, \quad p_1, p_3 > 0, \quad p_2^2 < p_1 p_3, \quad \varepsilon_0 > 0 \end{aligned} \tag{5.67}$$

The double layer adaptive gain law is formulated in [100] by

$$L_i^j(t) = \ell_i^j(t) + \ell_0 \tag{5.68}$$

for $i = 1,\ldots,q_1$ and $j = 1,\ldots,r_i$ where

$$\begin{aligned} \dot{\ell}_i^j(t) &= -\rho(t) sign\left(\xi_i^j(t)\right) \\ \dot{\rho}(t) &= \gamma\left|\xi_i^j(t)\right| \end{aligned} \tag{5.69}$$

$$\xi_i^j(t) = L_i^j(t) - \frac{1}{a\beta_0}\left|\bar{u}_{eq_i}{}^j(t)\right| - \varepsilon_1 \tag{5.70}$$

where β_0 and a are fixed positive scalars satisfy $0 < a < 1/\beta_0 < 1$ and ε_1 is a small positive scalar. The equivalent injection signal $\bar{u}_{eq}(t)$ can be approximated in real-time by low pass filtering of the switched signal i.e.

$$\dot{\bar{u}}_{eq_i}{}^{j}(t) = \frac{1}{\tau}\left(\beta_i^j(t) sign(s_i^j(t)) - \bar{u}_{eq_i}{}^{j}(t)\right) \tag{5.71}$$

where τ is a small positive constant.

Remark 5.8. The adaptive gains in (5.65) are non-overestimated [100].

The results are summarized in the following Proposition.

Proposition 5.3. The states x of linear CPS eq. (5.46) are estimated asymptotically in using the SMO in eq. (5.47) – (5.62) and the **adaptive** STW injection terms in eqs. (5.64)-(5.71), while the state attack d_x in eq. (5.64) are estimated asymptotically as

$$\hat{d}_x = \left((C_{1a}B)^T C_{1a}B\right)^{-1}(C_{1a}B)^T C_{1a}G_n(\upsilon_c)_{eq} \tag{5.72}$$

Remark 5.9. The novel observer design in eq. (5.72) makes $\hat{d}_x$ estimation possible with unknown boundaries $L_i^j > 0$.

Remark 5.10. Note that the Adaptive *Dual Layer Unit Vector Observer* technique [99] can be applied to the unit vector observer in (5.49) as well. This will provide a self-tuning capability to the observer (5.47), while the adaptive gain of the injection term will not be overestimated. The novel adaptive SMO that comprises the adaptive line-by-line STW algorithm (5.64) - (5.70) and the adaptive unit vector injection terms is capable of reconstructing both the states and state attacks without knowing the attack boundaries.

This novel result was submitted to the American Control Conference 2020.

5.2.4 Sensor attacks reconstruction

Consider the corrupted sensors of the system (5.46) as

$$y_2 = C_2 x + D_1 d_y(t) \tag{5.73}$$

where $y_2 \in \mathbb{R}^{p-q_1}, D_2 \in \mathbb{R}^{(p-q_1)\times(q-q_1)}, d_y(t) \in \mathbb{R}^{(q-q_1)}$. Note that, according to eq. (5.43), the number of sensor attacks is equal or greater than the number of virtual unprotected measurements, $p - p_1 \leq q - q_1$.

Using $\hat{x}$ estimated by the SMO (5.47), it follows

$$D_1 d_y(t) = y_2 - C_2 \hat{x} \tag{5.74}$$

Note that if

1. the number of attack signals and the number of corrupted sensors are the same: $q - q_1 = p - q_1$, and D_1 is invertible, then the sensor attacks is reconstructed easily as

$$d_y(t) = \left(D_2\right)^{-1}\left(y_2 - C_2 \hat{x}\right) \tag{5.75}$$

2. the number of attack signals are greater than the number of corrupted sensors, i.e. $q - q_1 > \bar{p} - q_1$, then a sparse vector $d_y(t)$ can be reconstructed.

Assumption (A 5.6): The sensor attack $d_y(t)$ is a j-sparse vector. i.e. There is only a limited number, j, of the non-zero sensor attacks $d_y \in \mathbb{R}^{q-q_1}$ at any time instant. Specifically, the index set of non-zero sensor attacks is presented as $\Phi_\Gamma = \{k_1, k_2, \ldots k_j\}$, $j \leq q - q_1$, where

$$2j + 1 \leq p - p_1 \tag{5.76}$$

Assumption (A 5.7): Matrix D_1 satisfies the RIP condition in definition 3.1.

The sensor attack $d_y(t)$ in (5.74) is reconstructed using the SR Algorithm presented in Section 3.1 as

$$\mu\dot{v}(t) = -\left\lceil v(t) + \left(D_1^T D_1 - I_{(q-q_1)\times(q-q_1)}\right)a(t) - D_1^T \gamma \right\rfloor^{\beta}$$

$$\hat{d}_y(t) = a(t) \tag{5.77}$$

where $v \in \mathbb{R}^q$ is the state vector, $\hat{d}_y(t)$ represents the estimate of the sparse sensor attack $d_y(t)$, $\mu > 0$ is a time-constant determined by the physical properties of the implementing system.

Note that $\lceil . \rfloor^{\beta} = |.|^{\beta} sign(.)$ and $a(t) = H_{\lambda}(v)$ where $H_{\lambda}(.)$ is defined as eq. (3.6), that is

$$H_{\lambda}(v) = \max\left(|v| - \lambda, 0\right)\mathrm{sgn}(v)$$

where $\lambda > 0$ is chosen with respect to the noise and the minimum absolute value of the nonzero terms.

Under assumption (A 5.7), the state v of eq. (5.77) converges in finite time to its equilibrium point v^*, and sensor attack estimation $\hat{d}_y(t)$ in eq. (5.77) converges in finite-time to sensor attack $d_y(t)$ in CPS eq. (5.44).

5.3 **Summary**

In this chapter, linearized cyber-physical systems under attacks are investigated when the number of sensors is greater than the number of potential attacks. Two approach are proposed to estimate the states and reconstruct the attacks.

At first, a novel SMO is proposed. The attacks are reconstructed on-line by using a filter that uses dynamic extension of the injection term and is proposed for the first time.

Next, an adaptive line-by-line STW SMO is developed to estimate the states of CPSs and reconstruct the plant attacks without any overestimated gain.

In next chapter, state estimation and attack reconstruction in nonlinear CPSs are investigated.

CHAPTER 6

Attack Reconstruction in Nonlinear Systems: the Number of Potential Attacks Is Greater Than the Number of Sensors

6.1 Introduction

The nonlinear CPS in eq. (2.10) is considered when the number of potential attacks are greater than the number of sensors, i.e.

$$\begin{aligned} \dot{x} &= f(x) + B(x)d(t) \\ y &= C(x) + Dd(t) \end{aligned} \quad where \quad q > p \tag{6.1}$$

Assumption (A 6.1) It is assumed that the attack vector $d(t)$ is sparse, meaning that numerous attacks are possible, but the attacks are not coordinated, and only few non-zero attacks happen at the same time. i.e. the index set of non-zero attacks is presented as $\Phi_{\Gamma} = \left\{k_1, k_2, \ldots k_j\right\}$, $j < q$, where

$$2j + 1 \leq p \tag{6.2}$$

The objective of this chapter is to reconstruct on-line the time varying attack sparse vector based on the sensor measurement y in CPS eq. (6.1).

6.2 System transformation

Feeding the sensor measurements under attack, y, of the CPS eq. (6.1) to the input of the low pass filter that facilitates filtering out the possible measurement noise gives

$$\dot{z} = \frac{1}{\tau}\left(-z + C(x) + D(x)d(t)\right) \tag{6.3}$$

whose output $z \in \mathbb{R}^p$, is available. Then, the CPS in eq. (6.1) is rewritten as

$$\begin{cases} \dot{\xi} = \eta(\xi) + \Omega d(t) \\ \psi = C\xi \end{cases} \tag{6.4}$$

where $\psi \in \mathbb{R}^p$, and

$$\begin{aligned} &\xi = \begin{bmatrix} z \\ x \end{bmatrix}_{(p+n)\times 1}, \quad \eta(\xi) = \begin{bmatrix} -\frac{1}{\tau}I & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} z \\ x \end{bmatrix} + \begin{bmatrix} \frac{1}{\tau}C(x) \\ f(x) \end{bmatrix} \\ &\Omega = \begin{bmatrix} \frac{1}{\tau}D(x) \\ B(x) \end{bmatrix} = \begin{bmatrix} \Omega_1 & \Omega_2 & \dots & \Omega_q \end{bmatrix}_{(p+n)\times q} \\ &C = \begin{bmatrix} C_1 & C_2 & \dots & C_{p+n} \end{bmatrix} = \begin{bmatrix} I_{p\times p} & 0_{p\times n} \end{bmatrix} \end{aligned} \tag{6.5}$$

Assumption (A 6.2) The transformed CPS in eq. (6.4) is assumed to have a vector relative degree $r = \{r_1, r_2, ..., r_p\}$, i.e.

$$\begin{aligned} &L_{\Omega_j} L_\eta^\lambda \psi_i(\xi) = 0 \quad \forall j = 1,..,q \ \ \forall \lambda < r_i - 1 \ \ \forall i = 1,...,p \\ &L_{\Omega_j} L_\eta^{r_i - 1} \psi_i(\xi) \neq 0 \quad \text{for at least one } 1 \leq j \leq q \end{aligned} \tag{6.6}$$

Assuming that the assumptions (A 3.4) and (A 3.5) in Section (3.3) are satisfied for system (6.4), then input-output dynamics of system (6.4) are presented as

$$\dot{\Upsilon}_i = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \Upsilon_i + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ L_f^{r_i}\psi_i(\xi) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \sum_{j=1}^{q} L_{\Omega_j} L_f^{r_i - 1}\psi_i(\xi) d_j \end{bmatrix}, \tag{6.7}$$

where

$$\Upsilon_i = \begin{bmatrix} \Upsilon_1^i(\xi) \\ \Upsilon_2^i(\xi) \\ \vdots \\ \Upsilon_{r_i}^i(\xi) \end{bmatrix} = \begin{bmatrix} \psi_i(\xi) \\ L_f \psi_i(\xi) \\ \\ L_f^{r_i-1}\psi_i(\xi) \end{bmatrix} \quad for \quad i = 1,\ldots,p \tag{6.8}$$

where $\psi_i(\xi)$ is the i^{th} entry of vector $\psi(\xi)$. Each of system output ψ_i at its own relative degree r_i, satisfies following equation

$$\dot{\Upsilon}_{r_i}^i(\xi) = L_f^{r_i}\psi_i(\xi) + \sum_{j=1}^{\alpha} L_{\Omega_j} L_f^{r_i-1}\psi_i d_j, \quad i = 1,\ldots,m\,. \tag{6.9}$$

Therefore, system eq. (6.4) can be rewritten as the following algebraic equation

$$Z_p = F(\xi) d(t), \tag{6.10}$$

where

$$Z_p = \begin{bmatrix} \dot{\Upsilon}_{r_1}^1 \\ \vdots \\ \dot{\Upsilon}_{r_p}^p \end{bmatrix} - \begin{bmatrix} L_f^{r_1}\psi_1(\xi) \\ \vdots \\ L_f^{r_p}\psi_p(\xi) \end{bmatrix}, \tag{6.11}$$

where $Z_p \in \mathbb{R}^p$, $F(\xi) \in \mathbb{R}^{p \times q}$, and

$$F(\xi) = \begin{bmatrix} L_{\Omega_1} L_f^{r_1-1}\psi_1 & L_{\Omega_2} L_f^{r_1-1}\psi_1 & \cdots & L_{\Omega_\alpha} L_f^{r_1-1}\psi_1 \\ L_{\Omega_1} L_f^{r_2-1}\psi_2 & L_{\Omega_2} L_f^{r_2-1}\psi_2 & & L_{\Omega_\alpha} L_f^{r_2-1}\psi_2 \\ \vdots & & & \vdots \\ L_{\Omega_1} L_f^{r_p-1}\psi_p & L_{\Omega_2} L_f^{r_p-1}\psi_p & \cdots & L_{\Omega_q} L_f^{r_p-1}\psi_p \end{bmatrix}. \tag{6.12}$$

Remark 6.1: The derivative $\dot{\Upsilon}_{r_1}^1,\ldots,\dot{\Upsilon}_{r_p}^p$ are computed exactly in finite time using higher order sliding mode differentiators [28] discussed in eqs. (3.30) and (3.31).

6.3 Sparse Recovery Algorithm

Assumption (A 6.3): The matrix $F(\xi)$ is supposed to satisfy the RIP condition in definition 3.1.

The attack $d(t)$ in (6.10) is reconstructed using the SR Algorithm presented in Section 3.1 as

$$\mu \dot{v}(t) = -\left\lceil v(t) + \left(F(\xi)^{T} F(\xi) - I_{q\times q}\right) a(t) \; - F(\xi)^{T} \gamma \right\rfloor^{\beta}$$
$$\hat{d}(t) = a(t) \qquad (6.13)$$

where $v \in \mathbb{R}^{q}$ is the state vector, $\hat{d}(t)$ represents the estimate of the sparse signal $d(t)$ of eq. (6.10), and $\mu > 0$ is a time-constant determined by the physical properties of the implementing system.

Note that $\lceil . \rfloor^{\beta} = |.|^{\beta}\, sign(.)$ and $a(t) = H_{\lambda}(v)$ where $H_{\lambda}(.)$ is a continuous soft thresholding function and defined as eq. (3.6), that is

$$H_{\lambda}(v) = \max\left(|v| - \lambda, 0\right) \mathrm{sgn}(v)$$

where $\lambda > 0$ is chosen with respect to the noise and the minimum absolute value of the nonzero terms.

Under Assumption (A 6.3), the state v of eq. (6.13) converges in finite time to its equilibrium point v^{*}, and $\hat{d}(t)$ in eq. (6.13) converges in finite-time to $\hat{d}(t)$ of eq. (6.10).

The results of this chapter have been published in [79].

6.4 **Summary**

In this chapter, considering nonlinear cyber-physical systems when the number of potential attacks is greater than the number of sensor measurements, attacks are reconstructed using higher order sliding mode differentiation techniques in concert with the SR algorithm, when only several unknown attacks out of all possible attacks are non-zero. The relative degree approach and SR algorithm has been applied to the system while recovering the attacks on the sensors and the plant.

In the next chapter, state and sensor attack reconstruction in nonlinear CPSs is investigated when the number of sensor measurements is greater than the number of potential attacks.

CHAPTER 7

Attack Reconstruction in Nonlinear Systems: the Number of Sensors is Greater than the Number of Potential Attacks

7.1 Introduction

Consider the nonlinear CPS model under the state and sensor attack in eq. (2.6) when the number of sensors is greater than the number of sensor attacks, that is

$$\begin{matrix} \dot{x} = f(x) + B_1(x) d_x(t) \\ y = C(x) + D_1 d_y(t) \end{matrix} \quad , \quad p > q - q_1 \tag{7.1}$$

where $y \in \mathbb{R}^p$, $d_x(t) \in \mathbb{R}^{q_1}$ and $d_y(t) \in \mathbb{R}^{q-q_1}$.

The objective in this chapter is to reconstruct the state attack $d_x(t)$, and sensor attack $d_y(t)$ in the nonlinear CPS eq. (7.1).

Since there are more sensors than potential sensor attacks in CPS eq. (7.1), $p > q - q_1$, there exists a nonsingular output transformation $M \in \mathbb{R}^{p \times p}$ so that

$$\bar{y} = M^{-1} y = M^{-1} C(x) + M^{-1} D_1 d_y \tag{7.2}$$

where the matrix M is selected to satisfy the condition

$$M^{-1} D_1 = \begin{bmatrix} \mathbf{0}_3 \\ D_2 \end{bmatrix} \tag{7.3}$$

where $\mathbf{0}_3 \in \mathbb{R}^{p_1 \times (q-q_1)}$, $D_2 \in \mathbb{R}^{(p-p_1) \times (q-q_1)}$, and $p - p_1 \leq q - q_1$.

The transformed sensor measurement vector $\bar{y}$ in eq. (7.2) is partitioned as

$$\bar{y} = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \end{bmatrix} \tag{7.4}$$

where $\bar{y}_1 \in \mathbb{R}^{p_1}$, $\bar{y}_2 \in \mathbb{R}^{p-p_1}$.

Next, CPS (7.1) is presented in a partitioned format in accordance with eqs. (7.1), (7.4) as

$$\begin{aligned} \dot{x} &= f(x) + B_1(x)d_x(t) \\ \bar{y}_1 &= C_1(x) \\ \bar{y}_2 &= C_2(x) + D_2 d_y(t) \end{aligned} \tag{7.5}$$

where $C_1 \in \mathbb{R}^{p_1}$ and $C_2 \in \mathbb{R}^{p-p_1}$.

Remark 7.1. The virtual measurement $\bar{y}_1$ in eq. (7.5) is not affected by the attack corruption signal and can be classified as a *protected* measurement.

Assumption (A 7.1): The number of protected measurements is equal or greater than the number of plant attacks, i.e.

$$p_1 \geq q_1 \tag{7.6}$$

Remark 7.2. Equation (7.3) gives that the number of unprotected measurements is equal or less than the number of attacks that may corrupt the measurements, i.e.

$$p - p_1 \leq q - q_1 \tag{7.7}$$

The considered problem is: given the nonlinear CPS dynamics in eq. (7.5) with virtual protected $\bar{y}_1 \in \mathbb{R}^{p_1}$ and unprotected $\bar{y}_2 \in \mathbb{R}^{p-p_1}$ sensors, and attack signals $d_x \in \mathbb{R}^{q_1}$ on the plant and $d_y \in \mathbb{R}^{q-q_1}$ on the sensors (sensor corruption signals), reconstruct the attack signals.

The attack reconstruction is to be accomplished in two steps:

Step 1: The plant state $x(t)$ and the attack $d_x(t)$ vectors are estimated by applying the HOSM observer described in Section (3.3) with respect to the protected output $\bar{y}_1$

only, so that

$$\hat{x}(t) \to x(t),\ \hat{d}_x(t) \to d_x(t) \tag{7.8}$$

in finite time, where $\hat{x}(t)$, $\hat{d}_x(t)$ are the estimation of CPS states and the reconstruction of plant attack respectively.

Step 2: Given the state $\hat{x}(t)$, which is estimated on-line, the unprotected sensor attack d_y is then estimated as $\hat{d}_y$ by applying a SR algorithm as it is described in section (3.1) [87].

7.2 **State Attack Reconstruction**

Consider the part of CPS eq. (7.5) associated with the virtual measurements protected from the attacks

$$\begin{aligned} \dot{x} &= f(x) + B_1(x)d_x(t) \\ \overline{y}_1 &= C_1(x) \end{aligned} \tag{7.9}$$

Note that only q_1 out of p_1 virtual protected measurements are employed, and that the other $p_1 - q_1$ virtual protected measurements can be used at the second step of the proposed algorithm. The aforementioned modifications are addressed by defining $\overline{y}_1$ and B_1 in eq. (7.9) are as $\overline{y}_1 = \left[\overline{y}_{11} \quad \overline{y}_{12} \quad ,..., \quad \overline{y}_{1q_1}\right]^T$, $B_1 = \left[b_1, b_2, ..., b_{q_1}\right] \in \mathbb{R}^{n \times q_1}$ where $b_i \in \mathbb{R}^n, \forall i = 1,...,q_1$ are smooth vector-fields defined on an open $\Omega \subset \mathbb{R}^n$.

The problem is to estimate the states of nonlinear CPS eq. (7.9) with unknown input, and reconstruct the state attack vector $d_x(t)$.

Assume that the CPS in eq. (7.9) has the vector relative degree $r = \left\{r_1, r_2, ..., r_{q_1}\right\}$ as it is defined in Assumption (A 3.2) in Chapter 3.

If the CPS in eq. (7.9) satisfies assumptions (A 3.3) and (A 3.4), then, the CPS given by eq. (7.9) with the involutive distribution $\Gamma = span\left\{b_1, b_2, ..., b_{q_1}\right\}$ and total relative degree $r = \sum_{i=1}^{q_1} r_i \leq n$ can be rewritten as

$$\dot{\delta}_i = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{r_i \times r_i} \delta_i + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ L_f^{r_i} \bar{y}_{1_i}(x) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \sum_{j=1}^{m} L_{b_j} L_f^{r_i - 1} \bar{y}_{1_i}(x) d_x(t) \end{bmatrix}, \forall i = 1, ..., q_1 \quad (7.10)$$

$$\dot{\gamma} = g(\delta, \gamma)$$

where

$$\delta = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_{q_1} \end{bmatrix}, \quad \delta_i = \begin{bmatrix} \delta_{i1} \\ \delta_{i2} \\ \vdots \\ \delta_{ir_i} \end{bmatrix} = \begin{bmatrix} \eta_{i1}(x) \\ \eta_{i2}(x) \\ \vdots \\ \eta_{ir_i}(x) \end{bmatrix} = \begin{bmatrix} \bar{y}_{1_i}(x) \\ L_f \bar{y}_{1_i}(x) \\ \vdots \\ L_f^{r_i - 1} \bar{y}_{1_i}(x) \end{bmatrix} \in \mathbb{R}^{r_i} \; \forall i = 1, ..., q_1, \quad \gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_{n-r} \end{bmatrix} = \begin{bmatrix} \eta_{r+1}(x) \\ \eta_{r+2}(x) \\ \vdots \\ \eta_n(x) \end{bmatrix} \quad (7.11)$$

The norm-bounded solution of the internal dynamics $\dot{\gamma} = g(\delta, \gamma)$ is assumed to be locally asymptotically stable [95] as it is mentioned in (A 3.5) in Chapter 3.

The variables $\eta_{r+1}(x), ..., \eta_n(x)$ are defined to satisfy

$$L_{b_j} \eta_i(x) = 0 \quad \forall i = r+1, ..., n, \; \forall j = 1, ..., q_1 \quad (7.12)$$

if assumption (A 3.4) is satisfied then it is always possible to find $n - r$ functions $\eta_{r+1}(x), ..., \eta_n(x)$ such that

$$\Psi(x) = col\left\{\eta_{11}(x), ..., \eta_{1r_1}(x), ..., \eta_{q_1 1}(x), ..., \eta_{q_1 r_{q_1}}(x), \eta_{r+1}(x), ..., \eta_n(x)\right\} \in \mathbb{R}^n \quad (7.13)$$

is a local diffeomorphism in a neighborhood of any point $x \in \bar{\Omega} \subset \Omega \subset \mathbb{R}^n$ which means

$$x = \Psi^{-1}(\delta, \gamma) \quad (7.14)$$

In order to estimate the derivatives $\delta_{ij}(t)$ $\forall i=1,\ldots,q_1$, $\forall j=1,\ldots,r_i$ of the outputs y_i in finite time, higher-order sliding-mode differentiators [96] presented at eqs. (3.30) and (3.31) are used.

The following exact estimates are available in finite time:

$$\hat{\delta}_i=\begin{bmatrix}\hat{\delta}_{i1}\\ \hat{\delta}_{i2}\\ \vdots\\ \hat{\delta}_{ir_1}\end{bmatrix}=\begin{bmatrix}\hat{\eta}_{i1}(\hat{x})\\ \hat{\eta}_{i2}(\hat{x})\\ \vdots\\ \hat{\eta}_{ir_1}(\hat{x})\end{bmatrix}\in\mathbb{R}^{r_i}\ \ \forall i=1,\ldots,q \quad \hat{\delta}=\begin{bmatrix}\hat{\delta}^1\\ \hat{\delta}^2\\ \vdots\\ \hat{\delta}^q\end{bmatrix}\in\mathbb{R}^{r_i} \tag{7.15}$$

Integrating the second equation in eq. (7.10) and replacing δ by $\hat{\delta}$, the internal dynamics is given as

$$\dot{\hat{\gamma}}=g\left(\hat{\delta},\hat{\gamma}\right) \tag{7.16}$$

and with some initial condition from the stability domain of the internal dynamics, a asymptotic estimate $\hat{\gamma}$ can be obtained locally as

$$\hat{\gamma}=\begin{pmatrix}\hat{\gamma}_1\\ \hat{\gamma}_2\\ \vdots\\ \hat{\gamma}_{n-r}\end{pmatrix}=\begin{pmatrix}\hat{\eta}_{r+1}(\hat{x})\\ \hat{\eta}_{r+2}(\hat{x})\\ \vdots\\ \hat{\eta}_n(\hat{x})\end{pmatrix} \tag{7.17}$$

Therefore, the asymptotic estimate for the mapping eq. (7.14) is identified as

$$\Psi(\hat{x})=col\left\{\hat{\eta}_{11}(\hat{x}),\ldots,\hat{\eta}_{1r_1}(\hat{x}),\ldots,\hat{\eta}_{q_1 1}(\hat{x}),\ldots,\hat{\eta}_{q_1 r_{q_1}}(\hat{x}),\hat{\eta}_{r+1}(\hat{x}),\ldots,\hat{\eta}_n(\hat{x})\right\} \tag{7.18}$$

The asymptotic estimate $\hat{x}$ of the state vector x of CPS (7.9) can be easily identified via eqs. (7.15) and (7.17) as

$$\hat{x}=\Psi^{-1}\left(\hat{\delta},\hat{\gamma}\right) \tag{7.19}$$

An asymptotic estimate $\hat{d}_x(t)$ of the cyber state attack $d_x(t)$ in eq. (7.9) can be identified as

$$\hat{d}_x(t) = L^{-1}\left(\Psi^{-1}\left(\hat{\delta},\hat{\gamma}\right)\right)\left[\begin{pmatrix} \dot{\hat{\delta}}_{1r_1} \\ \dot{\hat{\delta}}_{2r_2} \\ \vdots \\ \dot{\hat{\delta}}_{qr_q} \end{pmatrix} - \begin{pmatrix} L_f^{r_1} y_1\left(\Psi^{-1}\left(\hat{\delta},\hat{\gamma}\right)\right) \\ L_f^{r_2} y_2\left(\Psi^{-1}\left(\hat{\delta},\hat{\gamma}\right)\right) \\ \vdots \\ L_f^{r_q} y_q\left(\Psi^{-1}\left(\hat{\delta},\hat{\gamma}\right)\right) \end{pmatrix}\right] \tag{7.20}$$

where $L\left(\Psi^{-1}\left(\hat{\delta},\hat{\gamma}\right)\right) = \sum_{j=1}^{q} L_{b_j} L_f^{\,r_i-1} \bar{y}_{1_i}(x)$.

7.3 **Sensor Attacks Reconstruction**

After the state vector $x(t)$ and the plant attack $d_x(t)$ of CPS eq. (7.1) are reconstructed in eqs. (7.19) and (7.20), then, the sensor attacks $d_y(t)$ can be reconstructed as the following discussion:

Consider the attacked part of system eq. (7.5) as

$$\begin{aligned} \dot{x} &= f(x) + B_1(x)d_x(t) \\ \bar{y}_2 &= C_2(x) + D_2 d_y(t) \end{aligned} \tag{7.21}$$

where $y_2 \in \mathbb{R}^{p-q_1}$, $D_2 \in \mathbb{R}^{(p-q_1)\times(q-q_1)}$, $d_y(t) \in \mathbb{R}^{(q-q_1)}$.

Two cases that cover all possible situations are considered to reconstruct the sensor attack $d_y(t)$.

Case 1: If the number of sensor attacks and the number of corrupted sensors is the same, i.e. $p - q_1 = q - q_1$, and D_2 is invertible, then using $\hat{x}$ estimated by the SMO in eq. (7.19), there is a unique solution for estimation of sensor attack as

$$\hat{d}_y(t) = D_2^{-1}\left(y_2 - C_2(\hat{x})\right) \tag{7.22}$$

Case 2: If the number of sensor attacks is greater than the number of corrupted sensors, i.e. $p - q_1 < q - q_1$ and the following assumption is verified for sensor attack d_y .

Assumption (A 7.2): The attacks $d_y \in \mathbb{R}^{q-q_1}$ on the unprotected sensors $\bar{y}_2 \in \mathbb{R}^{p-p_1}$ are assumed not to be coordinated, meaning that there is only a small number of non-zero attacks at any point in time, i.e. the index set of non-zero attacks is presented as $\Phi_\Gamma = \{k_1, k_2, \ldots k_j\}$, $j \leq q - q_1$, where

$$2j + 1 \leq p - p_1 \tag{7.23}$$

Considering the corrupted sensor measurements dynamic of CPS (7.9) and using $\hat{x}$ estimated by the SMO in eq. (7.19), it is given that

$$\bar{y}_2 - C_2(\hat{x}) = D_2 d_y(t) \tag{7.24}$$

Assumption (A 7.3): Matrix D_2 satisfies the RIP condition in definition 3.1.

The attack $d(t)$ in (6.10) is reconstructed using the SR algorithm presented in Section 3.1 as

$$\begin{aligned} \mu \dot{v}(t) &= -\left\lceil v(t) + \left(D_2^T D_2 - I_{(q-q_1)\times(q-q_1)}\right) a(t) \ - D_2^T \gamma \right\rfloor^{\beta} \\ \hat{d}_y(t) &= a(t) \end{aligned} \tag{7.25}$$

where $v \in \mathbb{R}^q$ is the state vector, $\hat{d}_y(t)$ represents the estimate of the sparse signal $d_y(t)$, $\mu > 0$ is a time-constant determined by the physical properties of the implementing system.

Note that $\lceil . \rfloor^{\beta} = |.|^{\beta} sign(.)$ and $a(t) = H_\lambda(v)$ where $H_\lambda(.)$ is defined as eq. (3.6), that is

$$H_\lambda(v) = \max\left(|v| - \lambda, 0\right) \mathrm{sgn}(v)$$

where $\lambda > 0$ is chosen with respect to the noise and the minimum absolute value of the nonzero terms.

Under assumption (A 7.3), the state v of eq. (7.25) converges in finite time to its equilibrium point v^*, and sensor attack estimation $\hat{d}_y(t)$ in eq. (7.25) converges in finite-time to sensor attack $d_y(t)$ in CPS eq. (7.21).

The results presented in this chapter has been published in [80].

7.4 **Summary**

In this Chapter, nonlinear CPSs under deception attacks and sparse sensor attack when the number of sensor measurements is greater than the number of potential sensor attacks are considered. The states of system and the deception state attacks are reconstructed on-line using a HOSM observer. A SR algorithm is used to reconstruct the stealth sensor attacks to the unprotected sensors.

In the next chapter, the proposed algorithms in Chapters 4 to 7 are tested on the WECC power network system and the effectiveness of mentioned approaches are shown through the simulation results.

CHAPTER 8

CASE STUDY: CYBER ATTACK RECONSTRUCTION IN THE US WESTERN ELECTRICITY COORDINATING COUNCIL POWER SYSTEM

The objective of this chapter is to see how different approaches proposed in Chapters 4 to 7 are work to estimate the states and reconstruct the attacks in a CPS case study which is considered the US WECC power network system in this chapter.

8.1 **Mathematical Model of Electrical Power Network**

The descriptor (Differential Algebraic Equations (DAE)) swing mathematical model is adopted to describe the electromechanical behavior of electrical power networks [83, 84]. The DAE swing mathematical model for a power network stabilized by a linear output feedback controller is given by [84]:

$$\begin{bmatrix} I & 0 & 0 \\ 0 & M_g & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} \dot{\delta} \\ \dot{\omega} \\ \dot{\theta} \end{bmatrix} = -\underbrace{\begin{bmatrix} 0 & -I & 0 \\ L^{\theta}_{g,g} & E_g & L^{\theta}_{g,l} \\ L^{\theta}_{l,g} & 0 & L^{\theta}_{l,l} \end{bmatrix}}_{A}\underbrace{\begin{bmatrix} \delta \\ \omega \\ \theta \end{bmatrix}}_{x} + \underbrace{\begin{bmatrix} 0 \\ B_{\omega} \\ B_{\theta} \end{bmatrix}}_{B} d(t) + \begin{bmatrix} 0 \\ P_{\omega} \\ P_{\theta} \end{bmatrix} \tag{8.1}$$

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

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 the 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 sensor measurement vector, the vector $d \in \mathbb{R}^q$ is the attack vector, and $B \in \mathbb{R}^{(2a+b)\times q}$, $D \in \mathbb{R}^{p\times q}$ are the attack distribution matrices; P_ω, P_θ are *known* changes in the mechanical input power to the generators or real power demand at the loads. 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 $L^\theta_{g,g}, L^\theta_{g,l}, L^\theta_{l,g}, L^\theta_{l,l}$ form the following symmetric susceptance matrix

$$L^\theta = \begin{bmatrix} L^\theta_{g,g} & L^\theta_{g,l} \\ L^\theta_{l,g} & L^\theta_{l,l} \end{bmatrix} \tag{8.2}$$

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

Assumption (A 8.1) The matrix $L^\theta_{l,l}$ is nonsingular (such an assumption usually holds in practical electric power systems).

In the next section, assumption (A 8.1) is used in order to change DAE to Ordinary Differential Equation (ODE).

Note that the following terms that appear in the electric power network model (8.1)

$$\begin{bmatrix} 0 \\ B_\omega \\ B_\theta \end{bmatrix} d(t) + \begin{bmatrix} 0 \\ P_\omega \\ P_\theta \end{bmatrix} \tag{8.3}$$

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

8.2 **Transformation of DAE to ODE**

Assuming (A 8.1) holds, the variable θ can be expressed as

$$\theta = \left(R^{\theta}_{l,l}\right)^{-1}\left(-R^{\theta}_{l,g}\delta + P_{\theta} + B_{\theta}d\right) \tag{8.4}$$

substituting eq. (8.4) into eq. (8.1) it is obtained that

$$\begin{aligned} \begin{bmatrix} \dot{\delta} \\ \dot{\omega} \end{bmatrix} &= \begin{bmatrix} \varphi_{\delta}(\delta,\omega) \\ \varphi_{\omega}(\delta,\omega) \end{bmatrix} + \begin{bmatrix} 0 \\ P_{\theta\omega} \end{bmatrix} + \begin{bmatrix} 0 \\ B_{\theta\omega} \end{bmatrix} d(t) \\ y &= C \begin{bmatrix} \delta \\ \omega \end{bmatrix} + Dd(t) \end{aligned} \tag{8.5}$$

where

$$\begin{aligned} \begin{bmatrix} \varphi_{\delta}(\delta,\omega) \\ \varphi_{\omega}(\delta,\omega) \end{bmatrix} &= \begin{bmatrix} 0 & I_{p\times p} \\ M_g^{-1}\left(-R^{\theta}_{g,g} + R^{\theta}_{g,l}\left(R^{\theta}_{l,l}\right)^{-1} R^{\theta}_{l,g}\right) & -M_g^{-1}E_g \end{bmatrix} \begin{bmatrix} \delta \\ \omega \end{bmatrix} \quad . \\ P_{\theta\omega} &= M_g^{-1}\left(P_{\omega} - R^{\theta}_{g,l}\left(R^{\theta}_{l,l}\right)^{-1} P_{\theta}\right), \quad B_{\theta\omega} = M_g^{-1}\left(B_{\omega} - R^{\theta}_{g,l}\left(R^{\theta}_{l,l}\right)^{-1} B_{\theta}\right) \end{aligned} \tag{8.6}$$

8.3 **Parameterization of Mathematical Model of Western Electricity Coordinating Council Power System**

The electrical power network considered in this chapter is a classical nine-bus configuration adopted from [83, 84]. It consists of 3 generators $\{g_1, g_2, g_3\}$ and *6* load buses $\{b_1,\ldots,b_6\}$ as presented in Figure 1.6. Therefore, we have $\omega = \begin{bmatrix}\omega_1 & \omega_2 & \omega_3\end{bmatrix}^T \in \mathbb{R}^3$, $\delta = \begin{bmatrix}\delta_1 & \delta_2 & \delta_3\end{bmatrix}^T \in \mathbb{R}^3$, and $\theta \in \mathbb{R}^6$. The matrices M_g and E_g which are the diagonal matrices of the generator inertial and damping coefficients are shown in eq. (1.8), the Laplacian matrix associated with the susceptibility-weighted graph is the symmetric susceptibility matrix $L^{\theta} \in \mathbb{R}^{9\times 9}$ and is given in (1.10)-(1.11). The inputs P_{ω} and P_{θ} are due to *known* changes in the mechanical input power to the generators and real power demands at the loads and are defined as

$$P_{\omega}=\begin{bmatrix}0.716 & 1.63 & 0.85\end{bmatrix}^{T},\quad P_{\omega}=\begin{bmatrix}0 & -1.25 & -0.94 & 0 & -1 & 0\end{bmatrix}^{T}. \tag{8.7}$$

8.4 State Estimation and Attacks Reconstruction Using the Proposed Sliding Mode Observation Algorithms

The proposed approaches in the chapters 4 to 7 are applied to estimate the states and reconstruct the sensor and/or state attacks of the WECC power network described in Section 8.3.

8.4.1 Reconstruction of Attacks and Estimation of States: the Number of Sensors and the Number of Potential Attacks Is Equal

In this section, it is considered that the generator speed *deviations* from synchronous $\omega \in \mathbb{R}^{3}$ are measured. Therefore, the output of system eq. (8.1) is

$$y=\omega+d(t) \tag{8.8}$$

where $y=\begin{bmatrix}y_{1} & y_{2} & y_{3}\end{bmatrix}^{T}\in\mathbb{R}^{3}$, $\omega=\begin{bmatrix}\omega_{1} & \omega_{2} & \omega_{3}\end{bmatrix}^{T}\in\mathbb{R}^{3}$, and $d=\begin{bmatrix}d_{1} & d_{2} & d_{3}\end{bmatrix}^{T}\in\mathbb{R}^{3}$ represents the stealth attack to the sensors.

The problem is reconstructing the attacks in the Linear WECC power system when the number of measurements and the number of attacks is equal.

In this study, it is given that

$$B_{\omega}=I_{3},\ B_{\theta}=0_{6\times 3}. \tag{8.9}$$

Therefore, THE WECC power network eq. (8.1) is rewritten for this parameterized case as

$$\begin{bmatrix} I & 0 & 0 \\ 0 & M_g & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{\delta} \\ \dot{\omega} \\ \dot{\theta} \end{bmatrix} = - \begin{bmatrix} 0 & -I & 0 \\ R_{g,g}^{\theta} & E_g & R_{g,l}^{\theta} \\ R_{l,g}^{\theta} & 0 & R_{l,l}^{\theta} \end{bmatrix} \begin{bmatrix} \delta \\ \omega \\ \theta \end{bmatrix} + \begin{bmatrix} 0 \\ I \\ 0 \end{bmatrix} d(t) + \begin{bmatrix} 0 \\ P_{\omega} \\ P_{\theta} \end{bmatrix} . \qquad (8.10)$$

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

The system eq. (8.10) is reduced to

$$\begin{cases} \dot{\delta} = \omega \\ \dot{\omega} = \varphi_{\omega}(\delta, \omega) + P_{\theta\omega} + M_g^{-1} d(t) \\ y = \omega + d(t) \end{cases} . \qquad (8.11)$$

The WECC power system eq. (8.11) can be presented in a numerical format as follows

$$\begin{bmatrix} \dot{\delta}_1 \\ \dot{\delta}_2 \\ \dot{\delta}_3 \\ \dot{\omega}_1 \\ \dot{\omega}_2 \\ \dot{\omega}_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ -0.3145 & 0.1187 & 0.1158 & -1 & 0 & 0 \\ 0.4363 & -0.8474 & 0.4111 & 0 & -2 & 0 \\ 0.9046 & 0.8736 & -1.7782 & 0 & 0 & -3 \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \omega_1 \\ \omega_2 \\ \omega_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1.1886 \\ 18.6934 \\ -11.9475 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 8 & 0 & 0 \\ 0 & 29.4118 & 0 \\ 0 & 0 & 62.5 \end{bmatrix} d(t) \qquad (8.12)$$

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

In order to apply the sensor attack reconstruction algorithm proposed in Chapter 4 the system eq. (8.11) is presented in the form of eq. (4.9) as

$$\begin{cases} \dot{\delta} = y - d(t) \\ \dot{y} = \varphi_{\omega 1}\delta + \varphi_{\omega 2} y - \varphi_{\omega 2} d(t) + P_{\theta\omega} + M_g^{-1} d(t) + \dot{d}(t) \end{cases} \qquad (8.13)$$

where $\varphi_{\omega}(\delta, \omega)$ is presented in the form of

$$\varphi_{\omega}(\delta, \omega) = \varphi_{\omega 1}\delta + \varphi_{\omega 2} y - \varphi_{\omega 2} d(t) . \qquad (8.14)$$

8.4.1.1 Sliding Mode Observer: All Sensors Can be Attacked

Note that in this case, no sensors have been protected, and all sensors might be attacked, i.e. 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 WECC power system described in eq. (8.13) is designed in the format of eq. (4.11) as

$$\begin{cases} \dot{\hat{\delta}} = \hat{y} \\ \dot{\hat{y}} = \varphi_{\omega 1}\hat{\delta} + \varphi_{\omega 2}\hat{y} + P_{\theta\omega} + \upsilon \end{cases} \tag{8.15}$$

where υ is the injection term designed in a format of eq. (4.14). that is

$$\upsilon = \left(\rho + L_3\right)\frac{e_y}{\left\|e_y\right\|},\ \rho, L_3 > 0 .$$

Finally, in accordance with eq. (4.15), the sensor attack is reconstructed as

$$\hat{d} = \left(\frac{-\varphi_{\omega 1}}{s} - \varphi_{\omega 2} + M_g^{-1} + sI\right)^{-1} \upsilon_{eq} \ . \tag{8.16}$$

Remark 8.1 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 eq. (8.16) is discussed in Remark 4.3 in Chapter 4.

8.4.1.2 Sliding Mode Observer: Some Sensors Are Protected

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

$$Dd(t) = \left[0 \quad d_2 \quad d_3\right]^T \tag{8.17}$$

and the output/sensed equations in eq. (8.10) becomes

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}. \tag{8.18}$$

To apply the sensor attack reconstruction algorithms proposed in the Section 4.2.1.1 the WECC power network eq. (8.13) with one protected sensor/measurement is presented as

$$\begin{aligned} \dot{\delta} &= \begin{bmatrix} y_1 \\ y_2 - d_2 \\ y_3 - d_3 \end{bmatrix} \\ \dot{y} &= \varphi_{\omega 1}\delta + \varphi_{\omega 2}\begin{bmatrix} y_1 \\ y_2 - d_2 \\ y_3 - d_3 \end{bmatrix} + P_{\theta\omega} + M_g^{-1} d(t) + \dot{d}(t) \end{aligned} \quad . \tag{8.19}$$

Then, the state/attack observer is designed as follows

$$\begin{aligned} \dot{\hat{\delta}} &= \hat{y} \\ \dot{\hat{y}} &= \varphi_{\omega 1}\hat{\delta} + \varphi_{\omega 2}\hat{y} + P_{\theta\omega} + \upsilon \end{aligned} \tag{8.20}$$

where υ is the injection term.

The attack signals d_2, d_3 are exactly estimated as

$$\begin{bmatrix} \hat{d}_2 \\ \hat{d}_3 \end{bmatrix} = \left(\frac{-\varphi'_{\omega 1}}{s} - \varphi'_{\omega 2} + \left(M'_g\right)^{-1} + sI_{2\times 2} \right)^{-1} \upsilon_{eq} \tag{8.21}$$

where, in accordance with eq. (4.37),

$$\begin{aligned} &\varphi'_{\omega 1} = \begin{bmatrix} -0.8474 & 0.4111 \\ 0.8736 & -1.7782 \end{bmatrix}, \quad \varphi'_{\omega 2} = \begin{bmatrix} -2 & 0 \\ 0 & -3 \end{bmatrix} \\ &\left(M'_g\right)^{-1} = \begin{bmatrix} 29.4118 & 0 \\ 0 & 62.5 \end{bmatrix}, \quad \upsilon_{eq} = \begin{bmatrix} \upsilon_{2eq} \\ \upsilon_{3eq} \end{bmatrix} \\ &\upsilon_2 = \left(\rho_1 + L_{11}\right)\frac{e_{y_2}}{\left\| e_{y_2} \right\|}, \quad \upsilon_3 = \left(\rho_2 + L_{12}\right)\frac{e_{y_3}}{\left\| e_{y_3} \right\|} \end{aligned} \tag{8.22}$$

where $\rho_1, \rho_2, L_{11}, L_{12} > 0$ and $e_{y_2} = y_2 - \hat{y}_2$, $e_{y_3} = y_3 - \hat{y}_3$. The estimation of $\upsilon_{2eq}, \upsilon_{3eq}$ is discussed in Remark 4.3 in Chapter 4.

Remark 8.2 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 8.1 due to the reduced order of the matrix to be inverted.

8.4.1.3 Cleaning up the Measurements Corrupted by Attacks

As soon as the attacks are exactly reconstructed in (8.16) or (8.21), 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" WECC power system eq. (8.11) becomes

$$\begin{cases} \dot{\delta}_c = \omega_c \\ \dot{\omega}_c = \varphi_\omega(\delta_c, \omega_c) + P_{\theta\omega} + M_g^{-1}\left(d(t) - \hat{d}(t)\right) \\ y_c = \omega_c + \left(d(t) - \hat{d}(t)\right) \end{cases} \tag{8.23}$$

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 WECC power system eq. (8.23) converges to

$$\begin{cases} \dot{\delta}_c = \omega_c \\ \dot{\omega}_c = \varphi_\omega(\delta_c, \omega_c) + P_{\theta\omega} \\ y_c = \omega_c \end{cases} \tag{8.24}$$

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

9.4.1.4. Simulation Results

Simulation set-up: The simulation results have been obtained via MATLAB. Three simulation experiments have been performed for sensor attack reconstruction in the WECC power system in eq. (8.10).

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 $y^T = (\omega_1, \omega_2, \omega_3)^T$. The attack named *stealth attacks* [31, 32] that completely corrupts the measurement vector are considered as

$$d_1 = -1.1\omega_1 + 2\sin(t), \quad d_2 = -0.9\omega_2 + \cos(0.5t), \quad d_3 = -0.8\omega_3 + \sin(t). \qquad (8.25)$$

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

The SMO parameters used in the simulations are presented in the following table.

Table 1. Simulation Parameters

Sampling time (sec)	10^{-4}	Parameter γ in (4.27)	0.1
Integration algorithm	Euler	Parameter α in (4.23)	0.1
Parameter ρ in (4.14) and (4.22)	0.01	Parameter ε in (4.23)	0.1
Injection term gain L_3 in (4.14)	100	Parameter ℓ_0 in (4.26)	1
Parameter σ_0 in (4.27)	0.01		

Note that the gain of injection term, L_3, in eq. (4.14) is assumed unknown in the adaptive scheme. Also, no LPF is used for estimating υ_{eq}, since the transfer matrix in attack reconstruction formula eq. (8.16) has the LPF property. Therefore, υ is used in eq. (8.16), while υ_{eq} is recovered automatically.

In this study, the attack observations are done by both the fixed and adaptive gain SMO presented in Chapter 4. The results are presented in Figures 8.1 – 8.11.

The results of the experiments 1: The plot presented in Figure 8.1 demonstrates the convergence of the measured outputs y_1, y_2, y_3 (generator speed *deviations* from synchronous measured in rad/s) to zero as expected.

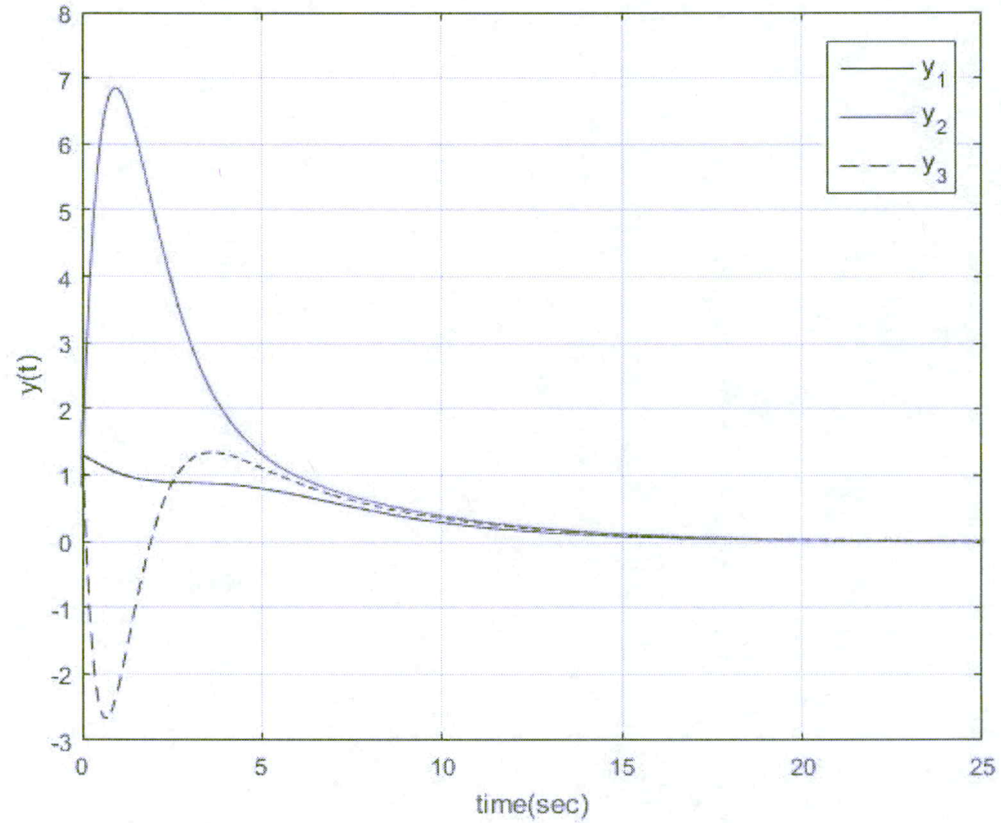


Figure 8.1 Outputs of WECC Power System without Attacks

The results of the experiments 2: Figure 8.2 demonstrates the measured outputs y_1, y_2, y_3 while the sensors are under *stealth attacks* eq. (8.25). It can be observed that the outputs are corrupted and do not converge to zero due to the *stealth attack.*

The results of the experiment3: Figure 8.3 shows the compensated outputs, i.e. when the attacks are reconstructed and the measurements are cleaned from the attacks.

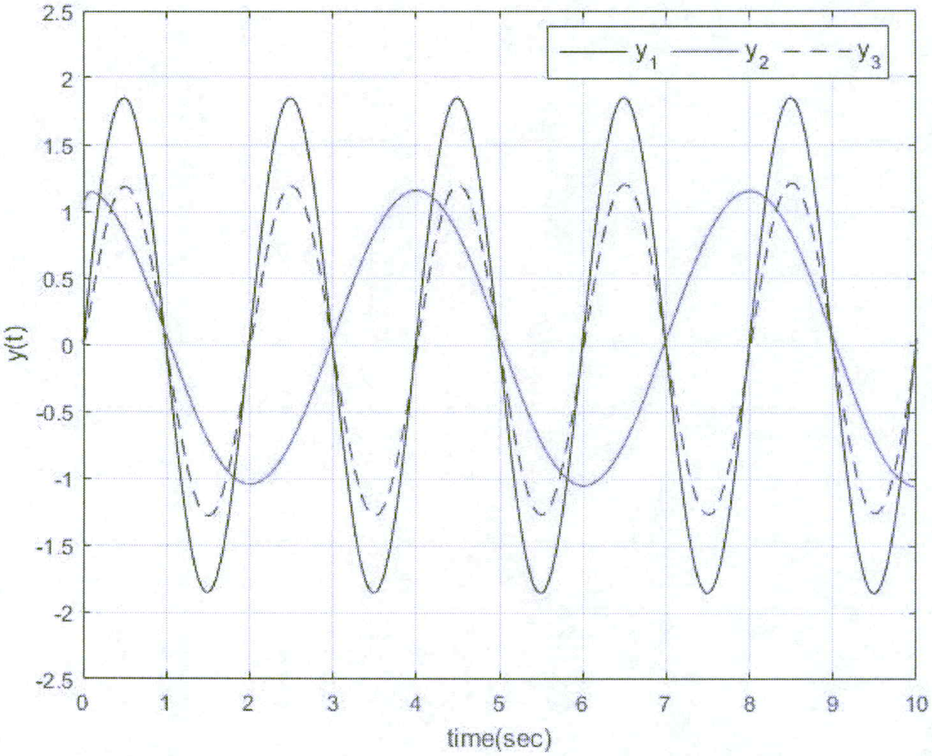


Figure 8.2 Outputs of WECC Power System under Stealth Attack

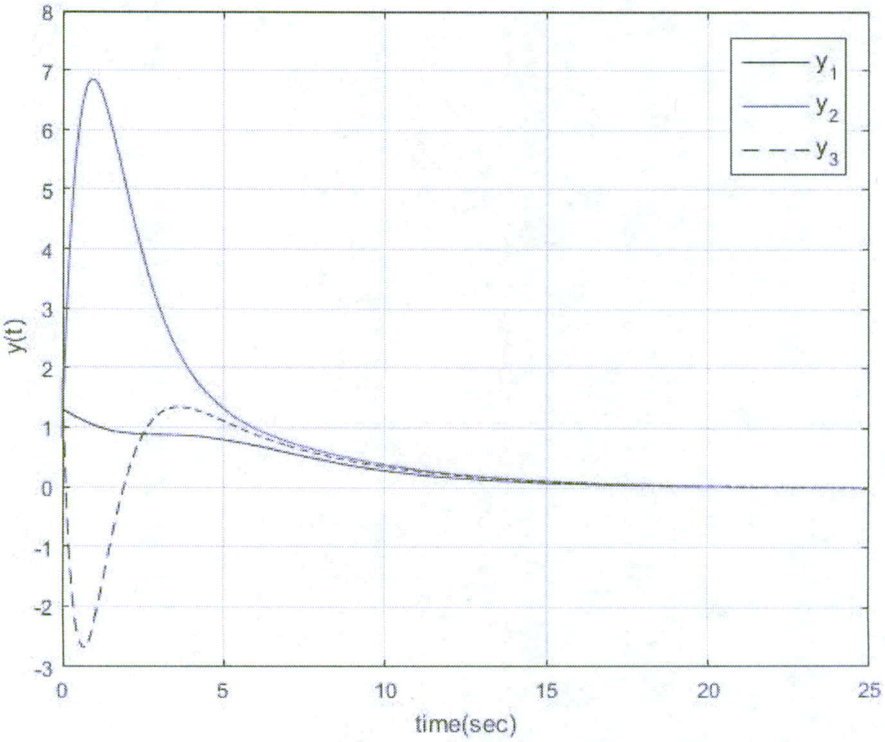


Figure 8.3 Outputs of WECC Power System After the Corrupted Measurements Are Cleaned

The cleaned output dynamics (Figure 8.3) practically coincide with the outputs of the systems without attack (see Figure 8.1) after a short transient. In Figures 8.4 – 8.6, the outputs of system in the three scenarios (without attack, corrupted by attack, and compensated after being attacked) are compared.

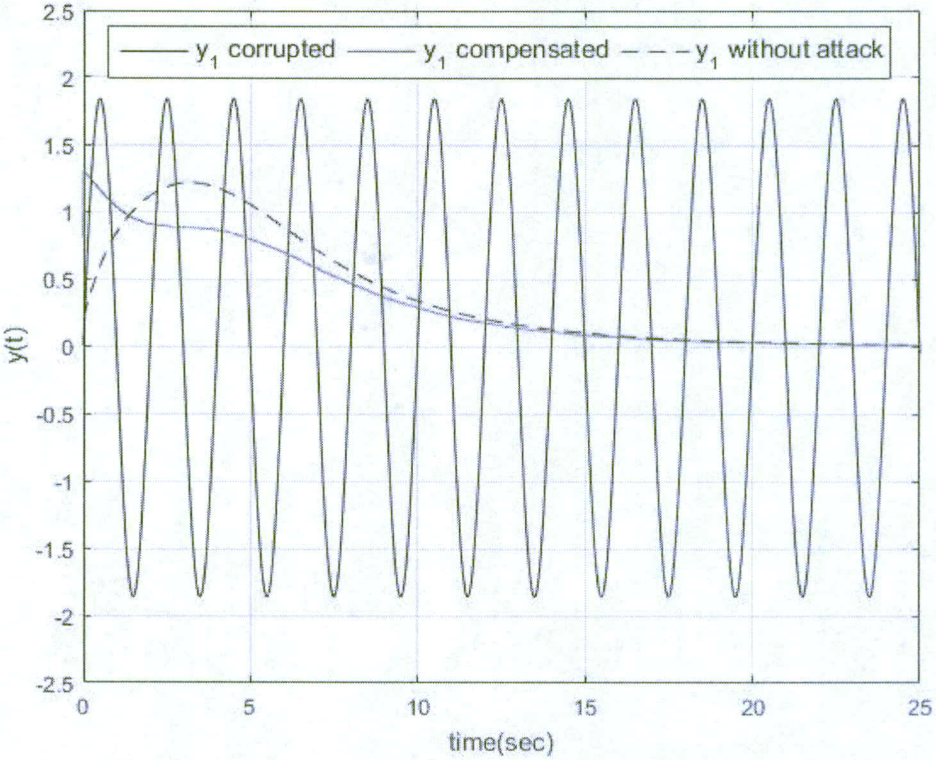


Figure 8.4 Comparing y_1 without Attack with Corrupted y_1 and Compensated y_1 After Being Attacked

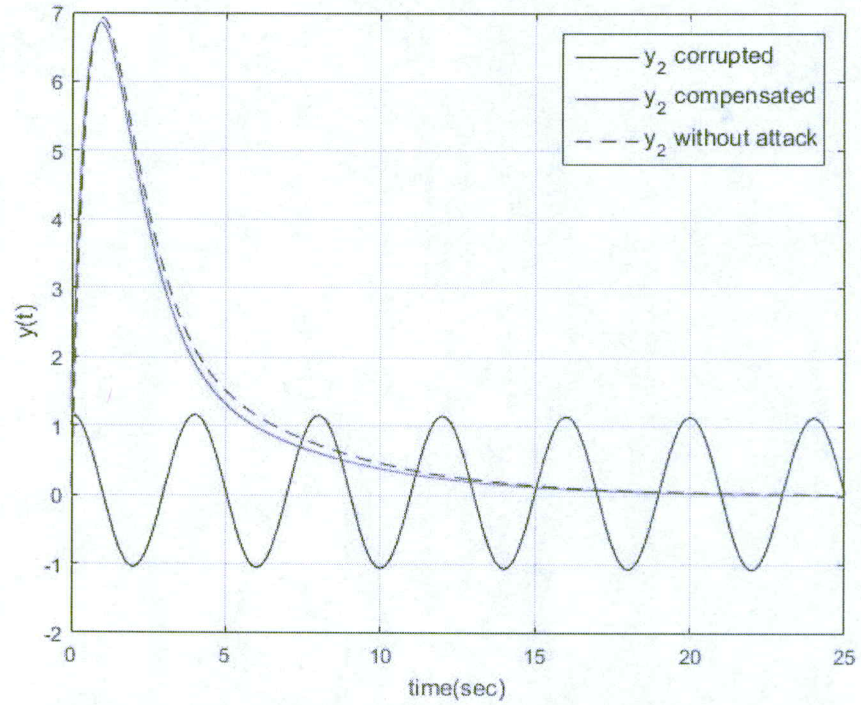


Figure 8.5 Comparing y_2 without Attack with Corrupted y_2 and Compensated y_2 After Being Attacked

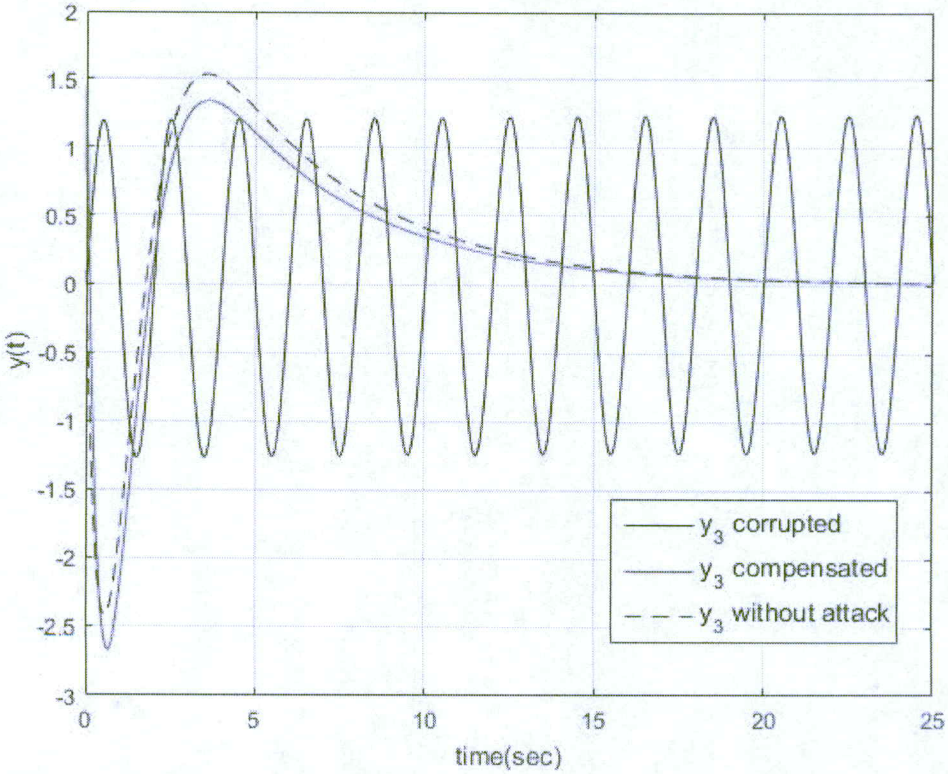


Figure 8.6 Comparing y_3 without Attack with Corrupted y_3 and Compensated y_3 After Being Attacked

It is shown in Figures 8.7 – 8.9 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 SMO and the adaptive SMO presented in Chapter 4. Figures 8.10 and 8.11 show the sliding mode injection terms used in both observers.

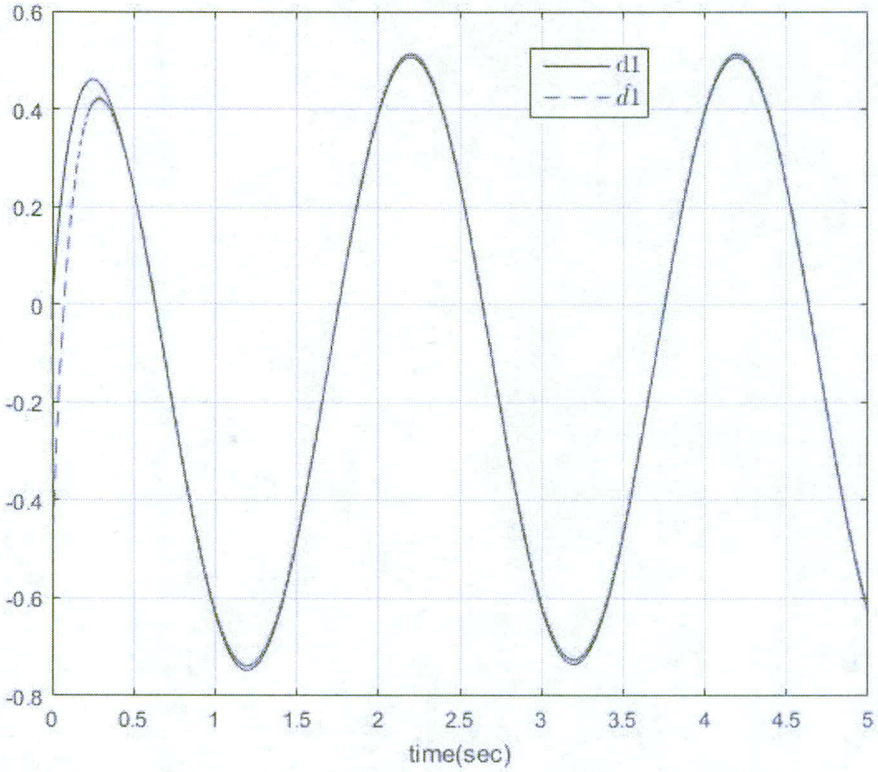


Figure 8.7 Comparing Sensor Attacks d_1 with Its Reconstruction $\hat{d}_1$

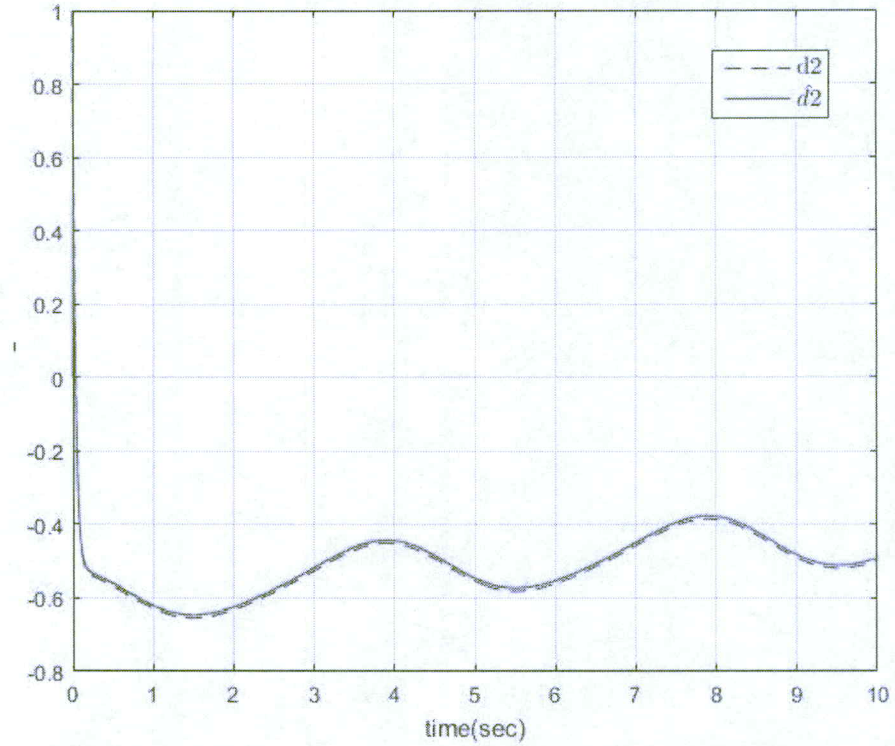


Figure 8.8 Comparing Sensor Attacks d_2 with Its Reconstruction $\hat{d}_2$

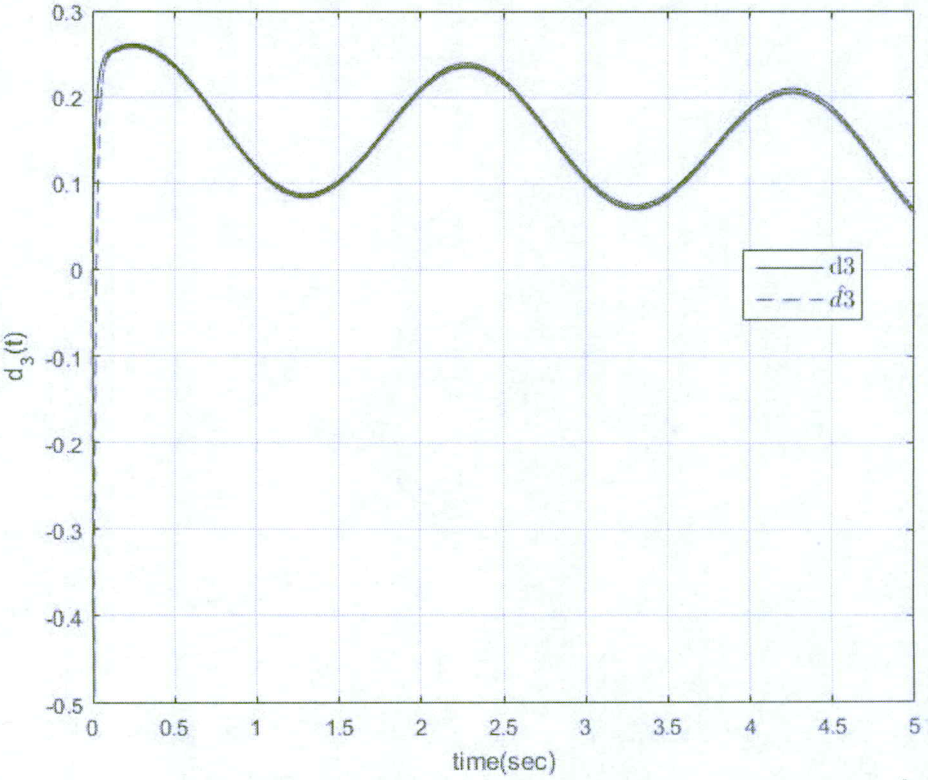


Figure 8.9 Comparing Sensor Attacks d_3 with Its Reconstruction $\hat{d}_3$

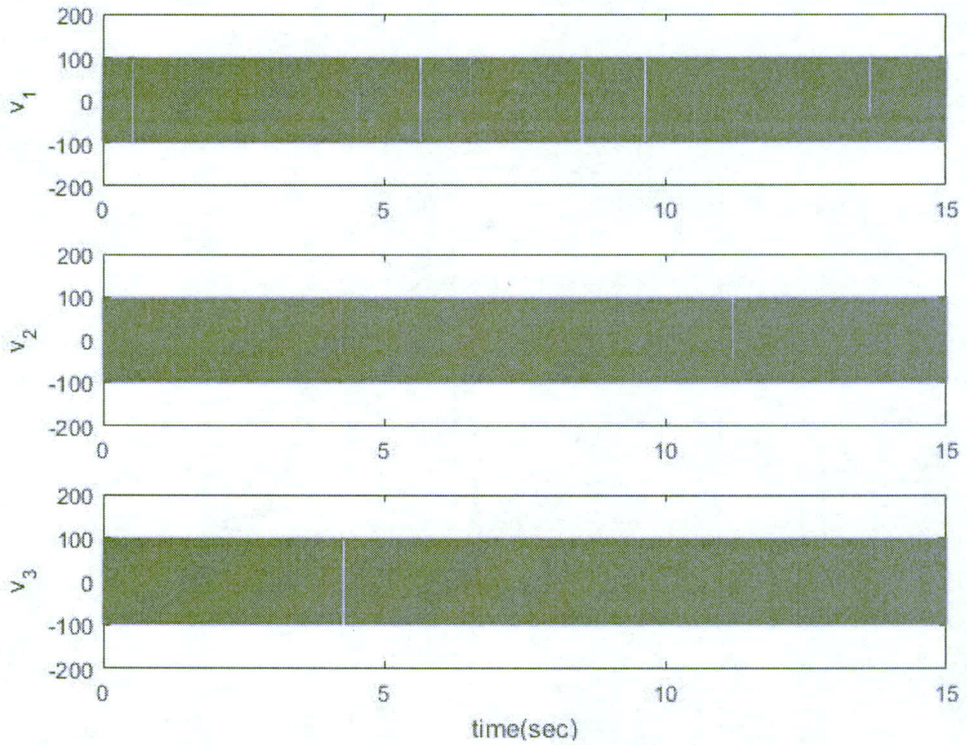


Figure 8.10 Fixed Gain Sliding Mode Injection Terms $\upsilon_1, \upsilon_2, \upsilon_3$

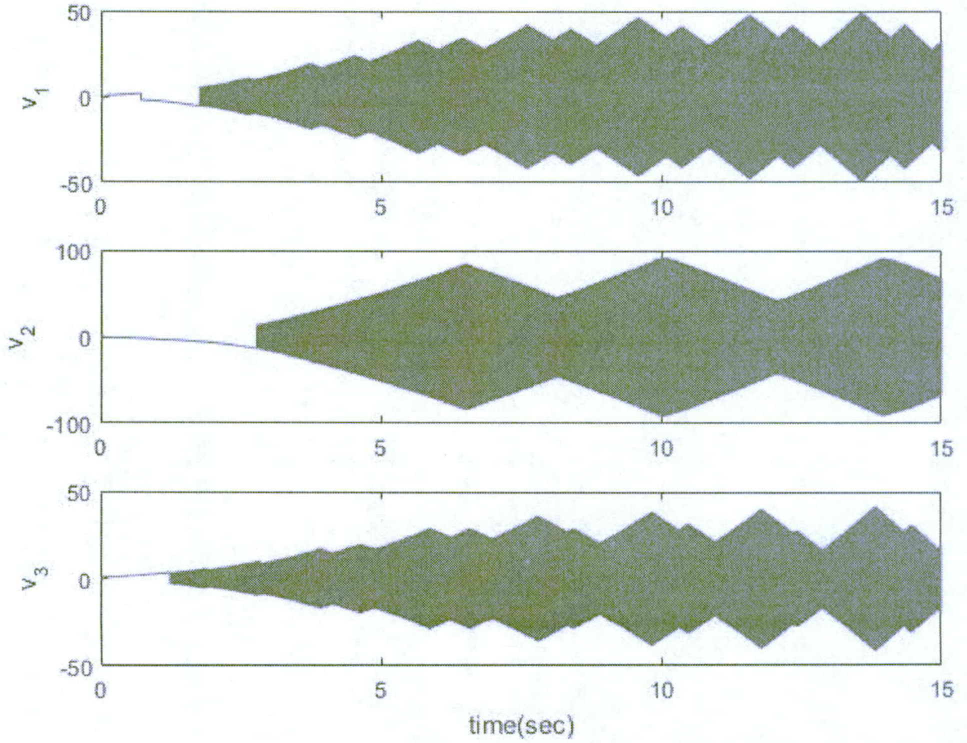


Figure 8.11 Adaptive Sliding Mode Injection Terms $\upsilon_1, \upsilon_2, \upsilon_3$

Remark 8.3 The output stabilization plots under both the fixed and adaptive gain observers look the same here, since the attack reconstruction fixed gain filter eq. (8.16) includes a LPF inherently and it behaves like a damper that mitigates the rippling in the attack estimation that may be generated by a large sliding gain. Note that the main advantage of the adaptive SMO is in self-tuning.

8.4.2 Reconstruction of Attacks and Estimation of States: the Number of Sensors is Greater than the Number of Potential Attacks

The application of two approaches proposed in Chapter 5 are tested for attack reconstruction on the mathematical model of US WECC power system eq. (8.10) under attack here.

8.4.2.1 Sliding Mode Observer with Dynamic Extension of Injection Term

The fixed and adaptive gain SMOs with filtering feature proposed in Chapter 5 is applied for attack reconstruction on the attacked WECC (8.10).The Simulation setup for designing the SMO in this section is as follows:

(a) There are three sensors that measure the generator speed deviation from synchronicity of three generators of WECC and two attacks that corrupt the second and third sensors as

$$\begin{aligned} &y_1 = \omega_1, \quad y_2 = \omega_2 + d_{\omega_2}, \quad y_3 = \omega_3 + d_{\omega_3} \\ &d_{\omega_2} = -\omega_2 + 2\sin(0.5\pi t), \quad d_{\omega_3} = -\omega_3 + \cos(0.5\pi t) + \sin(\pi t) \end{aligned} \tag{8.26}$$

(b) The matrices B_θ, B_ω, C, D in eq. (8.1) are given as follows for this study.

$$B_{\theta} = \mathbf{0}_{6\times 2}, B_{\omega} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, C = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, D = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{8.27}$$

Consider eqs. (8.5), (5.4) and (5.5), $x_1 \in \mathbb{R}^3$ and $x_2 \in \mathbb{R}^3$ are defined

$$x_1 = [\delta_1 \quad \delta_2 \quad \delta_3]^T, x_2 = [\omega_1 \quad \omega_2 \quad \omega_3]^T \tag{8.28}$$

Substitute (8.27) in (8.5) and (8.6), then it is clear that A_{11} in eq. (5.4) is not Hurwitz and it is needed to define a matrix $L \in \mathbb{R}^{3\times 3}$ as disused in eq. (5.6). If $L = I_{3\times 3}$ is selected, then the new variables $\bar{x}_1 \in \mathbb{R}^3$ and $\bar{x}_2 \in \mathbb{R}^3$ in eqs. (5.10) are defined as follows

$$\bar{x}_1 = \delta + \omega, \quad \bar{x}_2 = \omega \tag{8.29}$$

and according to eqs. (5.13), (5.14), eq. (8.29) can be rewritten as

$$\bar{x}_1 = [\delta_1 + \omega_1 \quad \delta_2 + \omega_1 \quad \delta_3 + \omega_1]^T, \quad \bar{x}_{21} = \omega_1, \quad \bar{x}_{22} = [\omega_2 \quad \omega_3]^T \tag{8.30}$$

(c) The matrices $A_{22}^s \in \mathbb{R}$, $A_{33}^s \in \mathbb{R}^{2\times 2}$ are selected as

$$A_{22}^s = -1, \quad A_{33}^s = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \tag{8.31}$$

Finally, attacks are reconstructed/estimated according to eq. (5.25) which is

$$\hat{d} = G^*(s)\upsilon_{eq}$$

where $G^*(s) \in \mathbb{R}^{2\times 2}$ is given by

$$G*(s) = \begin{bmatrix} G_{11}^*(s) & G_{12}^*(s) \\ G_{21}^*(s) & G_{22}^*(s) \end{bmatrix} \tag{8.32}$$

where

$$\begin{aligned}
G_{11}^{*}(s) &= \frac{s^5 - 67.67s^4 - 145.6s^3 - 90.81s^2 - 10.94s + 0.2193}{s^6 + 98.19s^5 + 2185s^4 + 2775s^3 + 526.6s^2 + 22.29s + 0.2193}, \\
G_{12}^{*}(s) &= \frac{0.7783s^4 + 24.67s^3 + 30.43s^2 + 3.281s}{s^6 + 98.19s^5 + 2185s^4 + 2775s^3 + 526.6s^2 + 22.29s + 0.2193}, \\
G_{21}^{*}(s) &= \frac{0.3663s^4 + 24.09s^3 + 30.36s^2 + 6.089s}{s^6 + 98.19s^5 + 2185s^4 + 2775s^3 + 526.6s^2 + 22.29s + 0.2193}, \\
G_{22}^{*}(s) &= \frac{-s^5 - 34.56s^4 - 102.5s^3 - 83.93s^2 - 13.64s - 0.2193}{s^6 + 98.19s^5 + 2185s^4 + 2775s^3 + 526.6s^2 + 22.29s + 0.2193}
\end{aligned} \tag{8.33}$$

8.4.2.1.1 Simulation Results

Simulation results shown in Figures 8.12 and 8.13 illustrate that sensor attacks $d_{\omega_2}, d_{\omega_3}$ are accurately estimated by $\hat{d}_{\omega_2}, \hat{d}_{\omega_3}$. The attack observation is done by both the fixed and adaptive gain SMO. Their injection terms are illustrated in Figure 8.14. As it is shown, the gain of SMO injection term changes when it is needed.

Figures 9.15-9.17 compare the corrupted measurements with cleaned up measurements from reconstructed attacks and measurements when there is no attack. These comparisons clearly show the importance of attack reconstruction and compensation. The presented simulation results are given by applying fixed gain SMO.

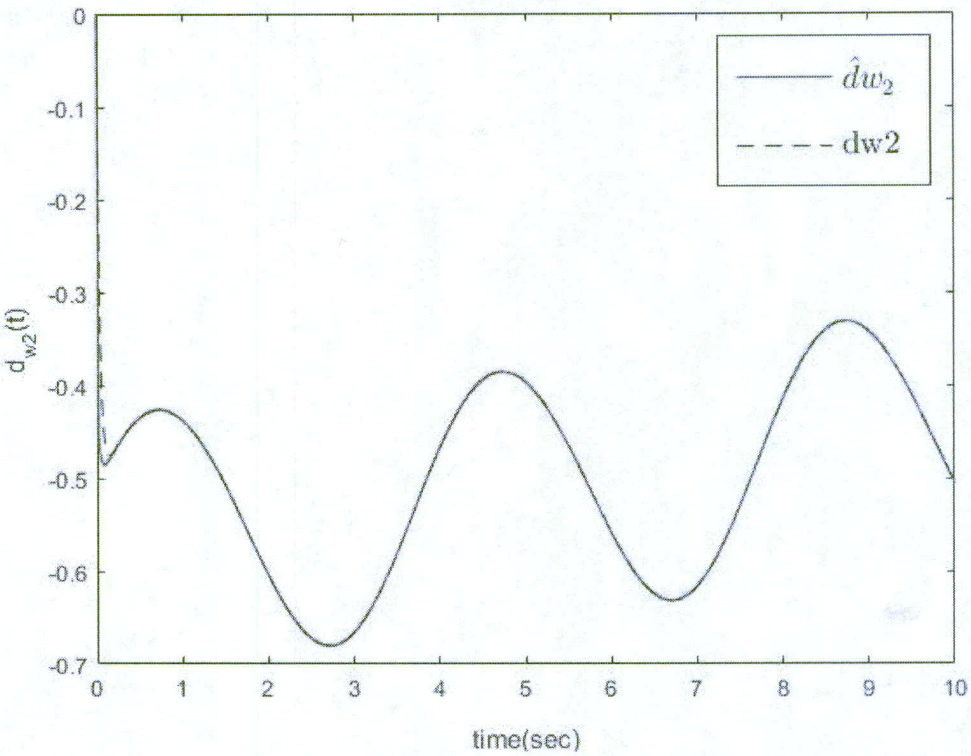


Figure 8.12 Comparing Sensor Attack d_{ω_2} with Its Reconstruction $\hat{d}_{\omega_2}$

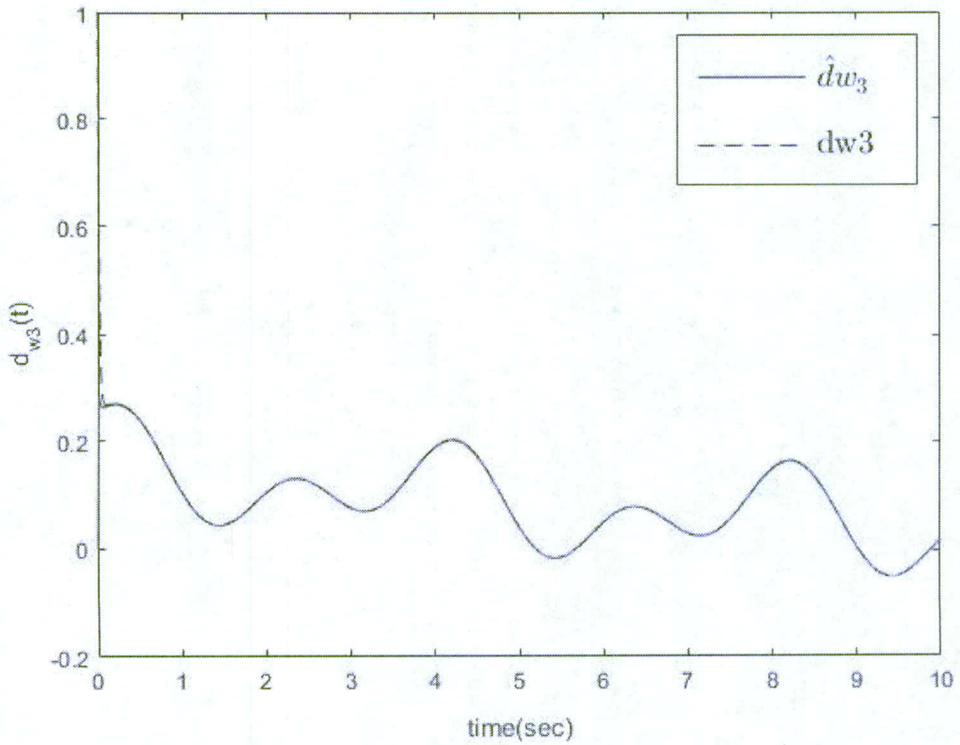


Figure 8.13 Comparing Sensor Attack d_{ω_3} with Its Reconstruction $\hat{d}_{\omega_3}$

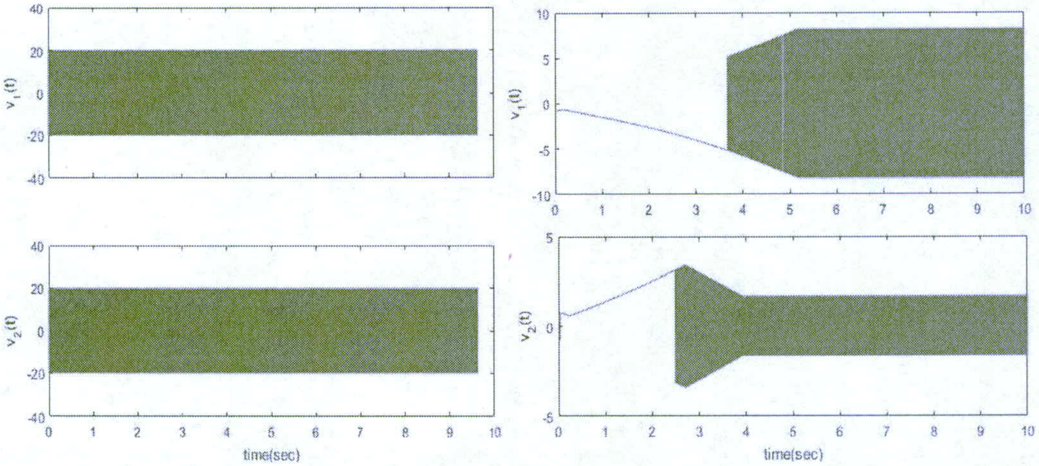


Figure 8.14 Fixed and Adaptive Gain Sliding Mode Injection Terms υ_1,υ_2

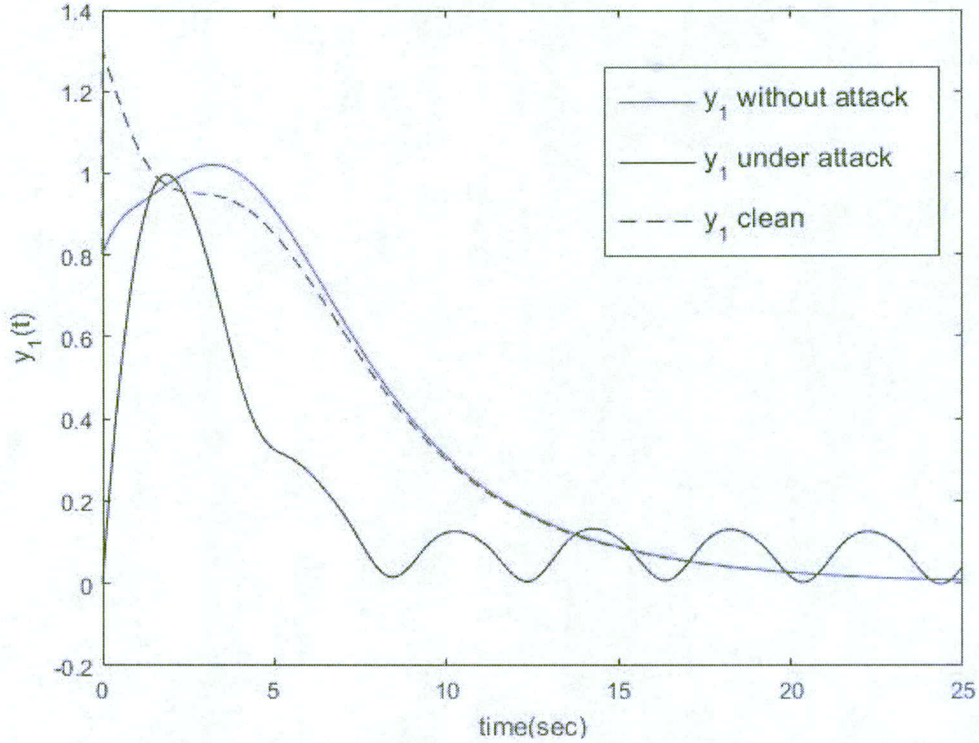


Figure 8.15 Comparing y_1 without Attack with Corrupted y_1 and Compensated y_1 after being Attacked

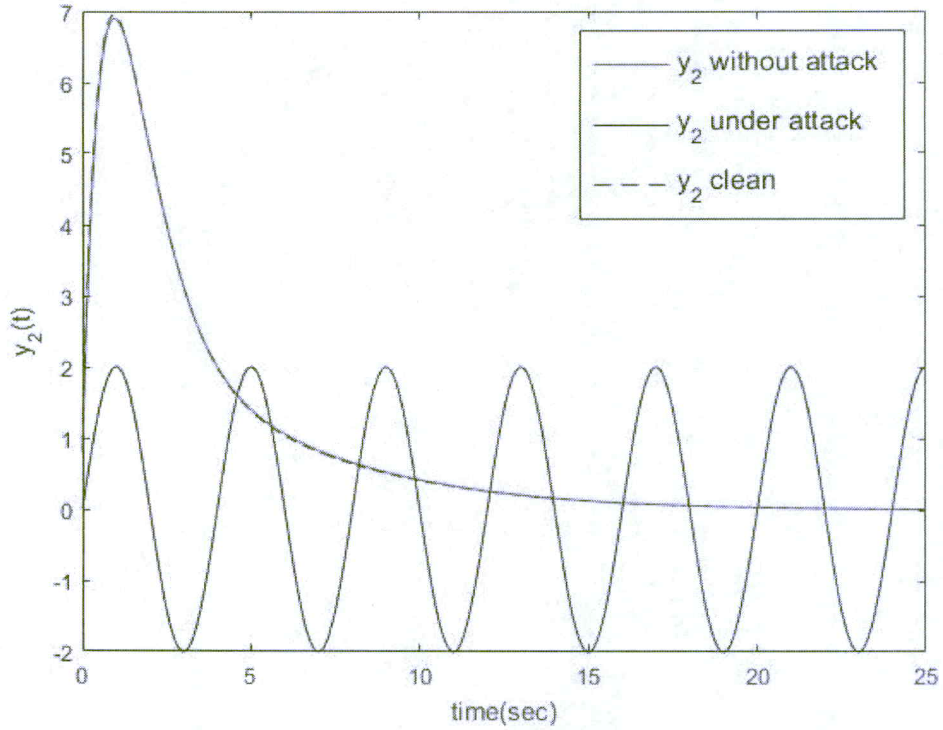


Figure 8.16 Comparing y_2 without Attack with Corrupted y_2 and Compensated y_2 After Being Attacked

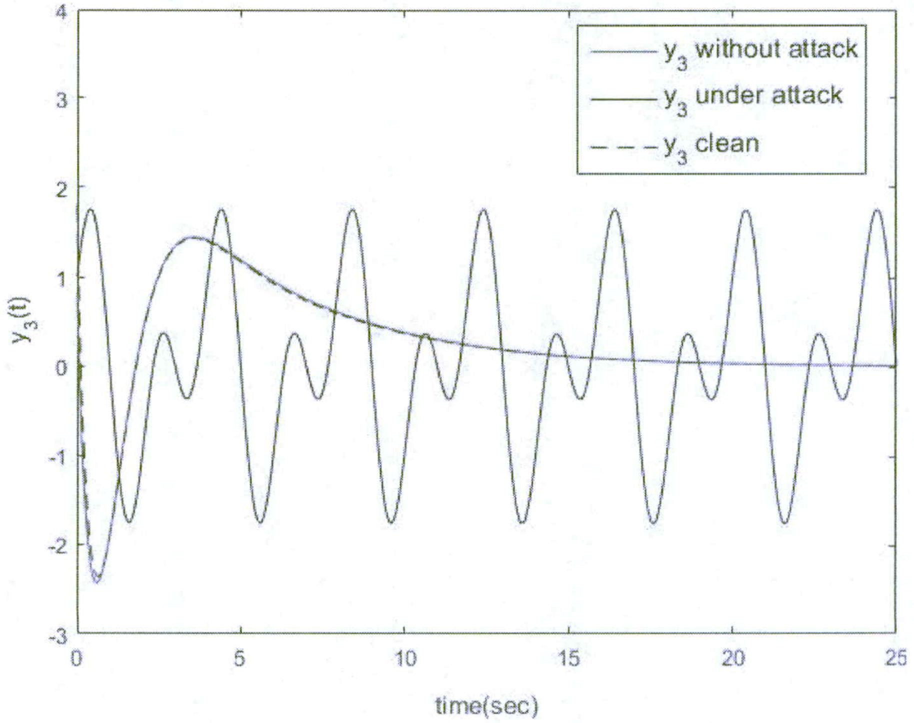


Figure 8.17 Comparing y_3 without Attack with Corrupted y_3, and Compensated y_3 After Being Attacked

8.4.2.2 Line-by-Line Super Twisting Sliding Mode Observer

In this section, we investigate the WECC power system eq. (8.10) when we have more sensors rather than plant attacks, i.e. there are 6 sensor measurements and 3 plant attacks. The matrices B and D in eq. (8.1) are defined in such a way that plant attack d_x and sensor attack d_y can be written separately as follows

$$\begin{bmatrix} I & 0 & 0 \\ 0 & M_g & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} \dot{\delta} \\ \dot{\omega} \\ \dot{\theta} \end{bmatrix} = -\begin{bmatrix} 0 & -I & 0 \\ R^{\theta}_{g,g} & E_g & R^{\theta}_{g,l} \\ R^{\theta}_{l,g} & 0 & R^{\theta}_{l,l} \end{bmatrix}\begin{bmatrix} \delta \\ \omega \\ \theta \end{bmatrix} + \begin{bmatrix} 0 \\ I \\ 0 \end{bmatrix} d_x(t) + \begin{bmatrix} 0 \\ P_{\omega} \\ P_{\theta} \end{bmatrix}$$
$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} C_{\delta} & 0 \\ 0 & C_{\omega} \end{bmatrix}\begin{bmatrix} \delta \\ \omega \end{bmatrix} + \begin{bmatrix} D_{\delta} \\ D_{\omega} \end{bmatrix} d_y(t) \tag{8.34}$$

where

$$C_{\delta} = I_3, \quad C_{\omega} = I_3, \quad D_{\delta} = 0_{3\times 6}, \quad D_{\omega} \in \mathbb{R}^{3\times 6} = \begin{bmatrix} 0 & 1 & 2 & 0 & 1 & 1 \\ 1 & 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}. \tag{8.35}$$

The novel adaptive STW observer discussed in section 5.2.3 is used to estimate the states of WECC power system eq. (8.34) under attack, and reconstruct the state attack d_x. Then, having 3 corrupted sensors and 6 potential sensor attacks, the SR algorithm is used to reconstruct the sparse sensor attack d_y in (8.34). Finally, the estimate of attack will be used to clean the sensors and system.

The WECC power system eq. (8.34) can be rewritten as

$$\begin{cases} \begin{bmatrix} \dot{\delta} \\ \dot{\omega} \end{bmatrix} = \begin{bmatrix} \omega \\ M_g^{-1}\left(-R^{\theta}_{g,g} + R^{\theta}_{g,l}\left(R^{\theta}_{l,l}\right)^{-1} R^{\theta}_{l,g}\right)\delta - M_g^{-1}E_g\omega + P_{\theta\omega} \end{bmatrix} + \bar{B}d_x(t) \\ \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \bar{C}_{\delta} \\ \bar{C}_{\omega} \end{bmatrix}\begin{bmatrix} \delta \\ \omega \end{bmatrix} + \begin{bmatrix} 0 \\ D_{\omega} \end{bmatrix} d_y(t) \end{cases} \tag{8.36}$$

where

$$\begin{aligned} & P_{\theta\omega}=M_g^{-1}\left(P_\omega-L_{g,l}^{\theta}\left(L_{l,l}^{\theta}\right)^{-1}P_\theta\right), B_{\theta\omega}=M_g^{-1}\left(B_\omega-L_{g,l}^{\theta}\left(L_{l,l}^{\theta}\right)^{-1}B_\theta\right) \\ & \bar{C}_\delta=\begin{bmatrix} I_3 & 0_3 \end{bmatrix},\quad \bar{C}_\omega=\begin{bmatrix} 0_3 & I_3 \end{bmatrix},\quad \bar{B}=\begin{bmatrix} 0_3 \\ M_g^{-1} \end{bmatrix} \end{aligned} \quad . \tag{8.37}$$

Remark 8.4: It can be verified that D_ω satisfies RIP condition defined in eq. (3.3).

A suitable choice of C_a and y_a are

$$C_a=\begin{bmatrix} C_1 \\ C_1A \\ C_2 \\ C_2A \\ C_3 \\ C_3A \end{bmatrix}=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},\quad y_a=\begin{bmatrix} y_1 \\ \mu\left(y_1-\hat{y}_1\right) \\ y_2 \\ \mu\left(y_2-\hat{y}_2\right) \\ y_3 \\ \mu\left(y_3-\hat{y}_3\right) \end{bmatrix} \tag{8.38}$$

It is easy to verify that

$$\begin{aligned} \bar{C}_{\delta 1}\bar{B}=0, &\quad \bar{C}_{\delta 1}A\bar{B}\neq 0 \\ \bar{C}_{\delta 2}\bar{B}=0, &\quad \bar{C}_{\delta 2}A\bar{B}\neq 0 \\ \bar{C}_{\delta 3}\bar{B}=0, &\quad \bar{C}_{\delta 3}A\bar{B}\neq 0 \end{aligned} \tag{8.39}$$

where $\bar{C}_{\delta i}$ is the i^{th} row of matrix $\bar{C}_\delta$.

Also, it is easy to check that $rank\,(C_aB)=rank\,(B)$.

The states $\hat{\delta}$, $\hat{\omega}$ and plant attacks $d_x(t)$ in eq. (8.36) are reconstructed as described in section 5.2 by using STW observer. The estimated $\hat{\omega}$ is used in eq. (8.36) to get

$$y_2-\hat{\omega}=D_\omega d_y(t). \tag{8.40}$$

The SR algorithm described in Section 3.1 can then be applied to reconstruct the sparse $d_y(t)$ in WECC power system eq. (8.36), where only one out of six potential attacks $d_{y1},\ldots,d_{y6}$ is non-zero.

Consider the following attacks which affect the states and the sensors of WECC power system eq. (8.36) starting at $t = 10\,\text{sec}$

$$d_x = \begin{bmatrix} d_{x1} \\ d_{x2} \\ d_{x3} \end{bmatrix} = 1(t-10).\begin{bmatrix} \sin(0.5t) \\ 1(t)-1(t-4)+1(t-8.5)-1(t-13)+1(t-17.5) \\ \cos(t)+0.5\sin(3t) \end{bmatrix} \qquad (8.41)$$

$$d_y(t) = 1(t-10).\begin{bmatrix} 0 & 0 & 0 & 0 & \sin(t) & 0 \end{bmatrix}^T . \qquad (8.42)$$

Note that the generator rotor angles $\delta_i \quad i = 1,2,3$ are supposed to converge to the constant values, while the generator speed deviations from synchronicity $\omega_i \to 0 \quad i = 1,2,3$ in the case of nominal performance (without attack).

8.4.2.1.2 Simulation Results The MATLAB software is used to simulate the system. The simulated plant attacks $d_{x_1}, d_{x_2}, d_{x_3}$ are accurately recovered in finite time and are shown in Figures 8.18 – 8.20.

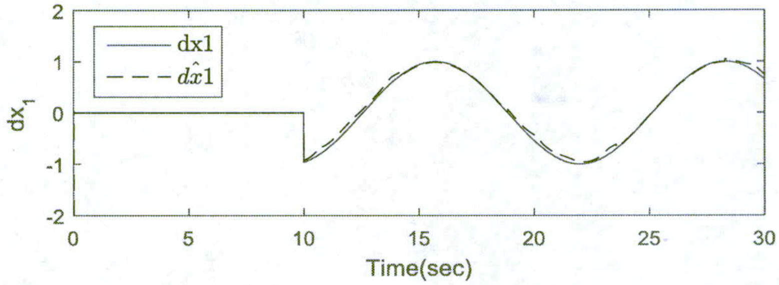


Figure 8.18 Plant Attack d_{x_1} Compare with Its Reconstruction $\hat{d}_{x_1}$

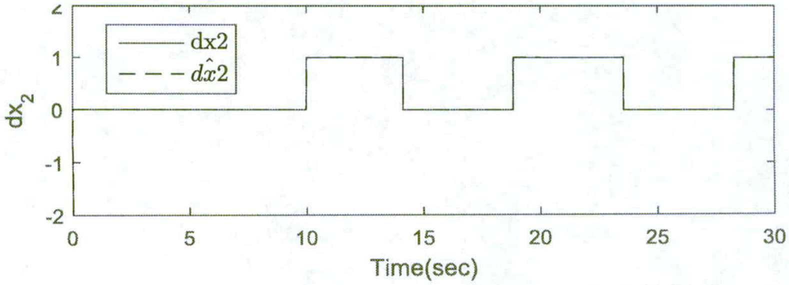


Figure 8.19 Plant Attack d_{x_2} Compare with Its Reconstruction $\hat{d}_{x_2}$

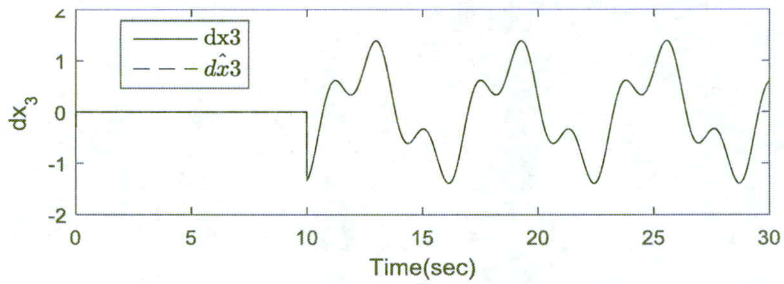


Figure 8.20 Plant Attack d_{x_3} Compare with Its Reconstruction $\hat{d}_{x_3}$

The reconstructed sensor attacks are shown in Figure 8.21.

The estimates $\hat{d}_x$ and $\hat{d}_y$ are used by the feedback control to compensate the state attacks, and to clean up the corrupted measurements respectfully. The results are depicted in Figures 8.22 and 8.23 where the WECC power system sensor measurements under attacks are compared to the sensor measurements without attacks, and the cleaned up sensor measurements.

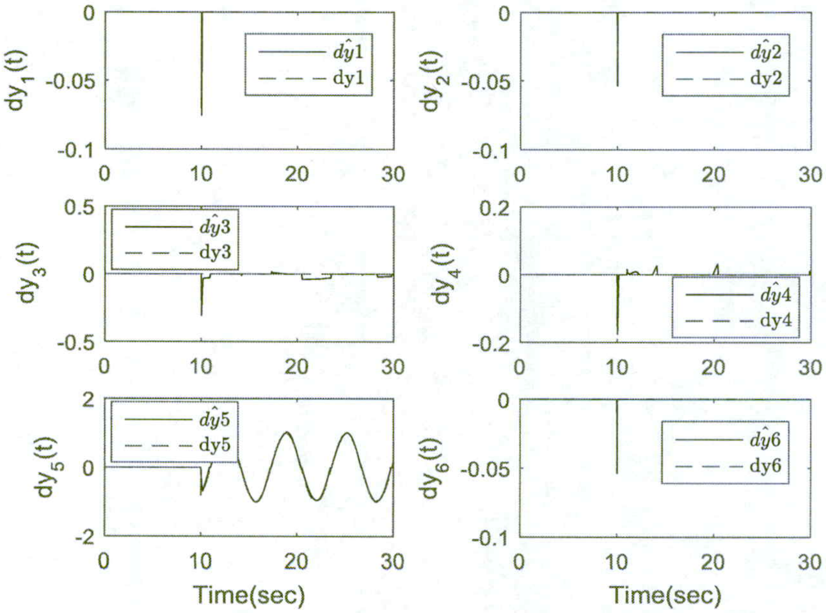


Figure 8.21 Sensor Attack d_y Reconstruction

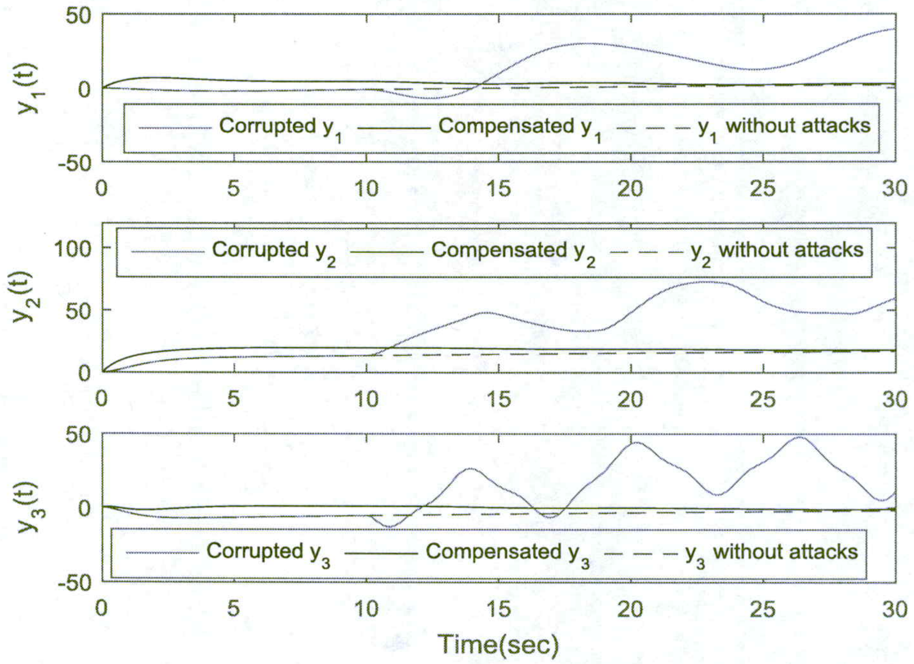


Figure 8.22 Corrupted WECC Power System Sensor Measurements y_1, y_2, y_3 Compared with the Compensated Measurements and to the Measurements without Attacks

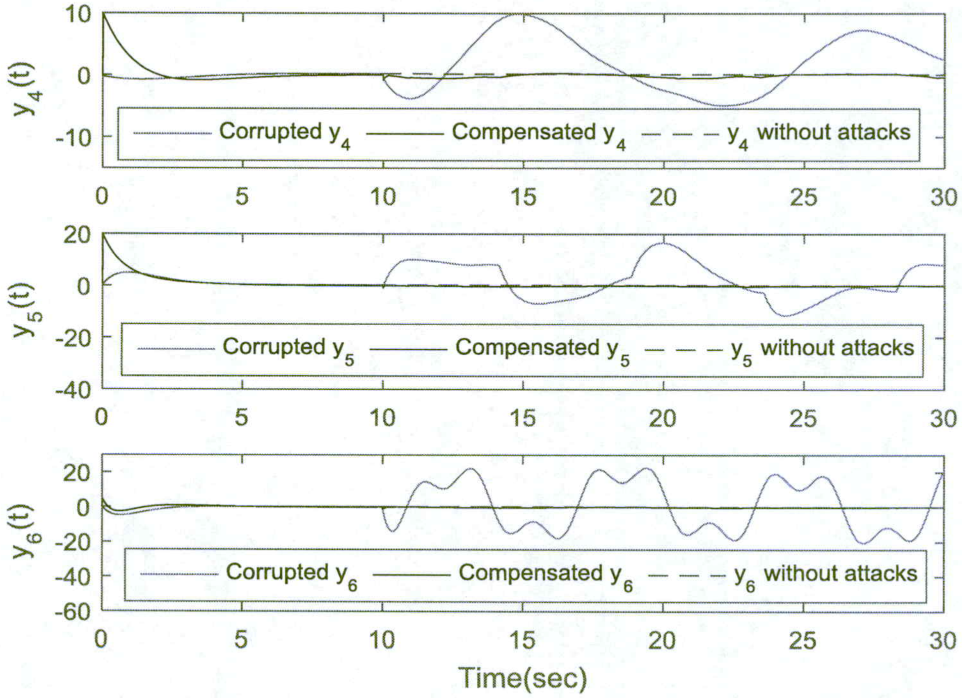


Figure 8.23 Corrupted WECC Power System Sensor Measurements y_4, y_5, y_6 Compared with the Compensated Measurements and to the Measurements without Attacks

8.4.3 Reconstruction of Attacks: the Number of Potential Attacks is Greater Than the Number of Sensors

In this study, the WECC power system eq. (8.1), is considered as a nonlinear electrical power system.

8.4.3.1 Sparse recovery Algorithm

It is assumed that the plant and the sensors are under attack, i.e. in the WECC power system eq. (8.1), is under attack signal $d = \begin{bmatrix} d_1^T & d_2^T \end{bmatrix}^T \in \mathbb{R}^{18}$ that corrupts the measurements $y = \begin{bmatrix} \delta \\ \omega \end{bmatrix} \in \mathbb{R}^6$, where $d_1 \in \mathbb{R}^{12}$, and $d_2 \in \mathbb{R}^6$ are the attacks of the plant and sensors respectively. Next, d_1, d_2 are further decoupled

$$d_1 = \begin{bmatrix} d_1^{\delta}{}_{(3\times1)} \\ d_1^{\omega}{}_{(3\times1)} \\ d_1^{\theta}{}_{(6\times1)} \end{bmatrix}, d_2 = \begin{bmatrix} d_2^{\delta}{}_{(3\times1)} \\ d_2^{\omega}{}_{(3\times1)} \end{bmatrix} \tag{8.43}$$

where d_1^{δ}, d_1^{ω}, d_1^{θ} are attacks on δ, ω, θ, and d_2^{δ}, d_2^{ω} are attacks on measurements of δ and ω respectively. Note that in this case study:

(a)
$$z = \begin{bmatrix} z_{1_{3\times1}} \\ z_{2_{3\times1}} \end{bmatrix} \in \mathbb{R}^6,\ \xi = \begin{bmatrix} z \\ \bar{x} \end{bmatrix} \in \mathbb{R}^{12},\ C = \begin{bmatrix} I_{6\times6} & 0_{6\times6} \end{bmatrix}. \tag{8.44}$$

(b) The matrices B and D in the WECC power system eq. (8.1) are defined as

$$\begin{aligned} B_{\delta} \in \mathbb{R}^{3\times18} &= \begin{bmatrix} I_{3\times3} & 0_{3\times15} \end{bmatrix} \\ B_{\omega} \in \mathbb{R}^{3\times18} &= \begin{bmatrix} 0_{3\times3} & I_{3\times3} & 0_{3\times12} \end{bmatrix} \\ B_{\theta} \in \mathbb{R}^{6\times18} &= \begin{bmatrix} 0_{6\times6} & I_{6\times6} & 0_{6\times6} \end{bmatrix} \\ D_{\delta} \in \mathbb{R}^{3\times18} &= \begin{bmatrix} 0_{3\times12} & I_{3\times3} & 0_{3\times3} \end{bmatrix}. \\ D_{\omega} \in \mathbb{R}^{3\times18} &= \begin{bmatrix} 0_{3\times15} & I_{3\times3} \end{bmatrix} \end{aligned} \tag{8.45}$$

The model of WECC power system eq. (8.1) with a LPF as it is written in eq. (6.4) can be presented as

$$\dot{\xi}=\begin{bmatrix}\frac{-1}{\tau} & 0 & \frac{1}{\tau} & 0\\ 0 & \frac{-1}{\tau} & 0 & \frac{1}{\tau}\\ 0 & 0 & 0 & 1\\ 0 & 0 & \underbrace{M_g^{-1}\left(-P_{g,g}^{\theta}+P_{g,l}^{\theta}\left(P_{l,l}^{\theta}\right)^{-1}P_{l,g}^{\theta}\right)}_{\varphi_{21}} & \underbrace{-M_g^{-1}E_g}_{\varphi_{22}}\end{bmatrix}\xi+\begin{bmatrix}\frac{1}{\tau}D_{\delta}\\ \frac{1}{\tau}D_{\omega}\\ B_{\delta}\\ B_{\theta\omega}\end{bmatrix}d+\begin{bmatrix}0\\0\\0\\ \underbrace{-M_g^{-1}P_{g,l}^{\theta}P_{l,l}^{\theta\,-1}P_{\theta}+M_g^{-1}P_{\omega}}_{P_{\theta\omega}}\end{bmatrix} \quad (8.46)$$

$$\psi=\begin{bmatrix}I_{6\times6} & 0_{6\times6}\end{bmatrix}\xi$$

Considering $\psi=\begin{bmatrix}\psi_1 & \psi_2\end{bmatrix}^T$ where $\psi_{1_{(3\times1)}}=z_{1_{(3\times1)}}$, $\psi_{2_{(3\times1)}}=z_{2_{(3\times1)}}$, then

$$\dot{z}_1=\frac{1}{\tau}\left(-z_1+\delta+d_2^{\delta}\right),\quad \dot{z}_2=\frac{1}{\tau}\left(-z_2+\omega+d_2^{\omega}\right). \quad (8.47)$$

In order to verify if the eq. (8.47) satisfies the RIP condition in Definition 3.1, eq. (3.2) or (3.3), model eq. (8.47) is rewritten in a format of eq. (6.10) as

$$\begin{bmatrix}\dot{z}_1+\frac{1}{\tau}z_1-\frac{1}{\tau}\delta\\ \dot{z}_2+\frac{1}{\tau}z_2-\frac{1}{\tau}\omega\end{bmatrix}=\underbrace{\begin{bmatrix}0_{3\times3} & 0_{3\times3} & 0_{3\times6} & \left(\frac{1}{\tau}\right)I_{3\times3} & 0_{3\times3}\\ 0_{3\times3} & 0_{3\times3} & 0_{3\times6} & 0_{3\times3} & \left(\frac{1}{\tau}\right)I_{3\times3}\end{bmatrix}}_{F(\xi)}\begin{bmatrix}d_1^{\delta}\\ d_1^{\omega}\\ d_1^{\theta}\\ d_2^{\delta}\\ d_2^{\omega}\end{bmatrix} \quad (8.48)$$

Apparently, $F(\xi)$ in (8.48) doesn't satisfy the RIP condition eq. (3.3), therefore another differentiation of $\dot{z}_1,\dot{z}_2$ is required:

$$\ddot{z}_1=\frac{1}{\tau}\left(-\dot{z}_1+\dot{\delta}+\dot{d}_2^{\delta}\right),\quad \ddot{z}_2=\frac{1}{\tau}\left(-\dot{z}_2+\dot{\omega}+\dot{d}_2^{\omega}\right) \quad (8.49)$$

Taking into account the output filter dynamics eq. (6.3), and bearing in mind that

$$\dot{\delta}=\omega+B_{\delta}d=\left(\tau\dot{z}_2+z_2-d_2^{\omega}\right)+B_{\delta}d \quad (8.50)$$

and

$$\begin{aligned}\dot{\omega} &= \varphi_{21}\delta + \varphi_{22}\omega + P_{\theta\omega} + B_{\theta\omega}d(t) \\ &= \varphi_{21}\left(\tau\dot{z}_1 + z_1 - d_2^{\delta}\right) + \varphi_{22}\left(\tau\dot{z}_2 + z_2 - d_2^{\omega}\right) + P_{\theta\omega} + B_{\theta\omega}d(t)\end{aligned} \tag{8.51}$$

where

$$B_{\theta\omega}d(t) = M_g^{-1}d_1^{\omega} - M_g^{-1}P_{g,l}^{\theta}\left(P_{l,l}^{\theta}\right)^{-1}d_1^{\theta} \tag{8.52}$$

then, eq. (8.49) is rewritten as

$$\tilde{Z}_m = \tilde{F}\tilde{d} \tag{8.53}$$

where

$$\tilde{Z}_m = \begin{bmatrix} \ddot{z}_1 + \frac{1}{\tau}\dot{z}_1 - \dot{z}_2 - \frac{1}{\tau}z_2 \\ \ddot{z}_2 + \frac{1}{\tau}\dot{z}_2 - \varphi_{21}\dot{z}_1 - \frac{1}{\tau}\varphi_{21}z_1 - \varphi_{22}\dot{z}_2 - \frac{1}{\tau}\varphi_{22}z_2 - \frac{1}{\tau}P_{\theta\omega} \end{bmatrix} \tag{8.54}$$

$$\tilde{F} = \begin{bmatrix} \frac{1}{\tau} & 0 & 0 & 0 & -\frac{1}{\tau} & \frac{1}{\tau} & 0 \\ 0 & \frac{M_g^{-1}}{\tau} & \frac{-M_g^{-1}P_{g,l}^{\theta}\left(P_{l,l}^{\theta}\right)^{-1}}{\tau} & \frac{-\varphi_{21}}{\tau} & \frac{-\varphi_{22}}{\tau} & 0 & \frac{1}{\tau} \end{bmatrix} \tag{8.55}$$

$$\tilde{d}_{24\times1} = \left[\left(d_1^{\delta}\right)^T \quad \left(d_1^{\omega}\right)^T \quad \left(d_1^{\theta}\right)^T \quad \left(d_2^{\delta}\right)^T \quad \left(d_2^{\omega}\right)^T \quad \left(\dot{d}_2^{\delta}\right)^T \quad \left(\dot{d}_2^{\omega}\right)^T\right]^T. \tag{8.56}$$

Now, the eq. (8.53) satisfies the RIP condition eq. (3.3), therefore the SR algorithm can be applied to eq. (8.53).

Remark 8.5 The derivatives $\ddot{z}_1, \ddot{z}_2, \dot{z}_1$ and $\dot{z}_2$ that appear in the entries of the virtual measurement vector $\tilde{Z}_m$ are obtained using HOSM differentiators [96].

Assumption (A 8.2) The sensor attack signals d_2^{δ} and d_2^{ω} are assumed to be slow with respect to system eq. (8.56) dynamics. In other words it is assumed $\dot{d}_2^{\delta} \approx 0$ and $\dot{d}_2^{\omega} \approx 0$.

Assumption (A 8.3) The attacks are assumed to be not-coordinated, and only two out of possible 18 attacks of following attack signal

$$d_{18\times 1} = \left[\left(d_1^{\delta}\right)^T \quad \left(d_1^{\omega}\right)^T \quad \left(d_1^{\theta}\right)^T \quad \left(d_2^{\delta}\right)^T \quad \left(d_2^{\omega}\right)^T \right]^T \tag{8.57}$$

are assumed to happen (it is not known which ones), the other 16 unknown attacks are assumed non-existent. These two attacks are recovered using the SR algorithm described in Section 3.2 applied to filtered WECC power system eq. (8.53).

8.4.3.2 Simulation Results

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

Simulation experiment 1: Two constant attacks $\left(d_1^{\omega}\right)_2 = -1$, which is the second entry of d_1^{ω}, and $\left(d_2^{\omega}\right)_1 = 1$ affected the filtered WECC power system eq. (8.46) at the time $t = 0.4\,\text{sec}$, and $\tau = 0.01$. The SR algorithm was used to recover the attacks. The results of the simulations are shown in Figure 8.24. The simulated two non-zero attacks, which are shown by dash line and dot line, are accurately recovered in finite time, while the estimated values of other zero attacks, which are shown by solid lines, converge to zero in finite time. In Figures 8.24 – 8.26, Attack1 and Attack2 are used to describe the real attack signals and $d1 - d18$ display the reconstructed plant and sensor attacks.

Simulation experiment 2: Two time-varying attacks, $\left(d_1^{\omega}\right)_1 = \sin(\pi t)$ and $\left(d_1^{\omega}\right)_2 = -\sin(\pi t)$ affect the filtered WECC power system eq. (8.37) at the time $t = 0.4\,\text{sec}$. The simulation results are shown in Figure 8.25. The simulated two time-varying non-zero attacks are accurately recovered in finite time, which are illustrated by dash line and dot line, while the estimated values of other 16 zero attacks appear to converge to zero in finite time. The solid lines illustrate them.

Simulation experiment 3: Two non-zero attacks attaks are generated and affected the filtered WECC power system eq. (8.37) at the time $t = 0.4\,\text{sec}$, the plant attack is time varying $\left(d_1^{\omega}\right)_2 = \sin\left(\pi t\right)$, and sensor attack is constant $\left(d_2^{\omega}\right)_1 = -1$. The simulation result in Figure 8.26 shows that 2 non-zero and 16 zero attacks were accurately recovered in finite time.

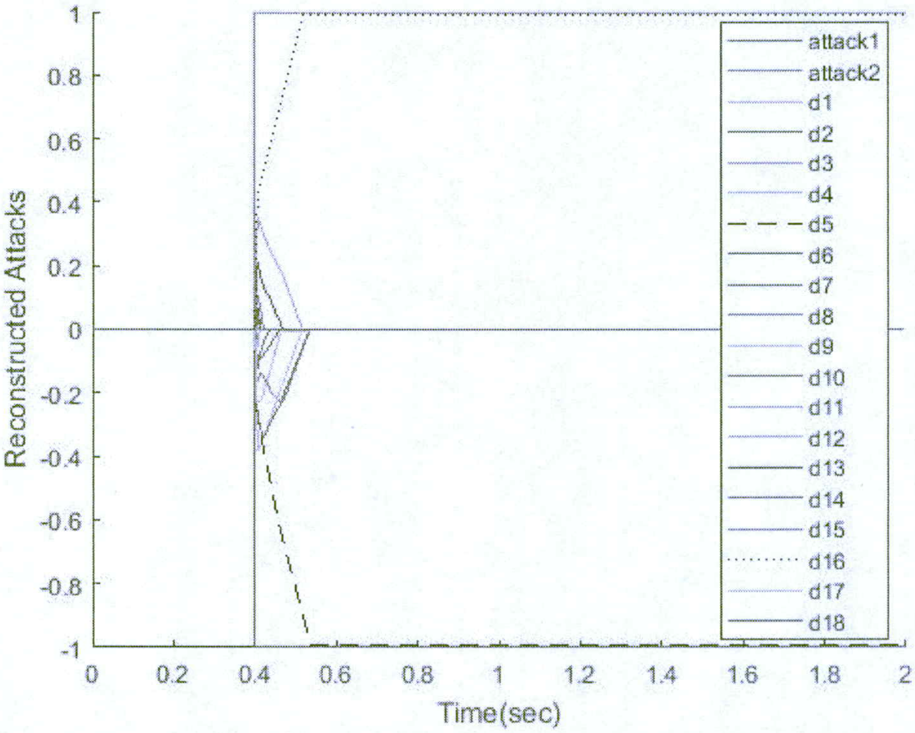
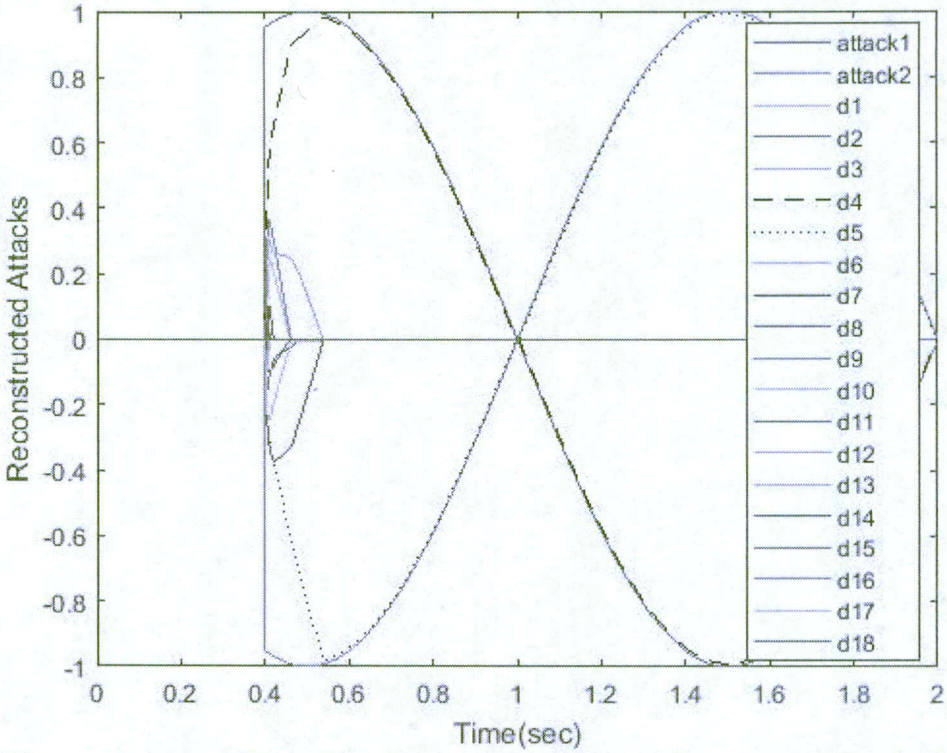
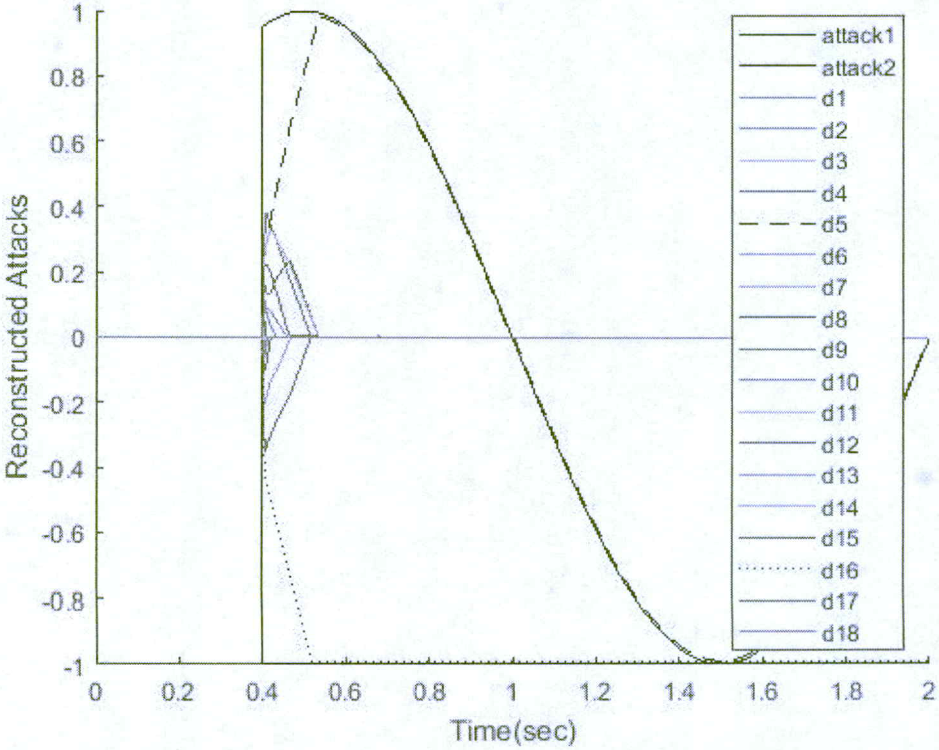


Figure 8.24 Reconstruction of Two Constant Plant Attack and Sensor Attak in a Sparse Attack Signal



8.25 Reconstruction of Two Time Varying Plant Attack in a Sparse Attack Signal



8.26 Reconstruction of Time Varying Plant Attack and Constant Sensor Attack in a Sparse Attack Signal

The Simulation results in Figures 8.24-8.26 show that SR algorithm can reconstruct the time varying sparse attack signal.in finite time.

8.4.4 Reconstruction of Attacks and Estimation of States: the Number of Sensors is Greater Than the Number of Potential Sensor Attacks

In this section, we investigate the WECC power system eq. (8.10) as a nonlinear system when we have more sensors rather than potential sensor attacks, i.e. there are 6 sensor measurements and 3 plant attacks. The matrices B and D in eq. (8.1) are defined in such a way that plant attack d_x and sensor attack d_y can be written separately as follows

$$\begin{bmatrix} I & 0 & 0 \\ 0 & M_g & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} \dot{\delta} \\ \dot{\omega} \\ \dot{\theta} \end{bmatrix} = -\begin{bmatrix} 0 & -I & 0 \\ R^{\theta}_{g,g} & E_g & R^{\theta}_{g,l} \\ R^{\theta}_{l,g} & 0 & R^{\theta}_{l,l} \end{bmatrix}\begin{bmatrix} \delta \\ \omega \\ \theta \end{bmatrix} + \begin{bmatrix} 0 \\ I \\ 0 \end{bmatrix} d_x(t) + \begin{bmatrix} 0 \\ P_{\omega} \\ P_{\theta} \end{bmatrix}$$
$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} C_{\delta} & 0 \\ 0 & C_{\omega} \end{bmatrix}\begin{bmatrix} \delta \\ \omega \end{bmatrix} + \begin{bmatrix} D_{\delta} \\ D_{\omega} \end{bmatrix} d_y(t) \tag{8.58}$$

where

$$C_{\delta} = I_3, \quad C_{\omega} = I_3, \quad D_{\delta} = 0_{3\times 6}, \quad D_{\omega} \in \mathbb{R}^{3\times 6} = \begin{bmatrix} 0 & 1 & 2 & 0 & 1 & 1 \\ 1 & 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}. \tag{8.59}$$

The WECC power system eq. (8.58) can be rewritten as

$$\begin{cases} \begin{bmatrix} \dot{\delta} \\ \dot{\omega} \end{bmatrix} = \begin{bmatrix} \omega \\ M_g^{-1}\left(-R^{\theta}_{g,g} + R^{\theta}_{g,l}\left(R^{\theta}_{l,l}\right)^{-1} R^{\theta}_{l,g}\right)\delta - M_g^{-1}E_g\omega + P_{\theta\omega} \end{bmatrix} + \bar{B}d_x(t) \\ \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \bar{C}_{\delta} \\ \bar{C}_{\omega} \end{bmatrix}\begin{bmatrix} \delta \\ \omega \end{bmatrix} + \begin{bmatrix} 0 \\ D_{\omega} \end{bmatrix} d_y(t) \end{cases} \tag{8.60}$$

where

$$\begin{aligned}&P_{\theta\omega}=M_g^{-1}\left(P_\omega-L_{g,l}^{\theta}\left(L_{l,l}^{\theta}\right)^{-1}P_\theta\right),B_{\theta\omega}=M_g^{-1}\left(B_\omega-L_{g,l}^{\theta}\left(L_{l,l}^{\theta}\right)^{-1}B_\theta\right)\\&\bar{C}_\delta=\begin{bmatrix}I_3 & 0_3\end{bmatrix},\ \ \bar{C}_\omega=\begin{bmatrix}0_3 & I_3\end{bmatrix},\ \ \bar{B}=\begin{bmatrix}0_3\\ M_g^{-1}\end{bmatrix}\end{aligned}\quad . \tag{8.61}$$

Remark 8.6: It can be verified that D_ω satisfies RIP condition defined in eq. (3.3).

Suppose that the following three plant attacks

$$d_x=\begin{bmatrix}d_{x1}\\ d_{x2}\\ d_{x3}\end{bmatrix}=(t-10)\begin{bmatrix}\sin(0.5t)\\ 0.5\cos(0.5t)\\ 0.5\sin(0.5t)+0.5\cos(0.5t)\end{bmatrix} \tag{8.62}$$

and the time-varying sensor attack

$$d_y=1(t-10)\cdot\begin{bmatrix}0 & 0 & 0 & 0.5\cos(0.5t) & 0 & 0\end{bmatrix} \tag{8.63}$$

affect system eq. (8.58) at $t=10$ sec.

The states $\hat{\delta}$, $\hat{\omega}$ and plant attacks $d_x(t)$ in eq. (8.62) are reconstructed by using HOSM observer as described in section 7.2. Then, the estimated $\hat{\omega}$ is used in eq. (8.58) to get

$$y_2-\hat{\omega}=D_\omega d_y(t). \tag{8.64}$$

The SR algorithm described in Section 3.1 can be applied to reconstruct the sparse $d_y(t)$ in WECC power system eq. (8.58), where only one out of six potential attacks $d_{y1},\ldots,d_{y6}$ is non-zero.

8.4.4.1 Simulation Results The MATLAB software is used to simulate the system. The simulated plant attacks d_{x_1},d_{x_2},d_{x_3} and sensor attack $d_{y1},\ldots,d_{y6}$ are accurately recovered in finite time and are shown in Figure 8.27 and 8.28 respectively. Reconstructed attacks are used foe cleaning the corrupted plant and measurements. Figure 8.29 and 8.30 compare corrupted measurements with compensated and without attack measurements.

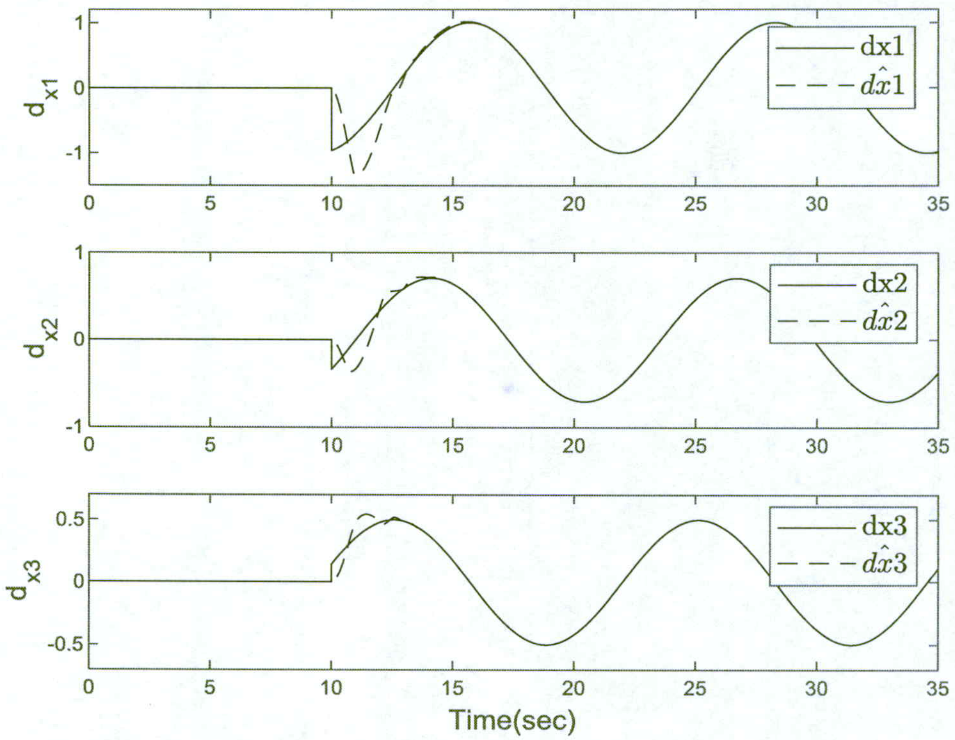


Figure 8.27 Plant Attack $d_{x_1}, d_{x_2}, d_{x_3}$ Compare with Its Reconstruction $\hat{d}_{x_1}, \hat{d}_{x_2}, \hat{d}_{x_3}$

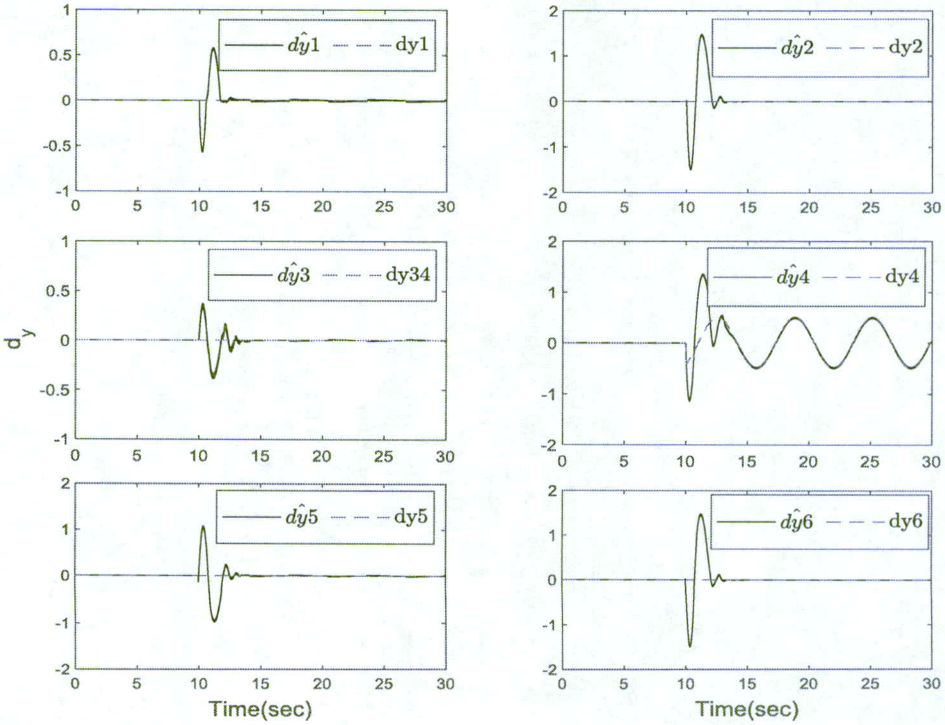


Figure 8.28 Sensor Attack d_y Reconstruction

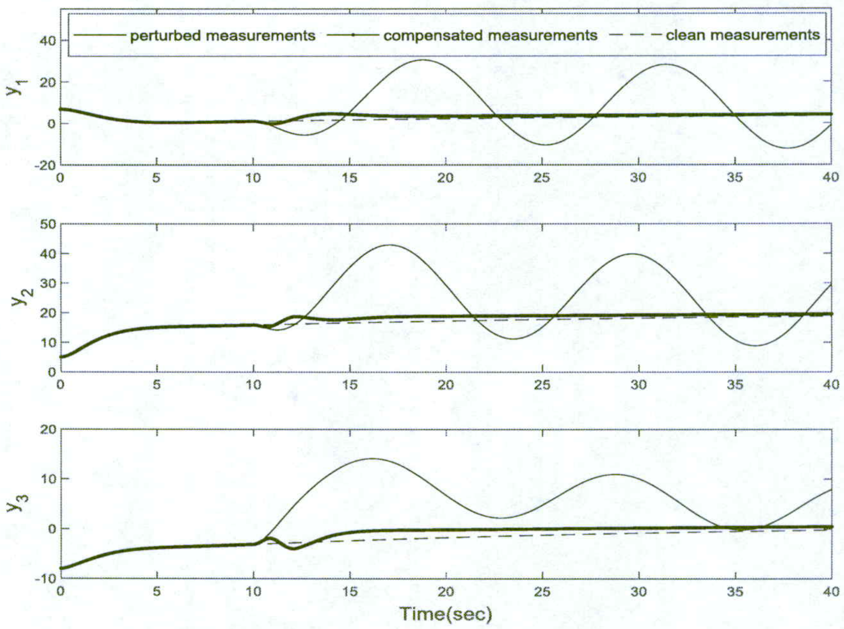


Figure 8.29 Corrupted WECC Power System Sensor Measurements y_1, y_2, y_3 Compared with the Compensated Measurements and to the Measurements without Attacks

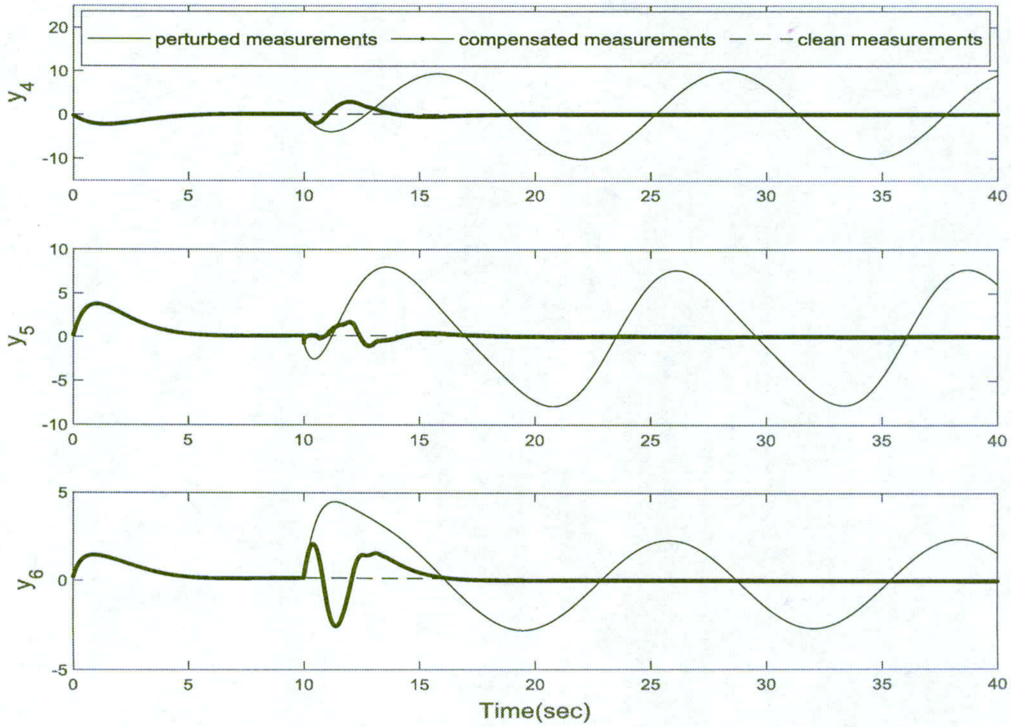


Figure 8.30 Corrupted WECC Power System Sensor Measurements y_4, y_5, y_6 Compared with the Compensated Measurements and to the Measurements without Attacks.

Therefore, simulation results illustrate that compensated measurements converge to the measurements without attack in finite time. As a result, actual measurements are recovered from corrupted ones in finit time by using the HOSM observer anf SR algorithm.

8.5 **Summary**

The effectiveness of the proposed algorithms in this dissertation to estimate the states and reconstruct the attacks are tested on the WECC power network system. The Simulation results confirm that the attacks degrade the performance of CPS under attack.

Illustrated figures imply that cleaning the measurements from the reconstructed attacks before using them in the feedback control can elevate CPS performance close to the one without attack.

CHAPTER 9

CONCLUSIONS AND FUTURE WORKS

9.1 Conclusions

Cyber Physical Systems (CPS) represent to the embedding of sensing, computation, communication and control into physical systems. Exchanging data among sensors, actuators and other networked components in common or wireless communication setting makes it possible for attackers to find access to sensing and actuation computing platforms and manipulate system measurements and control commands to severely compromise system performance.

There is a long list of publications which have focused on keeping the system safe from being attacked. However, how to ensure the CPS control system can continue functioning properly if attacks occur is another serious problem.

The literature that study the resilience-increasing mechanism for CPSs are mostly based on Game theory, Event- triggered Control, Mean Subsequence Reduced algorithms, and Trust-based approaches. The disadvantages of these works include the specific type of attack on the cyber layer is considered, the special structure of the CPS is investigated.

To address these challenges, SMC and HOSM control and observation algorithms are proposed in this dissertation to estimate the states of a CPS under the sensor and state attacks and reconstruct the attacks with arbitrary shape in finite time or asymptotically. By cleaning the measurements and compensating the state attacks by means of feedback control the CPS performance can remain as it was demonstrated prior to attacks.

This work has three major contributions listed as follow:

I. A novel observation algorithm based on a SR technique along with a sliding mode differentiator is proposed for reconstructing on-line the sparse attacks on nonlinear CPS when there are more potential attacks than sensors. The novel result of this work is presented in [79].

II. A new approach for on-line plant attack reconstruction and state estimation of a nonlinear CPS in finite time when the number of sensors is greater than the number of potential sensor attacks is proposed based on HOSM observer and differentiator. The result of this work is published in [80].

III. A novel SMO that includes the dynamic extension of the injection term is developed for the first time for on-line state estimation and attack reconstruction in a linearized CPS when the number of sensors is greater or equal to the number of potential attacks. Specifically, a novel adaptive sliding mode observation algorithm that reconstructs the smooth bounded attacks with unknown boundaries on their amplitude and rates is proposed. This novel dynamic filter that addresses the attack propagation dynamics is presented in [81, 82].

The proposed methodologies in this research are applied to the WECC power network system, whose sensors and/or states are under attack. Simulation results illustrate the efficacy of the developed observers for state estimation and attack reconstruction in CPSs.

9.2. Future Work

As far as we know, there are limited number of works which investigate the resilient control of nonlinear CPS. Attack reconstruction and compensation in nonlinear CPS is a good problem to work on in the future.

False-data injection attacks can be formulated against systems with unstable modes, and they aim to modify the system measurements to make some unstable modes unobservable. As far as we know, there is no solution to protect a CPS against false-data injection attack. We plan to work on in the future.

It is not always easy to compensate the attacks after finding their estimation. It can be considered as a future work as well.

APPENDIX A

A.1 Proof of Theorem 4.1 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 eq. (A.1) consider a following Lyapunov function candidate

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

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 (A 4.4), the derivative of the Lyapunov function candidate eq. (A.2) is given as

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

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 eq. (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 eq. (A.5) by taking Laplace transform and solving for d, we obtain the estimate $\hat{d}$ given by eq. (4.15) The theorem is proven [81].

A.2 Proof of Proposition 4.1: Consider the e_y dynamics from second equation of eq.(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 eqs. (4.22) - (4.27) 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 eqs. (4.22) - (4.27) follows the one of the Proposition 2 in [99, pp. 185-186]. Convergence $e_y \to 0$ in finite time yields eq. (A.5) and then eq. (4.15). Therefore, $d(t)$ is reconstructed as in eq. (4.15) with the adaptive injection term υ_{eq} or $\bar{\upsilon}_{eq}$. The proposition is proven [81].

A.3 Proof of Theorem 4.2: Taking into account eqs. (4.31), (4.34), and (4.36), 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 eq. (A.8) in order to prove the convergence $e_{y_1} \to 0$ in finite time.

The derivative of the Lyapunov function eq. (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 $\|\varphi_1\| \le L_{11}$ at least locally, where $L_{11} > 0$ is known, we obtain

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

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

Next, introduce a Lyapunov function candidate

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

that is applied to the third equation in eq. (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 $\|\varphi_2\| \le L_{12}$ 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 eq.(4.35). 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 eq. (4.35).

Then, the estimation error dynamics eq. (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 eq. (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 eq.(A.16) is estimated by $\hat{d} = \begin{bmatrix} \hat{d}_1 \\ \hat{d}_2 \end{bmatrix}$ as in eq. (4.37). Theorem 4.2 is proven [81].

A.4 Proof of Theorem 5.1

To develop a theory for selecting the gains to enforce sliding, consider the signals $\bar{e}_{11}$ and ϕ as defined in eq. (5.24) together with the dynamics

$$\dot{\bar{e}}_{11} = \begin{bmatrix} \bar{A}_{11} & 0 \\ \bar{A}_{21a} & A_{22}^{s} \end{bmatrix} \bar{e}_{11} - \begin{bmatrix} \bar{A}_{12b} - \bar{B}_1 D_2^{-1} \\ \bar{A}_{22b} - \bar{B}_{21} D_2^{-1} \end{bmatrix} D_2 d \tag{A.17}$$

then the transfer function matrix $G_{\phi}(s)$ mapping $d \mapsto \phi$ given by eq. (A.17) and eq. (5.24) is asymptotically stable since both $\bar{A}_{11}$ and A_{22}^{s} are Hurwitz. Additionally, since it is assumed that attacks are bounded, $\|d(t)\| < k_d$, for any initial condition $\bar{e}_{11}(0)$ at time $t = 0$, there exists a gain m_0 such that

$$\|\phi(t)\| \le m_0 k_d \tag{A.18}$$

for all $t \ge t_0$, where t_0 is finite reaching time.

The gain m_0 will now be employed in the modulation gain definition: specifically, chose the modulation gain in eq. (5.18) according to

$$\rho = m_0 k_d + \|D_2\| l_d \tag{A.19}$$

Define the Lyapunov function candidate $V = \frac{1}{2}e_{y_2}^T e_{y_2} = \frac{1}{2}\|e_{y_2}\|^2$ and the output estimation error $e_{y_2} = \bar{y}_2 - \bar{z}_{22}$, then it is easy to verify from eqs. (5.21) and (5.22) that

$$\dot{e}_{y_2} = \begin{bmatrix} \bar{A}_{21b} & \bar{A}_{22c} \end{bmatrix} \bar{e}_{11} - \left(\bar{A}_{22d} - \bar{B}_{22} D_2^{-1}\right) D_2 d + A_{33}^s e_{y_2} + D_2 \dot{d} + \upsilon \tag{A.20}$$

Replace ϕ defined in eq. (5.24), in eq. (A.20), then it is simplified to

$$\dot{e}_{y_2} = A_{33}^s e_{y_2} + \phi + D_2 \dot{d} + \upsilon \tag{A.21}$$

Since A_{ss}^3 is symmetric negative definite, for all $t \ge t_0$

$$e_{y_2}^T \dot{e}_{y_2} \le e_{y_2}^T \phi + e_{y_2}^T D_2 \dot{d} - (\rho + \eta)\|e_{y_2}\| \le \|e_{y_2}\|\left(\phi + \|D_2\|\|\dot{d}\| - (\rho + \eta)\right) \tag{A.22}$$

where $\|\phi\| < m_0 k_d$. Thus by choice of ρ in eq. (A.19) it follows

$$\dot{V} = e_{y_2}^T \dot{e}_{y_2} \le -\eta\|e_{y_2}\| = -\eta\sqrt{2}V^{1/2} \tag{A.23}$$

and a sliding motion is guaranteed in finite time.

On the sliding surface $e_{y_2} = \bar{y}_2 - \bar{z}_{22} = 0$, it follows from eq. (5.21) that $\bar{e}_{22} = -D_2 d$ and $\dot{\bar{e}}_{22} = -D_2 \dot{d}$.

After the system collapse during the sliding motion $e_{y_2} = \dot{e}_{y_2} = 0$, it is obtained from the last row of eq. (5.22) that

$$0 = \begin{bmatrix} \bar{A}_{21b} & \bar{A}_{22c} \end{bmatrix} \bar{e}_{11} - \left(\bar{A}_{22d} - \bar{B}_{22} D_2^{-1}\right) D_2 d + D_2 \dot{d} + \upsilon_{eq} \tag{A.24}$$

where υ_{eq} is the equivalent injection necessary to induce sliding mode. Defining

$\bar{d} = -D_2 d$ follows that eqs. (A.17) and (A.24) can be written as

$$\begin{bmatrix} \dot{\bar{e}}_1 \\ \dot{\bar{e}}_{21} \\ \dot{\bar{d}} \end{bmatrix} = \begin{bmatrix} \bar{A}_{11} & 0 & \bar{A}_{12b} - \bar{B}_1 D_2^{-1} \\ \bar{A}_{21a} & A_{22}^{s} & \bar{A}_{22b} - \bar{B}_{21} D_2^{-1} \\ \bar{A}_{21b} & \bar{A}_{22c} & \bar{A}_{22d} - \bar{B}_{22} D_2^{-1} \end{bmatrix} \begin{bmatrix} \bar{e}_1 \\ \bar{e}_{21} \\ \bar{d} \end{bmatrix} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ \mathrm{I}_{\mathrm{m}} \end{bmatrix}}_{B^*} \upsilon_{eq} \tag{A.25}$$

where the last equation in eq. (A.25) comes from rearranging eq. (A.24). Also by definition

$$\hat{d} = \underbrace{\begin{bmatrix} 0 & -D_2^{-1} \end{bmatrix}}_{C^*} \begin{bmatrix} \bar{e}_1 \\ \bar{e}_{21} \\ \bar{d} \end{bmatrix} \tag{A.26}$$

Therefore during the sliding motion, the signal v and $\hat{d}$ are connected via the transfer function $\hat{d} = G^*(s)\upsilon_{eq}$ which is defined in eqs. (5.25) - (5.28). If A^* in eq. (5.19) is Hurwitz then $G^*(s)$ represents a stable low pass filter. Although d is unknown, from eq. (5.25), it can be recovered from filtering the discontinuous injection υ through $G^*(s)$.

A.5 Proof of Proposition 5.1:

Consider the Lyapunov function as

$$V = \frac{1}{2}\sigma^2 + \frac{1}{2\gamma}\varphi^2 \tag{A.27}$$

where $\gamma > 0$, σ is defined in eq. (5.35), and

$$\varphi(t) = \left(qI_d / \alpha\right) - \ell(t) \tag{A.28}$$

Then the proof of the finite time convergence $e_y \to 0$ in eq. (5.23) by the adaptive injection term υ in (5.33)-(5.38) follows the one of the Proposition 2 in [99], pp. 185-186]. Therefore, $d(t)$ is reconstructed as in (5.31) with the adaptive injection term υ_{eq} in eq. (5.33). The proposition is proven [82].

REFERENCES

[1] P. Antsaklis, "Goals and challenges in cyber-physical systems research," *IEEE Transactions on Automatic Control,* vol. 59, issue 12, pp. 3117–3119, 2014.

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

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

[4] L. Shi, Q.Dai, Y. Ni, "Cyber–physical interactions in power systems: A review of models, methods, and applications," *Electric Power Systems Research*, vol. 163, Part A, pp. 396-412, October 2018.

[5] S. Amin, X. Litrico, S. S. Sastry, and A. M. Bayen, "Stealthy deception attacks on water SCADA systems," *In Proceeding of Hybrid Systems: Computation Control*, pp. 161–170, Stockholm, Sweden, April 2010.

[6] S. Dadras, S. Dadras, C. Winstead, "Identification of the attacker in cyber-physical systems with an application to vehicular platooning in adversarial environment," *Annual American Control Conference*, pp. 5560-5567, 2018.

[7] A Mitra, JA Richards, S Bagchi, S Sundaram, " Resilient distributed state estimation with mobile agents: overcoming Byzantine adversaries, communication losses, and intermittent measurements," Autonomous Robots, vol. 43, num. 3, pp. 743-768.

[8] E. Hashemi, M. Pirani, K. Amir, B. Fidan, S. Shen, B. Litkouhi, "Fault Tolerant Consensus for Vehicle State Estimation: A Cyber-physical Approach," *EEE Transactions on Industrial Informatics,* vol. 15 , issue 9, pp. 5129 – 5138, 2019.

[9] E. Byres and J. Lowe, "The myths and facts behind cyber security risks for industrial control systems." VDE Congress, 2004.

[10] E. Mousavinejad, F. Yang, Q. L. Han, Q. Qiu, and L. Vlacic, "Cyber Attack Detection in Platoon-Based Vehicular Networked Control Systems," *IEEE 27th International Symposium on Industrial Electronics (ISIE*), pp. 603-608, 2018.

[11] A. Farhat, and C. K. Cheok, "Improving adaptive network fuzzy inference system with Levenberg-Marquardt algorithm," *2017 Annual IEEE International Systems Conference (SysCon)*, April 2017, Montreal, Canada, pp. 1-6, 2017.

[12] A. Farhat, K. Hagen, K. C. Cheok, and B. Boominathan, "Neuro-fuzzy-based electronic brake system modeling using real time vehicle data," *EPiC Series in Computing. Proceedings of 34th International Conference on Computers and Their*

Applications, vol. 58, pp. 444-453, 2019.

[13] J. P. Conti, "The day the samba stopped," *Engineering & Technology,* vol. 5, no.4, pp. 46-47, 2010.

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

[15] J. Slay, and M. Miller, "Lessons learned from the Maroochy water breach," In *Proceedings of international conference on critical infrastructure protection,* pp. 73–82, 2007.

[16] D. U. Case, "Analysis of the cyber attack on the Ukrainian power grid," *Electricity Information Sharing and Analysis Center (E-ISAC), 2016.*

[17] A. Greenberg, Hackers remotely kill a jeep on the highway with me in it, Accessed: 2018-12-16. https://www.wired.com/2015/07/hackers-remotely-kill- jeep-highway, 2018.

[18] J. Harding, G. Powell, R. Yoon, J. Fikentscher. C. Doyle, D. Sade, et al, "Vehicle-to-vehicle communications: Readiness of V2V technology for application," *Technical report*. United States. National Highway Traffic Safety Administration, https://rosap.ntl.bts.gov/view/dot/27999, 2014.

[19] K. Hartmann, and C. Steup, "The vulnerability of UAVs to cyber attacks An approach to the risk assessment," In *Proceedings of international conference on cyber conflict,* pp. 1–23, 2013.

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

[21] N. W. Group. Internet security glossary. http://rfc.net/rfc2828.html, May 2000.

[22] Y. Mo and B. Sinopoli, "Secure control against replay attacks," *in*: *47th Annual Allerton Conference on Communication, Control, and Computing*, pp. 911–918, Monticello, IL, USA, Sep. 2009,

[23] M. Zhu and S. Martinez, "On the performance analysis of resilient networked control systems under replay attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 804–808, 2014.

[24] A. Khazraei, H. Kebriaei, R. F. Salmasi, "Replay attack detection in a multi agent system using stability analysis and loss effective watermarking," in: *Annual American Control Conference,* pp. 4778-4783, Seattle, WA, USA, May 2017.

[25] F. Miao, M. Pajic, G. J. Pappas, "Stochastic game approach for replay attack

detection," 52nd IEEE Conference on Decision and Control, pp. 1854-1859, Florence, Italy, 2013.

[26] R. Smith, "A decoupled feedback structure for covertly appropriating network control systems," *IFAC Proceeding Volumes*, vol.44, issue 1, pp. 90–95, 2011.

[27] A. O. Sá, L. F. Carmo, R. C. S. Machado, "Covert Attacks in Cyber-Physical Control Systems," *IEEE Transactions on Industrial Informatics*, vol. 13, issue 4, pp. , 2018.

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

[29] Md. A. Rahman, H. Mohsenian-Rad, "False data injection attacks with incomplete information against smart power grids," *2012 IEEE Global Communications Conference (GLOBECOM),* pp. 3153-3158, Anaheim, CA, USA, Dec. 2012.

[30] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security*, vol. 14, issue 1, Article no. 13, pp. 1-33, May 2011.

[31] N. Hashemi, C. Murguia, J. Ruths, "A Comparison of Stealthy Sensor Attacks on Control Systems. In: American Control Conference," pp. 973-979., Milwaukee, USA, July 2018.

[32] 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.

[33] L. Cazorla, C. Alcaraz, J. Lopez, "Cyber Stealth Attacks in Critical Information Infrastructures," *IEEE Systems Journal*, vol. 12, issue 2 , pp. 1778 – 1792, 2018.

[34] E. M. Rudd, A. Rozsa, M. Günther, and T. E. Boult, "A Survey of Stealth Malware Attacks, Mitigation Measures, and Steps Toward Autonomous Open World Solutions," *IEEE Communications Surveys & Tutorials*, vol. 19 , issue 2 , pp. 1145-1172, 2017.

[35] M. S. Haghighi, F. Farivar, A. Jolfaei, and M. H. Tadayon, "Intelligent robust control for cyber-physical systems of rotary gantry type under denial of service attack," *The Journal of Supercomputing*, Springer, pp.1-23, 2019.

[36] V. D. Gligor, "A note on denial-of-service in operating systems," *IEEE Transactions on Software Engineering*, pp. 320–324, 1984.

[37] S. Amin, A. Cárdenas, S. Sastry, "Safe and secure networked control systems under denial-of-service attacks," Hybrid Systems: Computation and Control, Springer Berlin/Heidelberg, pp. 31-45, 2009.

[38] E. Nozari, P. Tallapragada, and J. Cortés, "Differentially private average consensus: Obstructions, trade-offs, and optimal algorithm design," *Automatica,* vol. 81, pp.

221–231, 2017.

[39] Y. Mo, and R. M. Murray, "Privacy preserving average consensus," *IEEE Transactions on Automatic Control,* vol. 62, issue 2, pp. 753–765, 2017.

[40] C. Dwork, "Differential privacy," In *Encyclopedia of cryptography and security*," pp. 338–340. Springer, 2011.

[41] S. M. Dibaji, M. Pirani, D. B. Flamholz, A. M. Annaswamy, K. H. Johansson, A. Chakrabortty, "A systems and control perspective of CPS security," *Annual Reviews in Control*, vol. 47, pp. 394-411, 2019.

[42] N Hashemi, J Ruths, "Generalized chi-squared detector for LTI systems with non-Gaussian noise," *2019 American Control Conference*, pp. 404-410, 2019.

[43] S. H. Kafash, J Giraldo, C Murguia, AA Cardenas, J Ruths, "Constraining attacker capabilities through actuator saturation," *Annual American Control Conference*, pp. 986-991, Milwaukee, USA, July 2018.

[44] F. Pasqualetti F., F. Dorfler, F. Bullo , "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, pp. 2715–2729, 2013.

[45] A. Jafarnia-Jahromi, A. Broumandan, J. Nielsen, and G. Lachapelle, "GPS vulnerability to spoofing threats and a review of antispoofing techniques," *International Journal of Navigation and Observation* , pp. 1–16, 2012.

[46] 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*, vol. 35, issue 1, pp. 110-127, 2015.

[47] S. H. Kafash, N Hashemi, C Murguia, J Ruths, "Constraining Attackers and Enabling Operators via Actuation Limits," *IEEE Conference on Decision and Control* (CDC), pp. 4535-4540, Miami beach, USA, 2018.

[48] J. Giraldo, A. A. Cardenas, "Moving Target Defense for Attack Mitigation in Multi-Vehicle Systems," Proactive and Dynamic Network Defense, Springer, Cham, pp. 163-190, 2019.

[49] A. Gusrialdi and Z. Qu, Stoustrup J., Annaswamy A.M., Chakrabortty A., Qu Z. (Eds.), "Smart grid security: Attacks and defenses," Smart grid control, Springer, 2019.

[50] A. Teixeira, K.C. Sou, H. Sandberg and K.H. Johansson, "Secure control systems: A quantitative risk management approach, *IEEE Control Systems Magazine,* vol. 35, issue 1, pp. 24-45, 2015.

[51] Y. Lu, C.-Y. Chang, W. Zhang, L.D. Marinovici and A.J. Conejo, "On resilience

analysis and quantification for wide-area control of power systems," *Proceedings of IEEE conference on decision and control*, pp. 5799-5804, 2016.

[52] Z. Askarzadeh, R. Fu, A. Halder, Y. Chen, T. T. Georgiou, "Opinion Dynamics over Influence Networks," *American Control Conference,* pp. 1873-1878, Philadelphia, USA, 2019.

[53] E. Nekouei, M. Skoglund, K. H. Johansson, "Privacy of information sharing schemes in a cloud-based multi-sensor estimation problem," *Annual American Control Conference*, pp. 998-1002, 2018.

[54] M. Pajic, I. Lee, ad G. J. Pappas, "Attack-resilient state estimation for noisy dynamical systems," *IEEE Transactions on Control of Network Systems,* vol. 4, issue 1, pp. 82–92, 2017.

[55] E. Nekouei, T. Tanaka, M. Skoglund, K. H. Johansson, "Information-theoretic approaches to privacy in estimation and control," *Annual Reviews in Control*, vol. 47, pp. 412-422, 2019.

[56] S. Chen, M. Ma, and Z. Luo, "An authentication scheme with identity-based cryptography for M2M security in cyber-physical systems," *Security and Communication Networks*, vol. 9, issue 10, pp. 1146–1157, 2016.

[57] W. Diffie, and M. Hellman, "New directions in cryptography," *IEEE Transactions on Information Theory,* vol. 22, issue 6, pp. 644–654, 1976.

[58] F. Farokhi, I. Shames, and N. Batterham, "Secure and private control using semi-homomorphic encryption," *Control Engineering Practice,* vol. 67, pp. 13–20, 2017.

[59] S. M. Dibaji, H. Ishii, and R. Tempo, "Resilient randomized quantized consensus," *IEEE Transactions on Automatic Control,* vol. 63, issue 8, pp. 2508–2522, 2018.

[60] T. Chen, "Stuxnet, the real start of cyber warfare?[Editor's note] ," *IEEE Network,* vol. 24, issue 6, pp. 2–3, 2010.

[61] Q. Zhu, and T. Basar, "Robust and resilient control design for cyber-physical systems with an application to power systems," In *Proceedings of IEEE conference on decision and control,* pp. 4066–4071, 2011.

[62] W. Heemels, K. H. Johansson, and P. Tabuada, "An introduction to event-triggered and self-triggered control," In *Proceedings of IEEE conference on decision and control,* pp. 3270–3285, 2012.

[63] A. Cetinkaya, H. Ishii, and T. Hayakawa, "Networked control under random and malicious packet losses," *IEEE Transactions on Automatic Control,* vol. 62, issue 5, pp. 2434–2449, 2017.

[64] H. J. LeBlanc, H. Zhang, X. Koutsoukos, and S. Sundaram, "Resilient asymptotic

consensus in robust networks," *IEEE Journal on Selected Areas in Communications,* vol. 31 , pp. 766–781, 2013.

[65] S. M. Dibaji, H. Ishii, and R. Tempo, "Resilient randomized quantized consensus," *IEEE Transactions on Automatic Control,* vol. 63, issue 8, pp. 2508–2522, 2018.

[66] A. Ahmed, K. A. Bakar, M. I. Channa, K. Haseeb, and A. W. Khan, "A survey on trust based detection and isolation of malicious nodes in ad-hoc and sensor networks," *Frontiers of Computer Science*, vol. 9, issue 2, pp. 280–296, 2015.

[67] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cy- ber-physical systems under adversarial attacks," *IEEE Transactions on Automatic Control,* vol. 59, issue 6, pp. 1454–1467, 2014.

[68] X. Jin, W. M. Haddad, T. Yucelen, "An Adaptive Control Architecture for Mitigating Sensor and Actuator Attacks in Cyber-Physical Systems," *IEEE Transactions on Automatic Control*, vol. 62, pp. 6058–6064, 2017.

[69] S. Nateghi, Y. Shtessel, "Robust stabilization of linear differential inclusion using adaptive sliding mode control," *Annual American Control Conference*, pp. 5327-5331, Milwaukee, July 2018.

[70] R. J. Rajesh, Y. Shtessel, "Accuracy improvement of dynamic sensors using higher order sliding mode observers," *Annual American Control Conference*, pp. 5731-5736, Philadelphia, July 2018.

[71] P. Razzaghi, E. A. Khatib, Y. Hurmuzlu, "Nonlinear dynamics and control of an inertially actuated jumper robot," *Nonlinear Dynamics*, vol. 97, issue 1, pp. 161-176, 2019.

[72] M. Navabi, H. Mirzaei, "Robust optimal adaptive trajectory tracking control of quadrotor helicopter," *Latin American Journal of Solids and Structures*, vol. 14, issue 6, pp. 1040-1063, 2017.

[73] L. Fridman, A. Levant, J. Davila, "Observation of linear systems with unknown inputs via high-order sliding-modes," *International Journal of systems science,* vol. 38, issue 10. pp. 773-791, 2017.

[74] 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,* vol. 11, issue 11, pp. 1756–1766, 2017.

[75] C. Wu , Z. Hu, J. Liu, and L. Wu, "Secure Estimation for Cyber-Physical Systems via Sliding Mode," *IEEE Transactions on Cybernetics*, vol. 48, no. 12, pp. 3420 - 3430, 2018.

[76] X. Huang, D. Zhai, J. Dong, "Adaptive integral sliding-mode control strategy of data-driven cyber-physical systems against a class of actuator attacks," *IET Control*

Theory & Applications, vol. 12, issue 10, pp. 1440-1447, 2018.

[77] 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*, vol. 9, issue 2, pp. 886-899, 2018.

[78] S. Mousavian, J. Valenzuela, and J. Wang, "A probabilistic risk mitigation model for cyber-attacks to PMU networks, *IEEE Transactions on Power Systems*, vol. 30, issue 1, pp. 156–165, 2015.

[79] S. Nateghi, Y. Shtessel, J-p Barbot, G. Zheng, L. Yu, "Cyber-Attack Reconstruction via Sliding Mode differentiation and Sparse Recovery algorithm: Electrical Power Networks Application. In: *15th International Workshop on Variable Structure Systems and Sliding Mode Control*, pp. 285-290, Graz, Austria, July 2018.

[80] S. Nateghi, Y. Shtessel, J. P. Barbot, C. Edwards, "Cyber Attack Reconstruction of Nonlinear Systems via Higher-Order Sliding-Mode Observer and Sparse Recovery Algorithm," *IEEE Conference on Decision and Control*, pp. 5963-5968, Miami beach, USA, 2018.

[81] S. Nateghi, Y. Shtessel, and C. Edwards, "Cyber-Attacks and Faults Reconstruction using Finite Time Convergent Observation Algorithms: Electric Power Network Application," Journal of the Franklin Institute, 2019.

[82] S. Nateghi, Y. Shtessel, C. Edwards, JP Barbot, "Secure State Estimation and Attack Reconstruction in Cyber-Physical Systems: Sliding Mode Observer Approach," Control Theory in Engineering, 2019.

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

[84] 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.

[85] L. Silverman, "Inversion of multivariable linear system," *IEEE Transaction on Automatic Control*, vol. 14, pp. 270-276, 1969.

[86] M.K. Sain and J. Massey. "Invertibility of linear time- invariant dynamical systems of multivariable linear systems," *IEEE Transaction on Automatic Control*, vol. 14, pp. 141-149, 1969.

[87] J-P Barbot, D. Boutat and T. Floquet, "An observation algorithm for non-linear systems with unknown inputs," *Automatica*, vol. 45, issue 8, pp. 1970-1974, 2009.

[88] L. Yu, G. Zheng, and J-P. Barbot, "Dynamic Sparse Recovery with Finite-time Convergence," *IEEE Transaction on Signal Processing*," vol. 65, no. 23, pp. 6147-6157, 2017.

[89] T. Floquet, C. Edwards, S. K. Spurgeon, “On sliding mode observers for systems with unknown inputs,” *International Journal of Adaptive Control and Signal Processing*, vol. 21, pp. 638–656, 2007.

[90] A. Levant, “Sliding order and sliding accuracy in sliding mode control,” *International Journal of Control,* vol. 58, issue 6, pp. 1247–1263, 1993.

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

[92] C. Edwards, SK. Spurgeon, “On the development of discontinuous observers,” *International Journal of control*, vol. 59, issue 5, pp. 1211-1229, 1994.

[93] A. Isidori, *Nonlinear Control Systems* (3rd edition). Springer: Berlin, pp. 219–290, 1995.

[94] H. K. Khalil, *Nonlinear systems*, Prentice-Hall, New Jersey, 1996.

[95] L. Fridman, Y. Shtessel, C. Edwards, and X. G. Yan, “Higher Order Sliding Mode Observer for State Estimation and Input Reconstruction in Nonlinear Systems,” *International Journal of Robust and Nonlinear Control,* vol. 18, issue 4-5, pp. 399-412, Mar. 2008.

[96] A. Levant, “High-order sliding modes: differentiation and output-feedback control,” *International Journal of Control*, vol. 76, issue 9–10, pp. 924–941, 2003.

[97] E. J. Davision and S. H. Wang, “On pole assignment in linear multivariable systems using output feedback,” *IEEE Transaction on Automatic Control*, vol. 20, issue 4, pp. 516-518, 1975.

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

[99] C. Edwards, Y. B. Shtessel, “Adaptive continuous higher order sliding mode control,” *Automatica*, vol. 65, pp. 183–190, 2016.

[100] C. Edwards and Y. Shtessel, “Adaptive Dual Layer Super-Twisting Control and Observation,” *International Journal of Control*, vol. 89, issue 9, pp. 1759-1766, 2016.