Cyber-attack Detection and Resilient State Estimation in Power Systems

by

Ana Jevtić

M.Sc., University of Belgrade, 2013

Submitted to the Department of Electrical Engineering and Computer Science in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology

May 2020

Author:

Department of Electrical Engineering and Computer Science
May 15, 2020

Certified by:

Marija Ilić
Senior Research Scientist, Lincoln Lab and IDSS
Thesis Supervisor

Accepted by:

Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Cyber-attack Detection and Resilient State Estimation in Power Systems

by

Ana Jevtić

Submitted to the Department of Electrical Engineering and Computer Science on May 15, 2020 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Abstract

Many critical infrastructures, such as transportation and electric energy networks, and health care, are now becoming highly integrated with information and communication technology, in order to be more efficient and reliable. These cyber-physical systems (CPS) now face an increasing threat of cyber-attacks. Intelligent attackers can leverage their knowledge of the system, disruption, and disclosure resources to critically damage the system while remaining undiscovered. In this dissertation, we develop a defense strategy, with the ability to uncover malicious and intelligent attacks and enable resilient operation of cyber-physical systems. Specifically, we apply this defense strategy to power systems, described by linear frequency dynamics around the nominal operating point. Our methodology is based on the notion of data aggregation as a tool for extracting internal information about the system that may be unknown to the attacker. As the first step to resilience and security, we propose several methods for active attack detection in cyber-physical systems. In one approach we design a clustering-based moving-target active detection algorithm and evaluate it against stealthy attacks on the 5-bus and 24-bus power grids. Next, we consider an approach based on Interaction Variables (IntVar), as another intuitive way to extract internal information in power grids. We evaluate the effectiveness of this approach on Automatic Generation Control (AGC), a vital control mechanism in today's power grid. After an attack has been detected, mitigation procedures must be put in place to allow continued reliable operation or graceful degradation of the power grid. To that end, we develop a resilient state estimation algorithm, that provides the system operator with situational awareness in the presence of wide-spread coordinated cyber-attacks when many system measurements may become unavailable.

Thesis Supervisor: Marija Ilić
Title: Senior Research Scientist, Lincoln Lab and IDSS

To Mladen,
who taught me what resilience means

Acknowledgements

This thesis would not have been possible without the support of so many people, to whom I would like to express my deepest gratitude and appreciation. First and foremost, I am thankful to my advisor, Prof. Marija Ilić, for sharing with me her passion and enthusiasm for research, and her continuous support and guidance. Marija's endless patience and dedication, as well as her wealth of knowledge, were beacons of light directing me towards deep revelations and allowing me to reach my true potential. I am grateful for the freedom and encouragement she has given me to explore my own ideas, as well as guidance to help me stay on my path. It has been a privilege to work with Marija on this thesis.

I also extend my sincerest gratitude to my thesis committee members: to Prof. Luca Daniel, for his eagerness as an instructor and a mentor, and his valuable insight that helped shape this thesis; to Prof. Jacob White, for his continuous feedback, support and guidance, starting from my RQE all the way to my thesis defense.

I also wish to thank my friends and colleagues Rupamathi Jaddivada and Xia Miao, for many fun discussions and much needed relaxing outings. Thanks to the late night chats with Rupa, that I would choose over sleep anytime, and fun days at the office with Xia, my time at MIT was so much more enjoyable. Likewise, my thanks goes to all the fellow students, colleagues and friends at LIDS, for all the joy, support and words of comfort throughout this journey. Many thanks to all the staff at LIDS and at the EECS Graduate Office, especially to Rachel Cohen, who has always been there to help me navigate the administrative side of things, and Prof. Leslie Kolodziejski, who I could always depend on for advice and kindness. I could not forget to thank all of my teammates and coach Tony Lee from the Women's Volleyball Club, who have truly changed my life at MIT for the better, and who I will miss dearly.

Finally, my deepest debt of gratitude is to my family. To my amazing parents, Ljiljana and Milan, my brother and best friend Mladen, and my wonderful fiancé Stefanos, thank you for always supporting and encouraging me to pursue my dreams. There are no words to express the love I have for you and the gratitude I owe you. You were an endless source of inspiration, support and motivation to push onward, and I dedicate this dissertation to you.

Of course, this research would not be possible without the financial support from the Cybersecurity Center for Secure Evolvable Energy Delivery Systems (SEEDS) and Department of Energy (DoE) award DE-OE0000779.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Thesis Motivation

Cyber-physical systems (CPS) use advanced computation, sensing, communication and control technologies to efficiently operate a physical process. CPS is the underlying structure of many critical infrastructures, such as the electric power grid, transportation systems, gas and water networks. Resilient and reliable operation of these critical infrastructures is an important task, as new challenges arise [1–4].

CPS leverage advanced computation capability to efficiently operate the physical process so that certain objectives are reached. For example, in electric power grids, the system frequency needs to be maintained at nominal frequency (60 Hz in the U.S., 50 Hz in Europe and some other parts of the world) at all times, while keeping the operating cost as low as possible. Alternatively, safe navigation of a self-driving car or an Unmanned Aerial Vehicle (UAV) is an important objective in transportation systems.

To achieve these objectives, relevant data needs to be gathered from the physical process. Combined with the model of the physical system, acquired data allows for the current state of the system to be inferred. As the sensing technology is advancing, and the cost of implementation decreasing, the number of sensors deployed to monitor physical processes is on the rise.

Using the relevant data gathered from the physical system, the physical system can be controlled in order to reach desired performance objectives. Today, microcontrollers and programmable logic controllers (PLCs) are widely used in industry to implement advanced control logic and algorithms at the component or subsystem level. These devices are able to quickly process incoming data and compute control signals in an automated fashion.

Finally, a sophisticated communication infrastructure is used by CPS to transfer the sensor and control signals. Commonly, a Supervisory Control and Data Acquisition (SCADA) is used to manage the system in a hierarchical manner. It allows the system operator to remotely operate the system in the higher supervisory layer. In the lower layer, local devices, such as PLCs and Remote Terminal Units (RTUs) receive the commands from SCADA and directly interface with physical actuators (e.g. valves, motors, switches, etc.).

Unfortunately, there have been a number of successful attacks on cyber-physical systems [5–8] in the recent years. The motivation for these attacks can range from economic reasons [9] to purely malicious, terrorist ones. One well-known malicious attack was launched in the Ukrainian power grid in 2015 [7], where three distribution utilities were affected, and more than 200000 customers were without electricity for an extended period of time. Investigation revealed that this attack was launched via phishing emails delivering the BlackEnergy3 malware, which was eventually activated by an employee. The attack targeted the SCADA network via field devices with malicious firmware, allowing an attacker to remotely open the substation breakers. First stage of the attack was harvesting credentials to access the industrial control system (ICS) network. Then a telephone system was used to generate denial-of-service (DoS) attack, denying access to customers attempting to report the outages. The restoration efforts and mitigation procedures were further delayed by erasing master boot records on workstations via a modified KillDisk firmware attack. Another, more recent attack was the cyber incident that disrupted grid operation in the western US in March of 2019, via a DoS attack [8].

These events have led to the increase in awareness of the problem of securing critical infrastructure. Unfortunately, the opportunities for the attackers to manipulate these systems are still abundant. The control systems behind these critical infrastructures have long been protected by physically isolating the local control and communication networks from insecure global networks such as the Internet. However, parts of CPS are becoming increasingly exposed to the public via smart devices with Internet and other wireless connectivity. Some of examples of this are devices and applications that are already wide-spread and present in many homes and businesses: vehicle control through phone apps, devices such as smart meters, Nest, and Google

Home, that allow remote control of large appliances connected to the power grid.

An attacker can also take advantage of the many weaknesses of the communication network that already provides access to both sensing and control. One that is common in large scale systems with many heterogeneous components is outdated and improperly setup firewalls and malware protection. Another weakness is caused by attackers first targeting and hijacking trusted VPN (Virtual Private Network) or stealing valid employee credentials to gain access to SCADA communication. Another vulnerability is in the connected devices themselves. For example, in power grids, due to the sheer scale of the system, it is practically impossible to physically protect each component. Consequently, many substations, as well as (smart) meters and other sensors (PMUs) are left unattended and unguarded. An infected USB (as in Stuxnet attack [5]) can introduce malware to a field component, or be used to install a backdoor, which can later be used to manipulate the CPS. Similarly, malware in form of email attachments can be used to gain access to the corporate network, as in the Ukraine attack [7]. Finally, a disgruntled employee or malicious insider may lend or provide access rights to another entity, or even conduct the attack from the inside.

Given the abundant ”means, motive and opportunity”, it is imperative to design CPS in a way that incorporates fundamental security principles of *confidentiality*, *integrity* and *availability*, also referred to as the CIA triad [10]. All aspects of the CIA triad must be satisfied in order for the system security to be considered comprehensive and complete. *Confidentiality* means that data, resources and other sensitive information are protected from unauthorized access. Eavesdropping is an example of cyber-attacks that only target confidentiality, and is commonly a part of more complex attacks that can be designed if the other resources are utilized. *Integrity* refers to protecting the data from unauthorized changes, so that it is reliable and correct. For example, False Data Injection (FDI) attacks, including replay attacks utilized in Stuxnet, can be used to manipulate data integrity in order to cause disruption in the system. *Availability* means that authorized users have access to the systems and the resources they need. Denial-of-service, communication jamming and spoofing are attack examples that target data availability. Although data availability attacks are straightforward to detect, they may cause disruption in system operation, as the system cannot be controlled and operated reliably when feedback data and measurements are unavailable. Since cyber-attacks that target availability of data are easily recognizable, the attack detection methods proposed in this thesis will focus mainly on integrity attacks, namely False Data Injections (FDI). On the other hand, both data availability and integrity attacks will be considered in our analysis of resilient system operation.

FDI attacks affect the system by modifying control or measurement signals. The attacker can gain access to these signals in one of two ways. They can infiltrate the communication network, and then intercept real data packets while forwarding their own modified packets to the operator or the process. Alternatively, they can directly affect the sensors by manipulating their environment. Manipulated sensor signals will in turn cause wrong control signal computation, which can cause damage in the physical system. Same damaging effects can be caused by manipulating the control signal directly. Manipulating the control signal directly requires access to the communication between the local controller and the actuator. In the example of the power grid, this communication usually takes place within a secure facility, as generators are the main controllable components. It is also usually encrypted, and therefore regarded as less vulnerable when compared to the communication between the sensor and the operator. If integrity attacks are also stealthy, they may not be uncovered for an extended period of time, allowing the attacker great freedom to manipulate the system. This kind of attack was utilized both in Stuxnet (as a replay attack) and in the Ukraine incident. Thus, we focus on stealthy FDI attacks on sensor signals in the remainder of this thesis.

The danger multiplies when the attackers are resourceful, have detailed knowledge of the system, and can therefore launch highly effective and deceptive attacks. It has been well-established that passive detection theory is ineffective against stealthy attacks that closely mimic normal system behavior. Replay attacks [11] and zero-dynamics [12] attacks are examples of how an adversary can exploit knowledge of the system to launch attacks that can evade detection. Under these circumstances, the defender can assume an active role in cyber-security of the system, and exploit additional degrees of freedom that are unavailable to the attacker. The defender can design the control and detection strategy and utilize sensors in a way that altogether prevents the attacker from devising a stealthy attack. Thus, in this thesis, we propose intelligent system designs that prohibit the attacker from designing stealthy attack sequences.

1.2 Thesis Contributions

In this thesis, we argue that the defender must exploit additional degrees of freedom in system design in order to defend against stealthy cyber-attacks. For this purpose, we propose an aggregation framework that allows for the intelligent design of detection strategies that behave like a moving target. Utilizing the knowledge of the aggregate behavior of different parts of the system, and the ability to manipulate how aggregation is performed, the defender can actively change the detection strategy over time, continuously challenging the attacker. Further, knowledge of the aggregate system behavior can extend situational awareness of the operator during a wide-spread attack, when potentially large number of sensors may be affected. The aggregate system variables can then be used to replace unavailable or untrusted sensor readings. Using the aggregation framework, this thesis focuses on several challenges in power system operation posed by the actions of malicious intruders.

Attack detection

Attack detection is an essential first step to securing the electric power grid. Present methods rely on a fixed defense strategy, which, once discovered by the attacker, will no longer be effective. Hence, active and dynamic detection methods are needed to combat highly intelligent and invasive attacks.

Resilient SE

After an attack is detected, mitigation and self-healing procedures need to be initiated. However, until the intruder has been physically removed from the system, the system operator will have limited knowledge of the system's state and conditions, as many sensors may be unavailable. In that situation, critical processes that provide situational awareness, such as state estimation, need to be enhanced to provide resilience.

1.3 Thesis outline

This thesis will address each of the above issues, from theoretical perspective and through numerical simulations. We further introduce the problem of cyber-security in electric power systems in Chapter 2, where we describe the today's hierarchical control and its vulnerabilities, or attack surfaces for malicious intruders. In Chapter 3, we introduce the notion of extracting internal structure of the system through aggregation, to be leveraged for defensive purposes. Two aggregation methods are evaluated for the purpose of cyber-physical security. In Chapters 4 and 5, we propose two cyber-attack detection methods, based on aggregate behavior of the system. We address situational awareness and propose an algorithm that ensures resilient system operation in Chapter 6.

Chapter 2

The Vulnerability of Today's Hierarchical Power System Control

Current industry practice in power system stabilization and regulation is based on hierarchical frequency control. Figure 2.1 illustrates the control layers in today's power system operation with regard to different timescales. Nominal operating frequency, 60 Hz in the US, is maintained when supply (power generation) and demand (power consumption) are perfectly balanced. Thus, the system frequency is required to be at nominal value in real time, to ensure the supply and demand are balanced, and ensure system stability as well as quality of service.

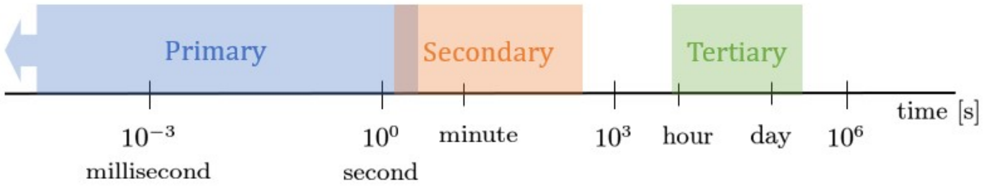


Figure 2.1: Timescales in different layer's of hierarchical control in power systems

During normal system operation, when all equipment is working as expected, the changes in the system are driven entirely by load dynamics.

Typical load curves exhibit variations at several different rates, ranging from very fast random variations (order of seconds), to hourly, daily, weekly and seasonal patterns. Based on historical load data, these hourly, daily, weekly, seasonal trends of the demand can be predicted ahead of time [13]. Figure 2.2 depicts a typical daily load curve[1], where the actual load is shown in blue, and load forecast in red. In the tertiary control level, generation is scheduled a day ahead of time to satisfy the forecast, predictable part of the total load demand. The challenge in frequency regulation arises from the unpredictable variations in demand, that occur at the timescale of minutes and seconds. Secondary and primary control layers are tasked with satisfying this unpredictable part of the demand in real time, represented by the deviation of demand from the forecast value, in order to maintain nominal frequency at all times.

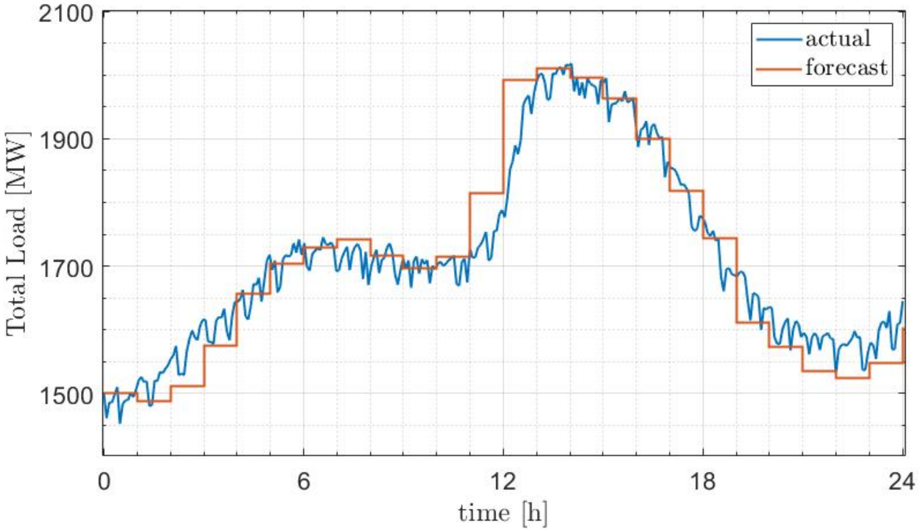


Figure 2.2: Forecast vs. actual load over 24h period [14]

Primary control in power systems is entirely decentralized, and its main purpose is to stabilize generator frequency and voltage dynamics. Each generator has a speed-governor controller, which stabilizes local frequency dynamics, and a field excitation controller, also known as Automatic Voltage Regulator (AVR), which stabilizes voltage dynamics. Since these controllers

[1] data taken from NY-ISO [14]

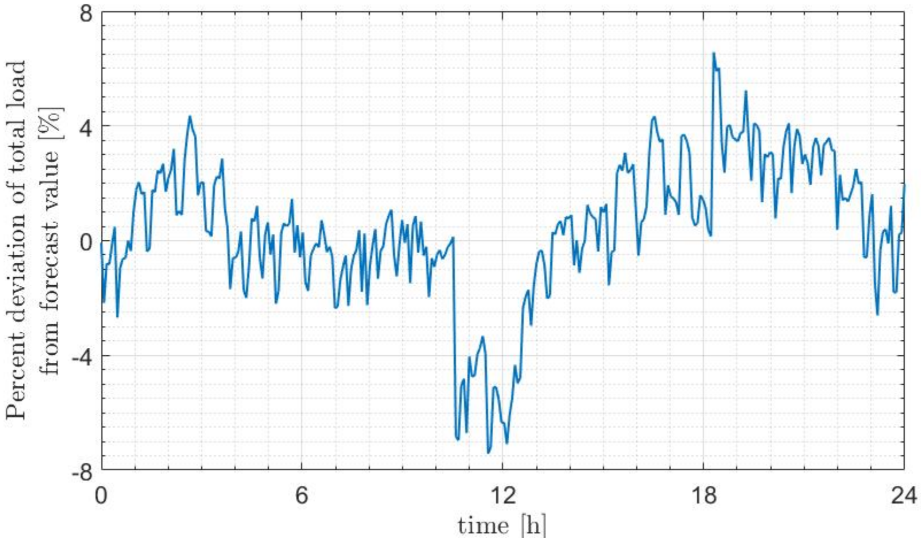


Figure 2.3: Deviation of load from forecast value [14]

are local and decentralized, they are tuned using an oversimplified model, a Thevenin equivalent [15,16], to represent the rest of the system. This tuning method ignores the dynamic interactions between different system components. In today's primary power system control, frequency is controlled via proportional output control, which may result in a static tracking error [17]. The output of interest is the deviation of local frequency from the nominal value of 60 Hz. However, there is no guarantee that the interconnected system will be stable when all the generators' controllers are tuned in this manner. Similarly, each AVR used for local voltage regulation is tuned under the assumption that the rest of the system is static. Thus, as in frequency regulation, there is no guarantee that the interconnected system voltage dynamics will be stable. It is important to note that there is a big disparity between literature, where many methods that provide stability guarantees have been proposed [18, 19], and industry practice. In the current industry implementation, this fast stabilization by primary-level controllers is assumed to be fully effective, even with the drawbacks of the current controller tuning procedures.

According to current industry standards, such as A1, A2, Control Performance Standard (CPS)-1, and CPS-2 [20, 21], frequency is required to be regulated at a slower rate and at the control area (CA) level. A control area usually encompasses a part of the grid owned by a single utility. Once the local generator frequency is stabilized by the primary controllers, Automatic Generation Control (AGC) is implemented at the control area level. AGC is intended for regulating the area frequency to the nominal 60 Hz. As such, it effectively compensates the slower deviations of supply from the required demand [22–24].

AGC performs frequency regulation by correcting the error in tracking the nominal frequency via proportional-integral (PI) control, after the frequency has been stabilized by primary control. The power imbalance in the area, caused by change in load or deviation of tie-line flows from scheduled values manifests itself through deviation in local frequency. Local measurements are used to compute Area Control Error (ACE), which is used to quantify this power imbalance. In a multi-area power system, each area is equipped with it's own AGC, responsible for maintaining the scheduled exchange with other areas, as well as supplying its own load. Thus, AGC can be viewed as centralized control at the area level, and decentralized control from the interconnected system level, as there is no coordination in control between areas.

In this thesis, we focus on primary and secondary control layers, that work together to satisfy the unpredictable changes in demand. In this setting, the predictable portion of the load is assumed to be supplied through economic dispatch when the day-ahead market is cleared (tertiary control). Then the unpredictable variations in load can be viewed as relatively small deviations around the forecast load, as seen in Figure 2.3 (for the daily demand shown in Figure 2.2). Secondary control, AGC, then computes set-points for the generators' primary controllers that will cancel these deviations. The overall power system hierarchical control scheme is depicted in Figure 2.4 This setting lends itself to using a linearized model of the system.

Large amounts of heterogeneous data are also becoming increasingly available from sources such as smart meters, distributed generation, transmission sensors, and smart home energy management systems (Nest, Google Home, etc.) in addition to the commonly used Supervisory Control and Data Acquisition (SCADA) measurements. This allows the critical system states to be monitored dynamically, with more precision and accuracy than ever before. The Phasor Measurement Unit (PMU) and Wide Area Measurement Systems (WAMS) are examples that have attracted attention from both academic and industry communities. Although there are a lot of sources of data

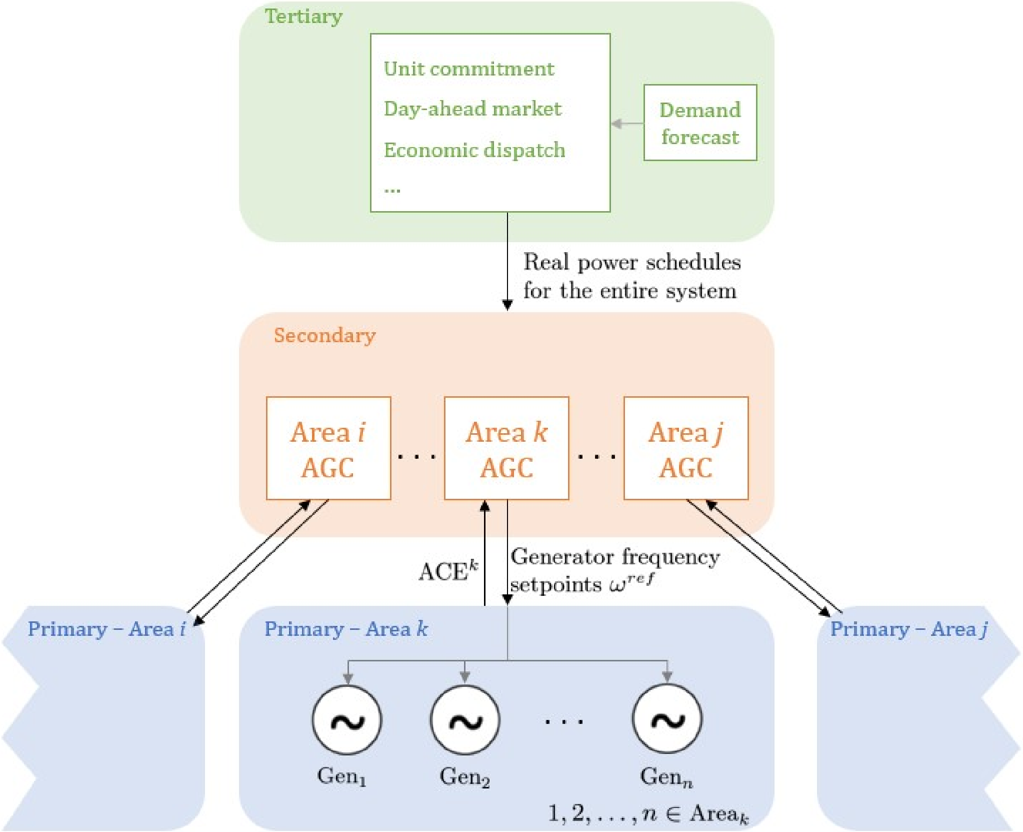


Figure 2.4: Illustration of hierarchical control in today's power systems

in distribution systems, and sophisticated sensors in the field, these data are frequently not used as they are not trusted. Instead, the industry relies on simple controllers that don't require a lot of data, which is a big disparity with the academic state-of-the-art. In this thesis, we consider the existing control architecture, but utilize the additional available data to ensure cyber-security of the grid.

While uncommon in power system stability analysis, modeling load dynamics is critical in cyber-security applications. That is especially the case for stealthy FDI attacks, when the attacker is attempting to inject malicious signals that resemble those of normal system operation. A false signal indicating that the system load is changing would prompt the generation to change accordingly. This wrong control signal can, in turn, cause serious consequences in the power grid, such as component disconnection or even cascading blackouts. Modeling that accurately captures load behavior is key to discerning true system response to load changes from the attacker's input. This has now become possible due to novel advanced sensing technologies, and their large scale deployment, enabling identification of such models. Thus, a structure-preserving model is derived to capture the behavior of loads and the system's response to that behavior.

The rest of this chapter is outlined as follows. In Section 2.1 we outline some of the vulnerabilities of today's power systems. Section 2.2 reviews dynamical models of relevant power system components. In particular, the model of the Generator-Turbine-Governor (G-T-G) set, commonly used in frequency stability analysis, the structure-preserving load model, as well as today's AGC control system model. Section 2.2.6 summarizes the entire chapter.

2.1 Vulnerability of today's power systems

Industrial control systems (ICS) are used to control and monitor the various parts of the grid, and may or may not be connected to the Internet. ICS that are not connected to the Internet, still rely on local area networks (LANs) or similar systems in order to control and monitor the process. Additionally, a large amount of electronic equipment, switches and circuit breakers are used to regulate different parts of the grid.

One example of ICS in power systems is SCADA, used to control geographically dispersed components, often spread out over large distances. Since its implementation in the 1970s, SCADA operation has been upgraded by connecting many of the older, legacy systems to the Internet, to improve the overall system efficiency and make it more intelligent. However, many of these legacy systems were not designed with security in mind, introducing new potential access points for a cyber-attack. One example is intrusion through data reporting routes, or malware injections via a thumb drive, like the Stuxnet worm [5]. More modern, "Smart Grid" components are designed using microprocessor and other hardware devices with advanced computing and networking capabilities. As such, they may be susceptible to manipulation over a network or the Internet [25].

ICS, including SCADA systems, have been designed to be efficient, rather than secure. The implemented control mechanisms have critical timing requirements, rigid performance specifications, and specific task priorities. They also have limited computing resources and communication bandwidth. This constrains use of existing IT cyber-security protocols, such as encryption and certificate authentication.

Each control area is usually managed by a single utility, with a SCADA at its control center (CC), performing centralized monitoring and control over long distance communication lines. Based on measurements received from remote terminal units (RTU), supervisory controls can be sent to field devices. Field devices perform actions such as opening and closing breakers, collecting measurements from sensors, and monitoring their behavior for alarm conditions [26]. An RTU is usually located at a substation or power plant, collecting data from the field devices. These field devices usually communicate with the RTU through encrypted, short-distance communication network. Substation RTUs collect measurement from entire neighborhoods or industrial complexes, and are usually somewhat physically secure, but impossible to actively monitor due to their number. The data collected by the RTU is then forwarded to the CC through long-distance, usually unencrypted, communication network. In this thesis, we focus on the unencrypted long-distance communication between RTUs and CC, and model loads at the substation level. The model of the aggregate substation level load is presented in the following section.

On the other hand, generator's primary controllers, governor and field excitation control, are local to the secure facility housing the generator itself. These facilities are not only physically secure, but actively monitored and guarded. Thus, we consider the local sensing and monitoring needed for primary control to be more secure, and leave the study of cyber-security of these systems to be addressed in future work.

2.2 Power system modeling

In this section we introduce the component and interconnected system models needed for analysis and cyber-security tool design. Let n_G be the number of generators and n_L number of loads in the system, and denote the set of generator buses by $\mathcal{G}$, and the set of load buses by $\mathcal{L}$.

2.2.1 Linearized model of the cyber-physical system

The dynamics of the power system are modeled using nonlinear differential equations:

$$\dot{x}(t) = f(x),\ x(t_0) = x_0 \tag{2.1}$$

where x is a vector of states, $f(\cdot)$ is a vector of nonlinear differential equations and x_0 is the state at time t_0, also called the initial condition. To perform the analysis needed for cyber-security tool design, this model is linearized around a given operating point [22,27]. Recall the Taylor series expansion of the function $f(\cdot)$ about x_0:

$$f(x) \approx f(x_0) + \frac{\partial f(x_0)}{\partial x}(x - x_0) + H.O.T. \tag{2.2}$$

where $x = \Delta x + x_0$. Then, Δx is the small deviation around the equilibrium x_0, and state of the linearized system:

$$\Delta\dot{x} = \frac{\partial f(x_0)}{\partial x}\Delta x,\ \Delta x(t_0) = x_0 \tag{2.3}$$

The fraction $\frac{\partial f(x_0)}{\partial x}$ is commonly referred to as the Jacobian. Finally, the standard state space model of the linearized system is given by:

$$\Delta\dot{x} = A\Delta x,\ \Delta x(t_0) = x_0 \tag{2.4}$$

where $A = \frac{\partial f(x_0)}{\partial x}$ is the system matrix. This commonly used approach to analysis is usually carried out by viewing the system as a whole. As such, the matrix A represents the closed loop system matrix, assuming the control has already been designed and tuned properly. Similarly, in this thesis, we consider cyber-physical security problems of systems for which appropriate control systems for stabilization and regulation already exist. Thus, the system is assumed to be stable (semistable) in absence of cyber-attacks. In other words, the matrix A is assumed to be negative semi-definite. The structure of the system matrix A will be discussed in the following sections.

2.2.2 Generator model with primary control

In this section we review the widely used Governor-Turbine-Generator (G-T-G) model. Dynamics of the mechanical subsystem of a generator are described by the swing equation:

$$J\dot{\omega}_G + D\omega_G = P_T - P_G + e_T a \tag{2.5}$$

where ω_G is the generator bus frequency, P_G the net real power injected into the network. Parameters J, D are generator's inertia and damping, and e_T is a parameter of the turbine.

Given the frequency set-point ω^{ref} from the secondary control layer, the speed-governor controls the generator frequency. Its states P_T and a denote the mechanical power of the generator and the turbine valve position, respectively. The governor dynamics are given by:

$$T_u \dot{P}_T = -P_T + K_t a \tag{2.6}$$

$$T_g \dot{a} = -ra - (\omega_G - \omega^{ref}) \tag{2.7}$$

The governor's and turbine's time constants are denoted by T_u and T_g while K_t and r are control gains.

2.2.3 Structure-preserving load model

In order to capture the behavior of interest, we adopt the structure-preserving load model [28, 29], briefly reviewed below. This thesis is focused on the transmission level system, considering aggregate loads at the substation level (everything below the substation is considered as one aggregate load). Another benefit of modeling loads as dynamic is that the sparsity of the overall system is preserved, as it is no longer needed to remove algebraic equations via Kron reduction, or a similar procedure. As seen in equations (2.8), the load dynamics in this model are driven by the mismatch between the electrical power delivered through the network (P_L) and the actual power (L) consumed by the load:

$$J\dot{\omega}_L + D\omega_L = P_L - L \tag{2.8}$$

where J and D are equivalent moment of inertia and damping coefficient of the aggregate load.

2.2.4 Interconnected system model

In order to derive the interconnected system, we treat the net real power injection P_i at each bus i as a coupling state variable. The vector of net real power injections at each bus P is composed of generator and load net real power injections:

$$P = \begin{bmatrix} P_G \\ -P_L \end{bmatrix} \tag{2.9}$$

where $P_G := [P_i]_{i \in \mathcal{G}}$ and $P_L := [P_i]_{i \in \mathcal{L}}$. For a lossless system, the dynamics of P can be obtained from the power flow equation:

$$P_i = \sum_{j \in \mathcal{N}} |Y_{ij}||V_i||V_j| \sin(\theta_i - \theta_j) \tag{2.10}$$

where $\mathcal{N}$ is a set of all buses connected to bus i through a transmission line of admittance Y_{ij}, and V_i is the voltage of bus i. To obtain the linearized DC power flow, certain assumptions are made:

- Transmission line resistance R is assumed to be negligible $(R \ll X)$
- Voltage angle differences are assumed to be small, i.e. $\sin\theta = \theta$, $\cos\theta = 1$
- Voltage profile assumed flat, i.e. bus voltage magnitudes are set to 1 p.u.

Then, the DC power flow equations can be expressed in matrix form as:

$$P = Y_{bus}\theta \tag{2.11}$$

The matrix Y_{bus}, also called admittance matrix, is defined for a transmission network topology characterized by its incidence matrix M as:

$$Y_{bus} = MBM^T \tag{2.12}$$

where $B = \text{diag}\{B_{ij}\}$, and B_{ij} is the susceptanse of the transmission line connecting buses i and j. Matrix Y_{bus} can be partitioned as:

$$Y_{bus} = \begin{bmatrix} Y_{GG} & Y_{GL} \\ Y_{LG} & Y_{LL} \end{bmatrix}$$

Differentiating the equation (2.11) we obtain:

$$\begin{bmatrix} \dot{P}_G \\ -\dot{P}_L \end{bmatrix} = Y_{bus} \begin{bmatrix} \omega_G \\ \omega_L \end{bmatrix} \tag{2.13}$$

Finally, the linearized closed-loop model of power system with primary control can be described in standard state space form by combining equations (2.5), (2.7), (2.8) and (2.13), with the state vector $x := [\omega_G, \omega_L, P_G, P_L, P_T, a]^T$, as:

$$\dot{x} = \mathcal{A}x + Bu + Gd \tag{2.14}$$

In this model, the control input from the secondary control layer is denoted with $u = [\omega_i^{ref}]_{i \in \mathcal{G}}$, and the disturbances with $d = [L_i]_{i \in \mathcal{L}}$. The system matrices are given by:

$$\mathcal{A} := \begin{bmatrix} -\mathbf{J}_G^{-1}\mathbf{D}_G & \mathbf{0} & -\mathbf{J}_G^{-1} & \mathbf{0} & \mathbf{J}_G^{-1} & \mathbf{J}_G^{-1}\mathbf{e}_T \\ \mathbf{0} & -\mathbf{J}_L^{-1}\mathbf{D}_L & \mathbf{0} & \mathbf{J}_L^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{Y}_{GG} & \mathbf{Y}_{GL} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{Y}_{LG} & -\mathbf{Y}_{LL} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{T}_u^{-1} & \mathbf{T}_u^{-1}\mathbf{K}_t \\ -\mathbf{T}_g^{-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{T}_g^{-1}\mathbf{r} \end{bmatrix}$$

$$B := \begin{bmatrix} \mathbf{0}_{(3n_G+2n_L)\times n_G} \\ \mathbf{T}_g^{-1} \end{bmatrix} \qquad G := \begin{bmatrix} \mathbf{0}_{n_G \times n_L} \\ -\mathbf{J}_L^{-1} \\ \mathbf{0}_{(3n_G+n_L)\times n_L} \end{bmatrix}$$

where:

$$\begin{aligned} &\mathbf{J}_G := \operatorname{diag}(\{J_i\}_{i\in\mathcal{G}}) && \mathbf{K}_t := \operatorname{diag}(\{K_{t,i}\}_{i\in\mathcal{G}}) \\ &\mathbf{J}_L := \operatorname{diag}(\{J_i\}_{i\in\mathcal{L}}) && \mathbf{T}_g := \operatorname{diag}(\{T_{g,i}\}_{i\in\mathcal{G}}) \\ &\mathbf{D}_G := \operatorname{diag}(\{D_i\}_{i\in\mathcal{G}}) && \mathbf{T}_u := \operatorname{diag}(\{T_{u,i}\}_{i\in\mathcal{G}}) \\ &\mathbf{D}_L := \operatorname{diag}(\{D_i\}_{i\in\mathcal{L}}) && \mathbf{r} := \operatorname{diag}(\{r_i\}_{i\in\mathcal{G}}) \\ &\mathbf{e}_T := \operatorname{diag}(\{e_{T,i}\}_{i\in\mathcal{G}}) && \end{aligned}$$

It should be noted that the system matrix $\mathcal{A}$ is quite sparse, which significantly reduces computation complexity, as it allows use of many fast computation algorithms, e.g. matrix multiplication, inversion, solving the Lyapunov and Sylvester equations, etc. Additionally, the model in (2.14) is structure-preserving and suitable for cyber-security tool design.

2.2.5 Secondary control layer - AGC

The control objective of each area's Automatic Generation Control (AGC) is to closely regulate the area frequency (to 60 Hz, in the U.S.), by maintaining scheduled power exchange with other areas, and supplying its own local loads. Today's AGC is designed with the assumption that the power system is in a quasi-static state, where the system's response remains close to the operating point. This assumption is used to derive the relation of the generator frequency ω_G to the generator's electrical power output P_G. This relationship is derived by setting the derivatives in equations (2.5) and (2.7) to zero, resulting in:

$$\omega_G = -\sigma P_G + (1 - \sigma D)\omega^{ref} \tag{2.15}$$

where σ is called the speed droop, and the equation (2.15) is referred to as the droop characteristic of the G-T-G set. Figure 2.5 depicts an example of the droop characteristic, and how a change in the generator power output P creates the need for updating the frequency reference ω^{ref}.

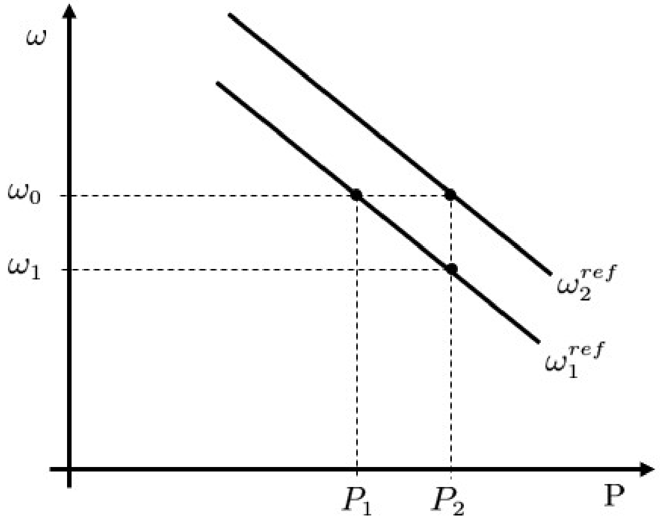


Figure 2.5: Illustration of the droop characteristic, and the effect of the increase in real power on frequency, and the resulting set-point change.

The speed droop σ is defined as:

$$\sigma = \frac{r}{rD + K_t + e_T} \tag{2.16}$$

Under this quasi-static assumption, the entire control area is reduced to an equivalent single machine, whose speed-droop is equal to the sum of speed droops of all generators in the area. Further, this results in frequencies of all buses within the area to be assumed uniform, i.e. $\omega^k = \omega^k_{G,1} = \cdots = \omega^k_{G,n_G}$, for an area k with n_G generators. The equivalent total droop of the area k is used to calculate the so-called area frequency bias:

$$\beta^k = \sum_{i \in Area_k} \frac{1}{\sigma_i} \tag{2.17}$$

This simplified representation of the area is used to derive the input to the AGC, Area Control Error (ACE), which represents the net power imbalance in the area:

$$ACE^k = (F - F^{sched}) + \frac{\beta^k}{2\pi}(\omega^k - 60Hz) \tag{2.18}$$

where F is the actual measured net tie-line power flow into the area, F^{sched} is the scheduled tie-line power exchange, and ω^k is the representative area frequency. Practically, frequency is measured at only one location, usually at the control center location, and used to compute the ACE signal. Therefore, ACE represents the net power imbalance caused by both external (inadvertent power exchange with other areas) and internal disturbances (local frequency deviation). Today's AGC is implemented as a simple PI controller, with an objective to drive ACE to zero, or, in other words, drive the power imbalance in the area to zero. The resulting control signal is the set-point adjustment for the governor controller of each generator participating in AGC.

2.2.6 Summary

In this chapter, we introduce today's hierarchical power systems control and the aspects of it that make it vulnerable to cyber-attacks. In particular, we focus our attention on secondary and primary control layers, where feedback control systems are implemented, as they are particularly vulnerable to cyber-attacks. Further, we review relevant dynamical models of power system components: the model of the Generator-Turbine-Governor (G-T-G) set, commonly used in frequency stability analysis, and the structure-preserving load model. We then derive the model of the interconnected physical system with primary control, and introduce today's secondary control scheme - the AGC control system. This model of the physical power system, as well as its primary and secondary control layers, enables further analysis and development of cyber-security methods developed in the following sections.

Chapter 3

Background on Aggregation Methods for Cyber-physical Security

With knowledge of the system topology, equipment and network parameters, control schemes and gains, a malicious intruder can design a stealthy attack that mimics normal system operation, but could cause devastating consequences. To combat an attacker with such extensive insight into the system, methods that take advantage of other, hidden, properties of the system need to be designed and implemented in order to ensure security of cyber-physical systems. Additionally, if this property is dynamic it allows the defender to construct a moving-target detection mechanism, altogether preventing an attacker from designing stealthy attacks.

Model order reduction methods used in power systems applications are predominantly based on decomposition. Decomposition refers to splitting the system into subsystems of lower dimensionality, and analyzing them independently. The separate solutions or conclusions made about the subsystems are then combined in some way to provide a solution or conclusion about the overall system. Decomposition methods used in power systems can be broadly classified into two categories: temporal and spatial decomposition. Most commonly used temporal decomposition method in power systems, namely Singular Perturbation Analysis (SPA) [32, 33], resulted from natural time scale separation. Power system dynamics encompass a wide variety of responses, ranging from very fast electrical phenomena evolving in microseconds to slower mechanical dynamics evolving over seconds or even minutes. Similarly, Modal Analysis (MA) [34] is used frequently to focus the analysis on modes of interest for the specific problem. For example, in transient stability analysis the faster modes are of interest, and slow ones can be assumed constant and disregarded from dynamical system. On the other hand, spatial decomposition usually entails dividing the power system into smaller geographical areas, and analyzing each subsystem independently, as if it was disconnected from the rest of the system. In general, the selection of the particular decomposition approach is highly dependent on the application.

For general linear systems, various model reduction methods have been thoroughly explored. Many of them can be classified as projection methods, and are equivalent to simple truncation in an appropriate basis of parts of the model treated as insignificant. Projection methods differ in the way that projection matrices are chosen, and some commonly used ones include Singular Value Decomposition (SVD) and Balanced Truncation (BT) [35,36], Moment Matching (MM) and Krylov methods [35, 37], etc.

There are several disadvantages to using decomposition and projection based model order reduction methods in cyber-security applications. First, in many instances, a priori error bounds are not known or cannot be computed. This is especially troubling for methods with high computational complexity, since the reduction procedure must be tuned by trial and error, possibly over many iterations. Second, and most evident in some temporal decomposition and projection methods, is that the physical meaning of subsystem state variables is lost. The projection matrices used in these methods are typically dense and not structured, resulting in loss of the sparse interconnection structure in power systems. In cyber-security applications, it is important to preserve sparsity through the reduction, so that the particular corrupted measurements in the original space can be related to those in the reduced space.

Conversely, methods for model order reduction through aggregation bring important benefits in cyber-security applications. Most importantly, the level of detail in modeling can be preserved through approximation. Even though the power systems are considered to be highly heterogeneous, they are also composed of many instances of the same component, such as generators, transformers, solar PVs etc. Therefore, a wide variety of dynamics can still be present in the reduced model when all the similar components are grouped and represented by an aggregate component. Secondly, sparsity of the system can be preserved through reduction. Sparsity of the reduced system is also a benefit from a computational viewpoint: we will take advantage of this property in cyber-attack detection, when the ability to perform fast analysis and computation will be of great importance to the feasibility of implementation.

Several aggregation methods have been developed for use in power system analysis. A well-known concept in power systems is coherency [38,39], which describes groups generators with similar dynamic response to a set of predetermined faults. Each coherent group of generators is then replaced by a large equivalent machine. This approach is commonly used for transient stability simulations, however, it is dependent on system conditions and time consuming to compute. Synchrony [40], or slow-coherency, is a similar concept, where generators are grouped with respect to a subset of modes, also called chords. In both coherency and syncrony-based approaches the less relevant parts of the system are replaced by static circuits. Further, both methods are designed specifically for generator equivalencing and do not consider other system components.

In this chapter, we review the technical background needed for development of security methods for cyber-physical systems proposed in the rest of this thesis. To that end, we propose an aggregation framework that can be used to extract this internal structure, and review two aggregation methods existing in the literature. In Section 3.1 we review the concept of Interaction Variables (IntVar), which represent aggregate behavior of control areas in power systems. In Section 3.2 we introduce a modified version of the clustering-based aggregation algorithm, which preserves sparsity of the system and is computationally efficient. In this approach, the outputs of the system are grouped based on the similarity of their responses to external disturbances. We also provide an interpretation and examples of the clustering-based aggregation in a 5-bus power system. Finally, Section 3.3 summarizes the chapter.

3.1 Control area aggregation via Interaction Variables

An important property of power system dynamics is its structural singularity, associated with the real power dynamics. Owing to this property, it can be shown that there exists a combination of states that stays constant when a control area is isolated and the reference signal is unchanged. This combination of states is termed Interaction Variable (IntVar), and is a fundamental concept behind inter-area dynamics in power systems.

Definition. *[22] Let the dynamics of control area i be described by a linear standard state space model:*

$$\dot{x}^i(t) = A^i x^i(t) + B^i u^i(t) + d^i(t) + F^i(t) \tag{3.1}$$

An interaction variable of control area i is defined as any linear combination of local states $z^i = T^i x^i$, $T^i \neq 0$, such that

$$\dot{z}^i(t) \equiv 0 \tag{3.2}$$

in the absence of control signal changes ($u^i(t) = 0$), disturbances in the area ($d^i(t) = 0$), and interactions with other areas ($F^i = 0$).

In this definition, the tie-line flows F between areas are considered an external input. Further, this definition indicates that any dynamics of IntVars are caused only by interactions with other areas and updates in the control signals. Also, IntVars are a function of local variables only, allowing a decoupled system representation. Physically, the IntVars represent the stored net energy imbalance in each control area, and inform the system operator about the amount of frequency control service needed in each control area. To derive the matrix T^i in the definition, we first combine equation (3.1) with $z = T^i x$:

$$\dot{z}^i = T^i A^i x^i + T^i B^i u^i + T^i d^i + T^i F^i \tag{3.3}$$

Under the conditions stated in the definition, $u^i(t) = 0$, $d^i(t) = 0$, $F^i = 0$, this simply gives:

$$\dot{z}^i = T^i A^i x^i \tag{3.4}$$

Thus, the matrix T^i can be obtained from:

$$T^i A^i = 0 \tag{3.5}$$

and corresponds to the left eigenvector corresponding to the zero eigenvalue of A^i. In power systems, the system matrix A^i of any area i has inherent structural singularity, as a direct consequence of the power conservation law, so T^i will always exist. More specifically, the interaction variable of a lossless system can always be defined as a linear combination of net real power injections of the components in the area [22], i.e.

$$T^i = [0 \ 0 \ldots t_p^i], \ t_p^i \in \mathbb{R}^{1\times n} \tag{3.6}$$

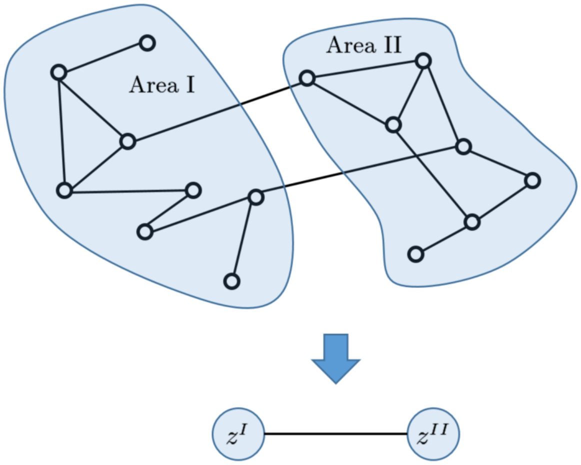


Figure 3.1: Illustration of aggregation via IntVar in a two-area system

and therefore

$$z^i = t_p^i P^i \tag{3.7}$$

where P^i is a vector of net real power injections at each bus in area i. A slightly more involved method for deriving T^i when grid losses are included can be found in the same reference.

By definition, interaction variables are defined for disconnected regions, so they are *local* variables associated with each area. Therefore, they represent aggregate behavior of an area when it is interconnected with the rest of the system. Figure 3.1 illustrates the aggregation of a two-area system using IntVar.

3.2 Clustering-based aggregation

In this section, we introduce a modified version of the aggregation method proposed in [41], for clustering measurements according to their dynamic response to the external input/disturbance $d(t)$ in the linear closed-loop system:

$$\Sigma : \begin{cases} \dot{x}(t) & = \mathcal{A}x(t) + Gd(t) \\ y(t) & = Cx(t) \end{cases} \tag{3.8}$$

We begin by introducing the definition of a cluster.

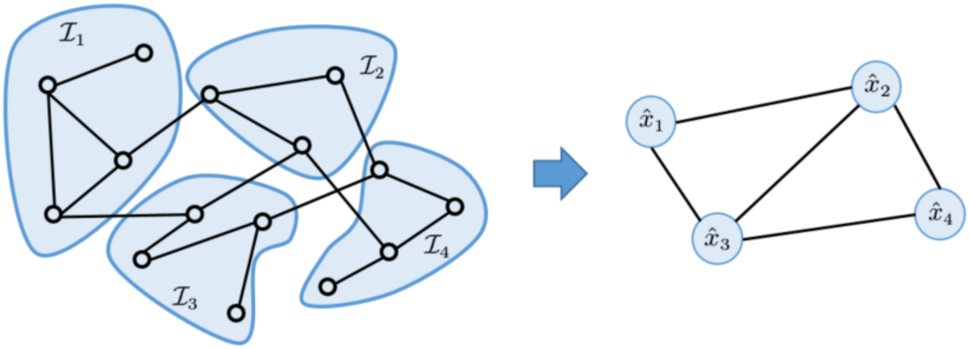


Figure 3.2: Illustration of clustering-based aggregation

Definition. *Let $\mathbb{L} = \{1, \dots, l\}$ be the set of measurement indices, and $\mathbb{K} = \{1, \dots, K\}$ the set of cluster indices. Then, clusters $\mathcal{I}_k$ (depicted in Figure 3.2), $k \in \mathbb{K}$, are defined as disjoint subsets of $\mathbb{L}$, that cover all the elements in $\mathbb{L}$, i.e. $\bigcup_{k\in\mathbb{K}} \mathcal{I}_k = \mathbb{L}$.*

With this definition in mind, we aim to partition the set $\mathbb{L}$ into clusters $\mathcal{I}_k$ such that

$$p_j g_i(s) = p_i g_j(s), \quad \forall i, j \in \mathcal{I}_k \tag{3.9}$$

where g_i is the i-th element of the input-output transfer matrix $g(s)$:

$$g(s) = C(sI_n - \mathcal{A})^{-1} G \tag{3.10}$$

of the system in (3.8). The measurements i and j belonging to the same cluster k will have a proportional, or in some cases identical, response to the input $d(t)$. In that sense, the proportionality of transfer functions g_i and g_j can also be expressed as proportionality to some scalar function $\bar{g}$ corresponding to cluster k. Therefore, we can define a condition for cluster formation in a compact way as follows.

Definition. *A set of measurements $\{y_i\}$ should form a cluster $\mathcal{I}_k$ if there exists a scalar function $\bar{g}(s)$ such that:*

$$(e^n_{\mathcal{I}_k})^T g(s) = p_k^T \bar{g}(s) \tag{3.11}$$

This definition provides intuition on the meaning of clustering in our application, but is not practical for designing a procedure that would form such clusters. For that reason, we derive an equivalent condition for cluster formation that is more practical to check, based on this definition of similarity and the notion of reachability.

Reachability Gramian of a semistable system

To that end, we first derive the reachability Gramian of a semistable system (3.8). The reachability Gramian is defined as

$$W_c = \int_0^\infty e^{\mathcal{A}t} G G^T e^{\mathcal{A}^T t} dt \tag{3.12}$$

When $\mathcal{A}$ is Hurwitz, the above integral converges, and W_c can also be found as a solution of the Lyapunov equation:

$$\mathcal{A} W_c + W_c \mathcal{A}^T + G G^T = 0 \tag{3.13}$$

However, in power systems, the system matrix $\mathcal{A}$ has an inherent structural singularity, as a direct consequence of power conservation law. Due to semistability of the system matrix $\mathcal{A}$, the integral in (3.12) may not converge. To compute the reachability Gramian of a semistable system, we first consider the decomposition of $\mathcal{A}$ where $0 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n$

$$\mathcal{A} = U \Lambda U^{-1} = [u_{max} \quad \bar{U}] \begin{bmatrix} 0 & \\ & \bar{\Lambda} \end{bmatrix} \begin{bmatrix} v_{max}^T \\ \\ \bar{V}^T \end{bmatrix}$$

where u_{max} and v_{max} are the right and left eigenvectors corresponding to the largest eigenvalue $\lambda_1 = 0$, and $\bar{\Lambda}$ is diagonal and Hurwitz. Let the stable subspace of Σ, $(\bar{\mathcal{A}}, \overline{G})$ given by:

$$\bar{\mathcal{A}} = \overline{V}^T \mathcal{A} \overline{U} \tag{3.14}$$

$$\overline{G} = \overline{V}^T G \tag{3.15}$$

Then, the reachability Gramian of the stable subspace is the solution of

$$\bar{\mathcal{A}}\, \overline{W}_c + \overline{W}_c \bar{\mathcal{A}}^T + \overline{G}\overline{G}^T = 0 \tag{3.16}$$

Substituting $\bar{\mathcal{A}}$ and $\overline{G}$ into (3.12) yields

$$\begin{aligned}\overline{W}_c &= \int_0^\infty e^{\bar{\mathcal{A}}t}\overline{GG}^T e^{\bar{\mathcal{A}}^T t}dt \\ &= \int_0^\infty \overline{V}^T e^{\mathcal{A}t}\bar{U}\overline{V}^T GG^T \overline{V}\bar{U}^T e^{\mathcal{A}^T t}\overline{V}dt \\ &= \overline{V}^T W_c \overline{V}\end{aligned}$$

and

$$W_c = \overline{VW}_c\overline{V}^T \tag{3.17}$$

is the reachability Gramian of the semistable system Σ and contains information on the degree of reachability of states with respect to the input $d(t)$. In the following theorem we show that the condition in (3.11) is equivalent to linear dependence of rows of a matrix Φ.

Theorem. *Consider the reachability Gramian W_c in* (3.17) *of the semistable system Σ in* (3.8)*. Furthermore, let the Cholesky factorization of W_c be given by $W_c = W_L W_L^T$, and $\Phi = CW_L$. Then, the condition in* (3.11) *is equivalent to*

$$(e^n_{\mathcal{I}_k})^T\Phi = p_k^T\bar{\phi} \tag{3.18}$$

where $\bar{\phi} \in \mathbb{R}^{1\times n}$ is a constant vector.

Proof. In order for (3.11) to hold, for each $i, j \in \mathcal{I}_k$ it must hold that

$$p_j\|g_i(s)\|_{\mathcal{H}_2} = p_i\|g_j(s)\|_{\mathcal{H}_2}.$$

Similarly, (3.18) is equivalent to

$$p_j\|\Phi_i\| = p_i\|\Phi_j\|$$

where Φ_i is the ith row of the matrix Φ. The $\mathcal{H}_2$-norm of a linear system can be computed as the $\mathcal{L}_2$-norm of its impulse response $h(t)$:

$$\|g(s)\|^2_{\mathcal{H}_2} = \|h(t)\|_2^2 = \mathrm{tr}\left\{C\int_0^\infty e^{\mathcal{A}t}GG^Te^{\mathcal{A}^Tt}dtC^T\right\}$$

Plugging in (3.17), we have

$$\|h(t)\|_2^2 = \mathrm{tr}\left\{C\overline{V}\left[\int_0^\infty e^{\bar{\mathcal{A}}t}\overline{GG}^Te^{\bar{\mathcal{A}}^Tt}dt\right]\overline{V}^TC^T\right\}$$

For $\|h(t)\|_2^2$ to be finite, the integral above must be finite. Since $\bar{\mathcal{A}}$ and $\overline{G}$ are the stable subspace of Σ, we have

$$\lim_{t\to\infty} e^{\bar{\mathcal{A}}t} = 0$$

and $\|h(t)\|_2^2$ is finite and equal to:

$$\begin{aligned} \|g(s)\|_{\mathcal{H}_2}^2 = \|h(t)\|_2^2 = \operatorname{tr}\{CW_cC^T\} = \operatorname{tr}\{CW_LW_L^TC^T\} = \\ = \|CW_L\|_F = \|\Phi\|_F \end{aligned} \tag{3.19}$$

where $\|\cdot\|_F$ is a vector norm applied to each row of Φ. Hence, (3.11) is equivalent to (3.18). □

However, in real systems, the identity in (3.11) is almost never the case. Therefore, we will relax the strict equality, and require

$$\|p_jg_i(s) - p_ig_j(s)\|_{\mathcal{H}_2} \le \varepsilon, \quad \forall i,j \in \mathcal{I}_k \tag{3.20}$$

to hold for each cluster. Equivalently, we can check for linear dependence between rows of matrix Φ:

$$\|p_j\Phi_i - p_i\Phi_j\| \le \theta \quad \forall i,j \in \mathcal{I}_k \tag{3.21}$$

where $\theta > 0$ and Φ_i is the i-th row of Φ. Here, θ is a parameter that allows us to control the coarseness of clustering. In other words, it allows us to find outputs that have a "similar", instead of equal, response, which relaxes the condition (3.11). The smaller θ is, more accurate the clustering will be, but the clusters may contain very few measurements, which is not desirable for attack detection purposes. On the other hand, if θ is too large, the aggregation error will be high, which may obscure stealthy attacks so that they remain undetected. This trade-off should be considered when choosing a particular value for θ depending on the particular application.

Finally, we can introduce the measurement clustering algorithm defined above. Assume k clusters have already been formed. First, we choose an index i that hasn't already been assigned to any cluster, and add it to cluster $k+1$. Then, we choose another index j that is not yet assigned to a cluster, and check condition (3.21) for i and j. If the condition is satisfied, we add j to cluster $k+1$. We repeat this process until all measurements are assigned to a cluster. This procedure is summarized in the algorithm below.

Algorithm 1 Clustering algorithm

```
**Initialize** cluster index $k = 0$, and cluster set $\mathbb{K} = \emptyset$
**repeat**
    **Choose** measurement index $i \in \mathbb{L}$ that hasn't been
        assigned to a cluster yet, and add it to cluster $\mathcal{I}_{k+1}$
    **Set** $k \leftarrow k + 1$, $\mathbb{K} \leftarrow \{\mathbb{K}, k\}$
    **Find** all $j \in \mathbb{L}$ that haven't been assigned to a cluster
        yet and that satisfy (3.21) and add them
        to $\mathcal{I}_{k+1} \leftarrow \{\mathcal{I}_{k+1}, j\}$
**until** all measurements are assigned to a cluster, i.e. $\bigcup_{k \in \mathbb{K}} \mathcal{I}_k = \mathbb{L}$
```

3.2.1 Interpretation and examples of clustering-based aggregation in power systems

It should be noted that the nature of power systems highly influences the clustering procedure. In other words, measurements of states tend to group according to type: generator net real power injection $P_{G,i}$ and mechanical power output $P_{T,i}$ measurements tend to be in the same group, while load net real power injections $P_{L,i}$ group together, and so do the frequencies ω_i. This is illustrated on the 5-bus test system [42]. Figure 3.3 depicts the one-line diagram and the graph representation of the full system.

Also, component and network parameters have an influence on cluster boundaries. Two generators with the same inertia, damping and controller gains will naturally have very similar dynamic response. Also, frequencies of components connected by a line with a large susceptance (i.e. small impedance), tend to have similar magnitude of oscillation. Examples of how the clusters change depending on line parameters are depicted in Figures 3.4 and 3.5.

3.3 Summary

In this chapter, we review the technical background needed for development of security methods for cyber-physical systems proposed in the rest of this thesis. We argue that methods that take advantage of internal structure of the system need to be designed and implemented in order to ensure security of cyber-physical systems. In this chapter, we propose an aggregation frame-

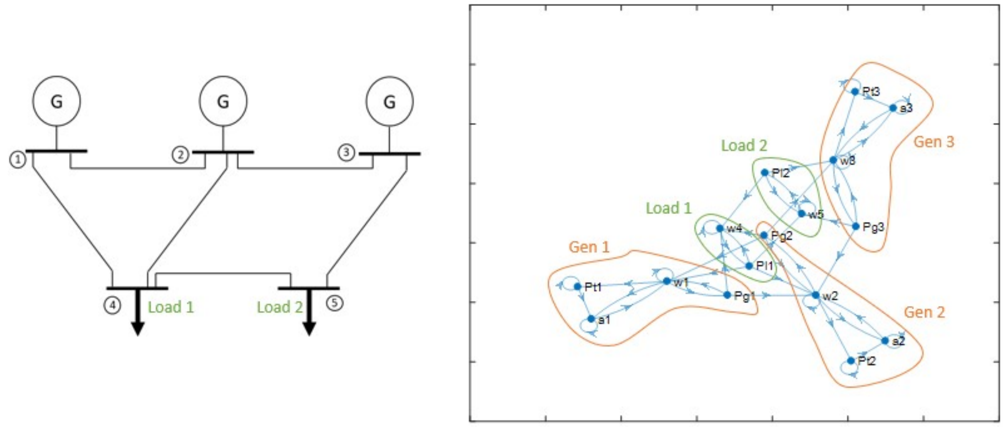


Figure 3.3: One-line diagram (left) and graph representation (right) of the 5-bus system

work that can be used to extract this internal structure. To that end, we review two aggregation methods existing in the literature. First, we introduce the notion of Interaction Variables, that represent aggregate imbalance in a control area. Then, we introduce a modified version of the clustering-based aggregation algorithm. Using this approach, outputs of the system are grouped according to their response to an exogenous disturbance. Finally, we provide an interpretation and examples of the clustering-based aggregation in a 5-bus power system.

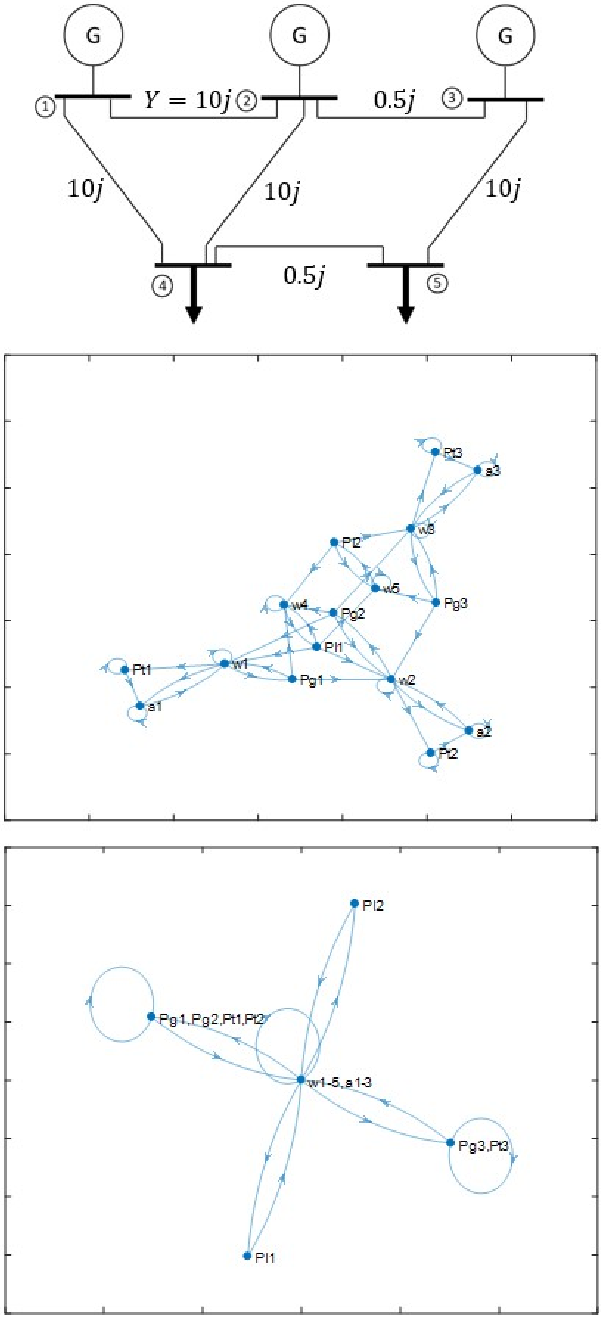


Figure 3.4: One-line diagram of the 5-bus system with line parameters (top), graph representation of the full system (middle), graph representation of the clustered system (bottom). States of generators 1 and 2 are clustered due to the weak line connection to bus 3.

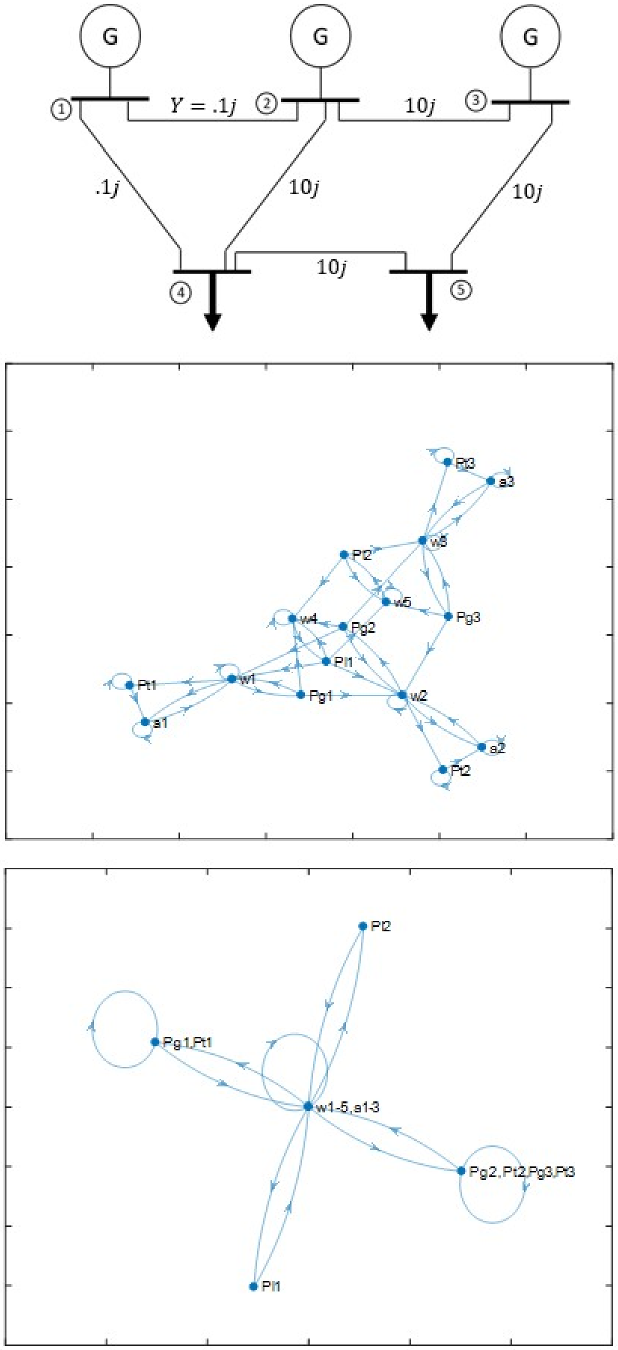


Figure 3.5: One-line diagram of the 5-bus system with line parameters (top), graph representation of the full system (middle), graph representation of the clustered system (bottom). States of generators 2 and 3 are clustered due to the weak line connection to bus 1.

Chapter 4

Moving-target Active Attack Detection: a Clustering-based Approach

The first step to responding to an attack is detection. The threat of stealthy attacks motivates the study of active detection schemes, where the defender modifies parts of the system to discover adversarial behavior. For instance, under certain scenarios, an attack is stealthy only if a defender uses particular control and detection policies. In this case, a defender can actively detect an adversary by changing their strategy. In this chapter, we adopt a moving-target approach to active detection. We show that using a constantly changing detection policy allows detection of stealthy attacks. As a basis of our proposed moving-target detection filter, we use the concept of output clustering introduced in Chapter 3. Clustering of the outputs gives the defender an upper hand, by providing additional information on the system, unknown to the attacker.

We review the state-of-the-art of cyber-attack detection in Section 4.1. The system and attack models are introduced in Sections 4.2 and 4.3, respectively. Our proposed moving-target approach is derived in Section 4.4, and tested on the 5-bus test system in Section 4.5, and the IEEE RTS 24-bus system in Section 4.6. Finally, the chapter is concluded in Section 4.7. The results in this chapter are partially based on [43].

4.1 State-of-the-Art on attack detection in cyber-physical systems

The literature on this topic is constantly growing, but securing power systems against cyber-attacks is still an open problem [44], [45], [46]. The complex and highly distributed nature of the electric power system, as well as diversity of its components and control designs needed to operate it, make cyber-physical security a challenging problem.

A malicious intruder can have detailed knowledge of the system model and parameters, including any defense strategies, and can therefore launch highly effective and deceptive attacks. Stealthy attacks [9], replay attacks [11], and zero-dynamics attacks [12] are all examples of how an adversary can exploit knowledge of the system to launch attacks that will not be detected by the existing systems in power system control centers. Specifically, state estimation (SE) in power systems is supported by a static failure detector, called Bad Data Detection (BDD). Static attack detectors do not consider system dynamics, but only the outputs of the system, which they check for consistency at every time step [47, 48]. Limitations of these techniques have been often underlined, especially by a knowledgeable attacker [49–51]. It is important to mention, however, that static estimation and detection algorithms have been in use in power systems for many years for practical reasons. For one, there were fewer and less frequent measurements available in the past, due to low bandwidth for communication between the field devices and the control center. Another consequence of that is that a sufficiently detailed model of the system's dynamics was hard to obtain and tune. With the recent technological improvements in communication networks, the advent of advanced sensors (e.g. PMUs), and model identification techniques [52], these limitations can be overcome.

Dynamic detection has been approached via heuristics and expert systems [53]. Reliability and accuracy when dealing with unforeseen system anomalies, as well as the absence of analytical performance guarantees, are some of the shortcomings of these methods. A different approach, based on comparing a discrete-time state transition map to a series of past measurements via Kalman filtering, can be found in [54] and references therein. Typically, these transition maps are based on heuristic models valid only around a specific operating point. However, such models poorly describe the complex dynamics of the power system and suffers from drawbacks similar to those of expert systems methods. In [55] a graph-theoretic framework is proposed to evaluate the impact of cyber-attacks on the smart grid.

Other recent approaches to dynamic detection consider continuous-time power system models and apply dynamic techniques [56, 57]. While [57] adopts an oversimplified model neglecting the network and load models, the references [49, 56, 58] use a more accurate network descriptor model. References [59], [60], [61] all provide dynamic attack detection algorithms for various particular settings. However, [49] also shows that no dynamic detector can counter stealthy attacks, as these attacks alter the output of the system in a way that could also be a result of normal system behavior. For that reason, another class of detectors has been introduced, namely *active* attack detectors. In contrast to passive detectors, active detectors perturb the system either through topology changes, or by injecting random signals into the network, in order to expose stealthy attacks.

One recent approach to active defense introduces an additional random signal, or "watermark", to the control signal as a form of authentication [62], [63], [11]. In normal operation, this watermark should also be present in the measurement signal, so it's absence suggests that the system has been tampered with. This is a good defense strategy, especially against replay attacks, but it is not effective in the case the attacker has extensive knowledge of the system model and the watermark. Another approach is to reveal the stealthy attacks by modifying the system's structure. Specifically, new measurements can be added incrementally to reveal stealthy zero-dynamics attacks [64]. Even though this strategy effectively increases the robustness of the system, it is only successful for attacks that are constructed off-line, and once they are launched, the adversary can't gain new information about the changes in the structure of the system. Coding sensor outputs [65], [66] is an economical way of detecting stealthy FDI attacks when the attacker knows the system model without the coding scheme. However, similarly to the previous approaches, this strategy fails when the attacker has extensive knowledge of the system. In [67] the authors posture that the attacker without previous knowledge can first identify the system model by observing the control and measurement signals. Then they provide a control design method which renders the system unidentifiable, but with a performance trade-off.

4.2 System description

We consider a cyber-physical system, where the physical layer of the system can be described by a set of linear dynamic equations:

$$\dot{x}(t) = Ax(t) + Bu(t) + Gd(t) \tag{4.1}$$

where the states of system are denoted $x \in \mathbb{R}^n$, the disturbance is $d \in \mathbb{R}^m$, and the control signal is $u \in \mathbb{R}^p$. In the cyber layer, a large network of field sensors is deployed to monitor the operation of the system in (4.1):

$$y(t) = Cx(t) \tag{4.2}$$

The collected measurements are then used to compute the control signal for the actuators. Depending on the purpose, a control algorithm can be implemented as either using output or state feedback. In output feedback control measurements are directly used to compute the control signal, i.e.

$$u(t) = f(y(t)) \tag{4.3}$$

where $f()$ is a linear function.

On the other hand, in state feedback control applications, a current state of the system first needs to be estimated. In that case, a state estimator such as an observer or Kalman filter is used to infer the current state $\hat{x}(t)$ from the received measurements $y(t)$. Then, the estimate $\hat{x}(t)$ is used to compute the control signal:

$$u(t) = f(\hat{x}(t)) \tag{4.4}$$

where $f()$ is a linear function.

For the purposes of cyber-security, we assume that the control algorithm and state estimator, if used, are designed appropriately, so that the closed-loop system is stable and regulated to achieve desired performance objectives. Thus, we can describe the stable closed-loop system as:

$$\Sigma : \begin{cases} \dot{x}(t) & = \mathcal{A}x(t) + Gd(t) \\ y(t) & = Cx(t) \end{cases} \tag{4.5}$$

where $\mathcal{A}$ is a stable closed-loop system matrix.

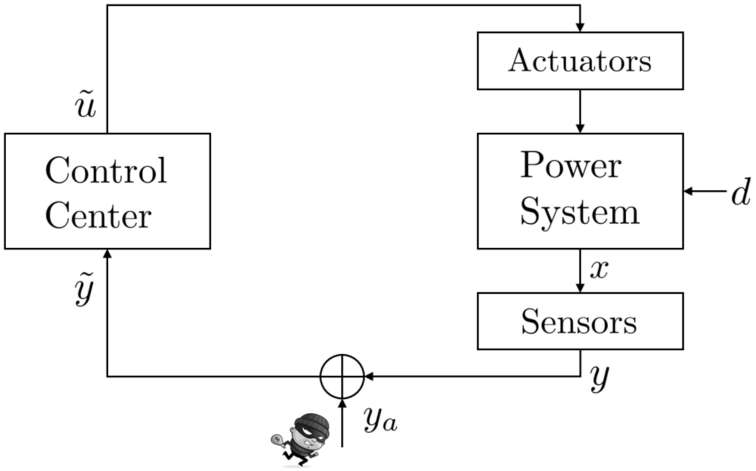


Figure 4.1: Block diagram of attacked system in (4.8). The attacker injects signal y_a into measurements y in order to manipulate the system.

4.3 Attack model

In case of False Data Injection (FDI) attacks, an attacker is assumed to be able to modify the measurement signal $y(t)$ arbitrarily, by injecting a the signal $y_a(t)$:

$$\tilde{y}(t) = y(t) + y_a(t) \tag{4.6}$$

As a consequence, the controller will receive and base its decisions on the corrupted measurement signal $\tilde{y}(t)$:

$$\tilde{u}(t) = f(\tilde{y}) = f(y(t) + y_a(t)) \tag{4.7}$$

Since f is a linear function, the control signal can then be decomposed as $\tilde{u}(t) = u(t) + u_a(t)$, where $u(t) = f(y(t))$ and $u_a(t) = f(y_a(t))$. Thus, the closed-loop attacked system can be described using:

$$\Sigma_a : \begin{cases} \dot{x}(t) & = \mathcal{A}x(t) + Bu_a(t) + Gd(t) \\ \tilde{y}(t) & = Cx(t) + y_a(t) \end{cases} \tag{4.8}$$

where $\mathcal{A}$ is the closed-loop system matrix. Figure 4.1 depicts the block diagram of the attacked system.

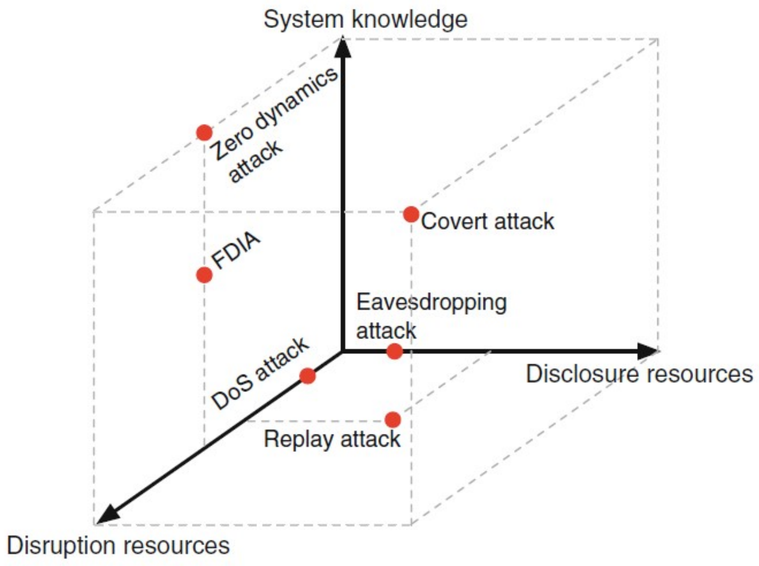


Figure 4.2: Attack space for cyber-physical systems [68]

Attacker's knowledge and resources

Many well-known attack schemes can be categorized based on the resources available to the attacker as depicted in Figure [68]. *Knowledge of the system model* is arguably the most important component of the attack model since it can be used to construct complex and stealthy attacks with significant impacts on the physical system. Disclosure resources refer to components such as communication channel that can be accessed during the attack and enable the adversary to violate data *confidentiality* by gathering sensitive information about the system. The disclosure attack, or eavesdropping, cannot inflict damage to the physical system, but can be used to construct more complex attacks to affect the system. On the other hand, disruption resources violate the *integrity* of data, and can be used by the adversary to manipulate the system operation.

In this thesis, we consider the family of stealthy attacks, also called covert attacks, which are designed to replicate normal system behavior, and will not be detected via existing static detection schemes (BDD in power systems).

Definition. *For the attacked linear system in* (4.8), *the attack* y_a *is called* *stealthy if and only if* $\tilde{y}(x_1, 0, t) = \tilde{y}(x_2, y_a, t)$ *for some initial conditions* $x_1, x_2 \in \mathbb{R}^n$ *and for* $t \geq t_0$.

Further, we assume the adversary has the following knowledge and resources:

- The attacker has access to the real time sensor measurements, and knows the true sensor outputs $y(t)$.
- The attacker has the ability to compromise integrity of the real time sensor measurements. Specifically, they can replace the true signal $y(t)$ with the arbitrary signal $\tilde{y}(t)$.
- The attacker has knowledge of the system matrices A, B, C, G, and can construct stealthy attacks.
- The attacker can gain knowledge of the current defense policy, and how it is generated. However, the attacker does **not** know when the policy has changed.

Attack detection problem

If there is no attack on the system, y_a will be a zero vector. Otherwise, $y_a \neq \mathbf{0}$. Thus, for a system defined as above, a cyber-attack on measurements can be detected using the residual:

$$r = \tilde{y} - y \tag{4.9}$$

when the residual r is nonzero. In a more general case, where process and measurement noise is considered, a threshold ϵ can be used to adjust for effects of noise:

$$r = |\tilde{y} - y| > \epsilon \tag{4.10}$$

However, the true measurements y are not known in presence of FDI attacks. In the rest of this chapter, we present a moving-target active attack detection filter. We use aggregation as a tool to uncover an internal structure in the vector y. If the same structure is not present in the vector $\tilde{y}$, we can infer that one or more measurements have been corrupted by the attacker.

4.4 Active clustering-based detection filter

In this section, we will introduce a cyber-attack detection algorithm that employs the output clustering outlined in Algorithm X. Two properties of this clustering method are key in our cyber-attack detection filter design. Firstly, we know that, once clustering is performed on the system in normal operating conditions, the outputs within the clusters will be approximately proportional to each other at all times t. That enables us to perform quick consistency checks to ensure the safety and reliability of the system. Secondly, the result of clustering will change over time as operating conditions change, which means our proposed detection filter will behave as a moving-target.

In the analysis in Chapter 3, we have shown that clusters can be formed such that measurements i, j within the cluster $\mathcal{I}_k$ are approximately proportional, i.e. $a_i y_i(t) \approx a_j y_j(t) \approx \cdots \approx z^{(k)}(t)$. This relation of measurements within the same cluster can also be written as:

$$\hat{y}^{(k)}(t) = \left[p_1^{(k)} \, \dots \, p_{|\mathcal{I}_k|}^{(k)}\right]^T z^{(k)}(t) \tag{4.11a}$$

$$\text{such that} \quad \|y^{(k)}(t) - \hat{y}^{(k)}(t)\| \leq \theta, \quad \theta \geq 0 \tag{4.11b}$$

where $p_i = a_i^{-1}$ for all $i \in \mathcal{I}_k$, and $y^{(k)}$ is a subset of measurements y belonging to cluster $\mathcal{I}_k$, $y^{(k)} = (e^n_{\mathcal{I}_k})^T y$, and all its elements are approximately proportional to a single scalar variable $z^{(k)}$. Parameter θ is a design parameter, which will be discussed in later sections.

Specifically, we use the knowledge of the fact that incoming measurements $\tilde{y}^{(k)}(t) = y^{(k)}(t) + y_a^{(k)}(t)$ belonging to cluster k also have the property in (4.11a) only if $y_a^{(k)} \equiv 0$. Therefore, we define the set of residuals $r_{i,j}(t)$ that exploit this property as:

$$r_{i,j}(t) = ||p_j \tilde{y}_i(t) - p_i \tilde{y}_j(t)||, \quad \forall i, j \in \mathcal{I}_k \tag{4.12}$$

The residuals $r_{i,j}$ defined above will have a value larger than some threshold ε only in presence of cyber-attacks.

Finally, we show that the original system outputs $y = Cx$ can be approximated by $\hat{y} = \Pi^T z$, where $z = \Pi y = \Pi C x$. The clustering matrix $\Pi \in \mathbb{R}^{K \times n}$ is defined as:

$$\Pi := \text{Diag}\{p_1, p_2, \dots, p_K\} E \in \mathbb{R}^{K \times n} \tag{4.13}$$

where E is a permutation matrix defined as $E = [e^n_{\mathcal{I}_1}, \dots, e^n_{\mathcal{I}_K}]^T$. The input-output transfer matrix associated with $\hat{y}$ can then be defined as

$$\hat{g}(s) = \Pi^T \Pi \left(sI_n - \mathcal{A}\right)^{-1} G \tag{4.14}$$

The following theorem establishes that $\hat{y}$ is a good approximation of y, and that it can be used in our cyber-attack detection methodology.

Theorem. *Consider a semistable linear system in* (4.5). *Consider also a clustering-based approximation $\hat{y}$ obtained using the aggregation matrix* Π. *Then, the error system of the approximation $g_e(s) = g(s) - \hat{g}(s)$ is asymptotically stable.*

Proof. By definition (4.13), Π is a unitary matrix, i.e. $\Pi\Pi^T = I_K$, and $\Pi^T\Pi$ is an orthogonal projection onto colspace(Π^T). Note also that, by definition, $v_{max} \in$ colspace(Π^T). We then define $\bar{\Pi}$ as an orthogonal complement of Π, such that $[\Pi^T\ \bar{\Pi}^T]^T$ is unitary, and $I - \Pi^T\Pi = \bar{\Pi}^T\bar{\Pi}$. Consider now the error system $g_e = g - \hat{g}$ of the approximation:

$$\begin{aligned} g_e(s) &= C(sI - \mathcal{A})^{-1}G - \Pi^T\Pi C(sI - \mathcal{A})^{-1}G = \\ &= (I_n - \Pi^T\Pi)C(sI - \mathcal{A})^{-1}G = \bar{\Pi}^T\bar{\Pi}g(s) \end{aligned} \tag{4.15}$$

Then $\Pi^T\Pi v_{max} = v_{max}$, or equivalently

$$\bar{\Pi}v_{max} = 0.$$

This implies that there is pole-zero cancellation in $\bar{\Pi}g(s)$ associated with the zero eigenvalue. Therefore, all poles of $\bar{\Pi}g(s)$ have negative real parts, and the error system g_e is asymptotically stable. □

Now, we introduce the moving-target detection algorithm based on dynamic clustering of system outputs. Firstly, the moving-target nature of our proposed method stems from the natural fluctuations occurring in power systems. As the underlying power system model is nonlinear, the system matrices $\mathcal{A}$ and G are only valid around a certain operating point x^0, and we will denote them with $\mathcal{A}(x^0)$ and $G(x^0)$. As a result, cluster boundaries have to be recomputed approximately every one hour, for the current operating conditions x^0. In this method, additional computations are traded for increased difficulty of compromising the integrity of the system for the attacker. In other words, even if we assume the attacker has complete knowledge of the system and the detection strategy at one point in time, that knowledge will eventually become unusable to construct stealthy attacks, as our detection strategy is dynamic.

Once the new operating point is received from the Control Center, the linearized system matrices and clusters need to be recomputed. Then, at every time step of the control processes, the incoming measurements must be verified using residuals of the clustering-based detection filter given in (4.12) before the control action is computed and performed.The proposed cyber-attack detection algorithm, based on dynamic clustering, is outlined in Algorithm 2.

Algorithm 2 Clustering-based Cyber-attack Detection Method

```
repeat
    Get new operating conditions $x^0$
    Compute matrices $\mathcal{A}(x^0)$, $G(x^0)$, and find cluster
      sets $\mathcal{I}_k$ according to Algorithm 1
    for every time-step of the control process do
        for all measurements $y_i, i \in \mathcal{I}_k$ do
            if condition in (4.12) is satisfied then
                → cyber-attack detected
            end if
        end for
    end for
until new operating conditions $x^0$ are received
```

4.5 Numerical example on the 5-bus system

In this section, we present the effectiveness of the proposed detection algorithm on the 5-bus test system [42], shown in Figure 4.3. This system has three synchronous generators, on buses 1, 2 and 4, and two loads, on buses 3 and 5. In this example, we consider a simulation study of the system's response to load variation at bus 3. First, output clustering of frequency and real power injection measurements is performed, and given in Table 4.1.

	Clustered states
Cluster 1	P_{G_1}, P_{G_2}, P_{L_1}
Cluster 2	P_{G_3}, P_{L_2}
Cluster 3	ω_{G_1}, ω_{G_2}, ω_{L_1}
Cluster 4	ω_{G_3}, ω_{L_2}

Table 4.1: Result of clustering; 5-bus system example.

Figures 4.4 and 4.5 depict the response of the system to this load variation (solid lines), as well as the cluster variable for both clusters (dotted lines).

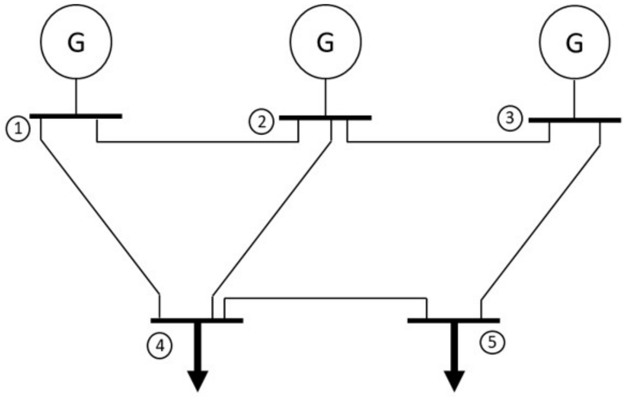


Figure 4.3: 5-bus test system [42]

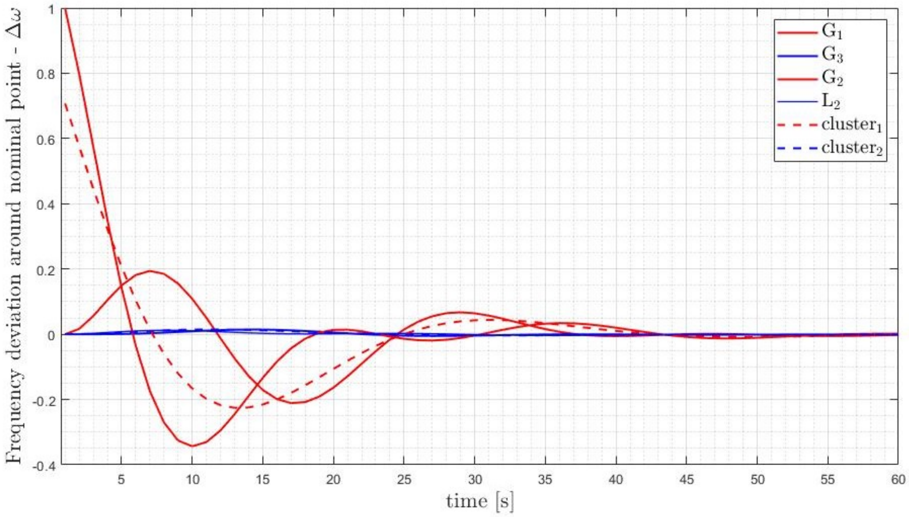


Figure 4.4: Frequency response to the load variation (solid lines: cluster 1 - red, cluster 2 - blue), and cluster variables (dotted lines)

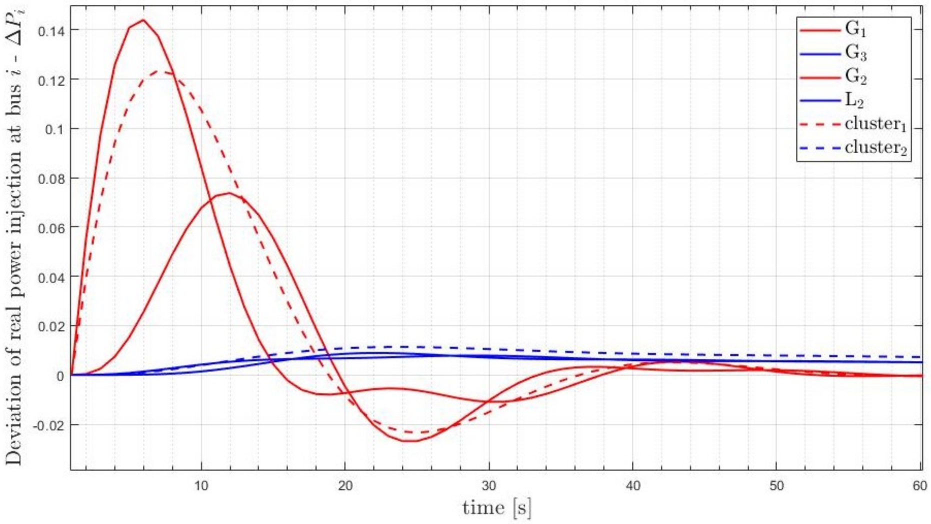


Figure 4.5: real power response to the load variation (solid lines: cluster 1 - red, cluster 2 - blue), and cluster variables (dotted lines)

4.6 Numerical example on the 24-bus system

The IEEE RTS 24-bus system [69], depicted in Figure 4.6, consists of 10 generators, equipped with governor control, and 14 loads.

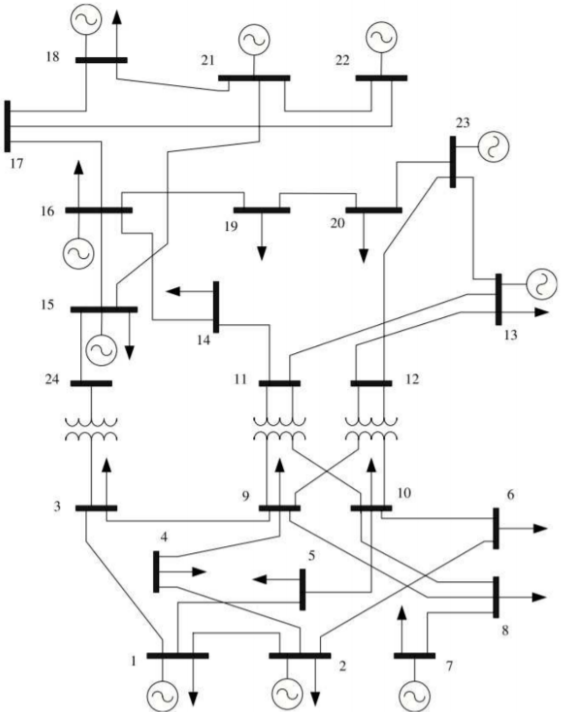


Figure 4.6: IEEE RTS 24-bus system [69]

For this system, we will first consider two scenarios with different loading conditions, to demonstrate the detection algorithm as well as the moving-target defense strategy:

- Scenario 1 - the system is at high loading condition. From $t = 0$ to 20 s, the loading is nominal. At time $t = 20$ s, load at bus 3 increases by 0.5 p.u., and at time $t = 200$ loading returns to nominal value.

- Scenario 2 - the system is at low loading condition. From $t = 0$ to 20 s, the loading is nominal. At time $t = 20$ s, load at bus 3 increases by 0.5 p.u., and at time $t = 200$ loading returns to nominal value.

We use these two scenarios to demonstrate the clustering method introduced in Chapter 3, and how cluster boundaries change with the operating conditions. This is demonstrated in Table 4.2, where the net real power injection of generator 8, P_{G_8}, and its mechanical power output P_{T_8} are clustered with respective states of generators 3 and 10 under Scenario 1, and generators 4 and 5 under Scenario 2.

	Clustered states
Scenario 1	$P_{G_3}, \boxed{P_{G_8}}, P_{G_{10}}, P_{T_3}, \boxed{P_{T_8}}, P_{T_{10}}$
Scenario 2	$P_{G_4}, P_{G_5}, \boxed{P_{G_8}}, P_{T_4}, P_{T_5}, \boxed{P_{T_8}}$

Table 4.2: P_{G_8} and P_{T_8} belong to different clusters as operating conditions change

Additionally, for the same $\theta = 5e^{-3}$, the clustering procedure resulted in 21 clusters under Scenario 1, and 23 under Scenario 2.

In Figure 4.7 we show the dynamic response of one of the clusters under Scenario 1. In this particular case we can also see the effect of the coarseness parameter θ: for higher detection accuracy, we can decrease θ in which case the cluster in Figure 4.7 would split into two. Appropriate attack analysis and parameter tuning is necessary in general case, but we use reasonable values in a common attack scenario to demonstrate our method.

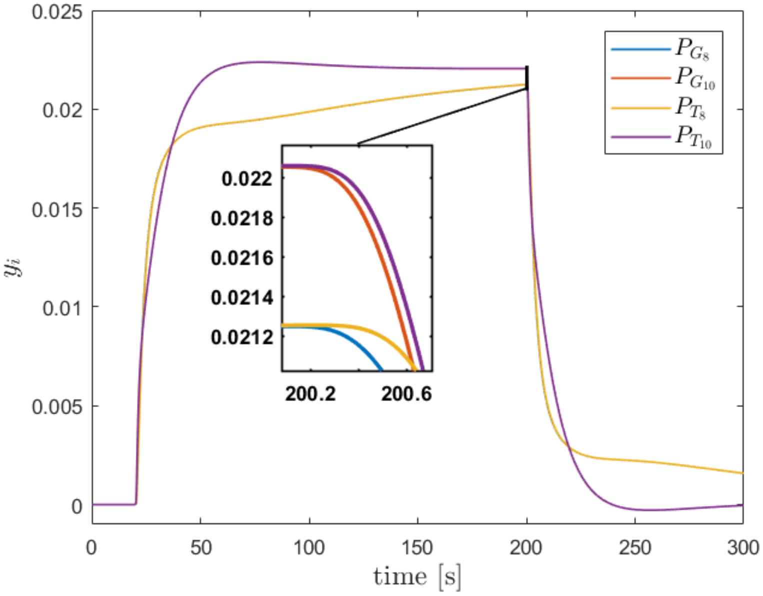


Figure 4.7: Dynamic response of measurements in one of the clusters under Scenario 1

4.6.1 Implementation of the attack detection methodology and result analysis

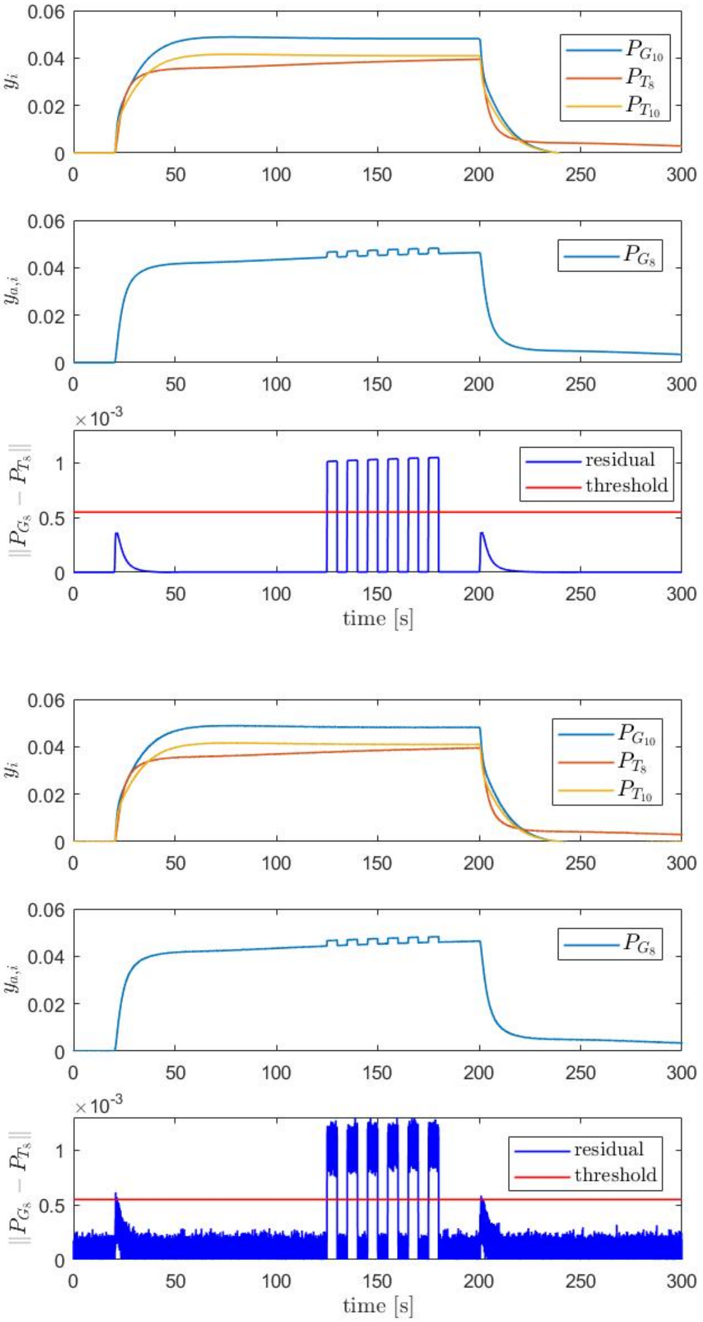
Scaling attack

We first consider Scenario 3 to demonstrate attack-detection capabilities of our proposed method on an optimal scaling attack defined in [70]:

- Scenario 3: the system is at high loading condition. From $t = 0$ to 20 s, the loading is nominal. At time $t = 20$ s, load at bus 3 increases by 0.5 p.u., and at time $t = 200$ loading returns to nominal value; at time $t = 125$ s, a sequence of 6 scaling attacks are launched on the measurement $P_{G,8}$, each lasting 5 seconds, with total duration of the attack $T_a = 55$ s

where a scaling attack can be represented in terms of system in (4.8) as $\tilde{y} = y + y_a$, where $y_a = k \cdot y$. We use scaling coefficient $k = 0.1$, which corresponds to a 10 % increase in value of $P_{G,8}$ at the time of the attack.

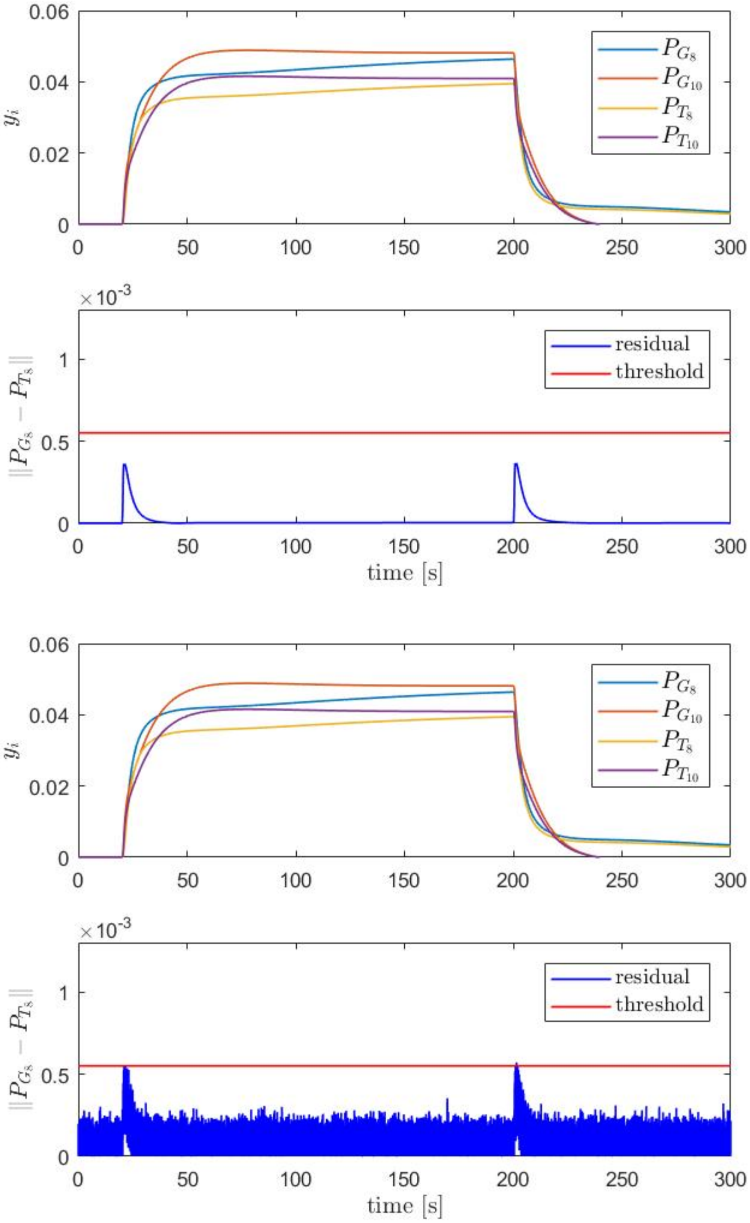
In Figure 4.8 a) we consider the noiseless scenario, and compare the attacked measurement $P_{G,8}$ (middle plot) with other measurements belonging to the same cluster (top plot), to obtain the residual in the bottom plot, which only crosses the chosen detection threshold for each of the attacks. In Figure 4.8 b) we consider the same attack scenario in presence of noise. Note that false positive alarms become very likely in this case, using the appropriate threshold designed for the deterministic scenario. If the noise parameters are known, additional statistical methods (e.g. hypothesis testing, etc.) may be employed to distinguish between noise and signal. In Figures 4.9 a) (no measurement noise) and 4.9 b) (with measurement noise) we show cluster measurements (top) and residuals (bottom) in absence of cyber-attacks. In noiseless scenario, the residual does not cross the detection threshold even when system conditions change, i.e. when there is a load disturbance in the system. In presence of noise, false positives are possible, and additional statistical methods can be employed to improve the performance of the attack detection filter.



(a) Top: Other cluster measurements; Middle: Measurement of $P_{G,8}$ under a scaling attack; Bottom: detection residual (blue) and threshold (red)

(b) Top: Other cluster measurements (noisy); Middle: Measurement of $P_{G,8}$ (noisy) under a scaling attack; Bottom: detection residual (noisy) and threshold (red)

Figure 4.8: Group of measurements belonging to one of the clusters under Scenario 3, under scaling attack in a) noiseless and b) noisy setting.



(a) Top: Measurements in one of the clusters under Scenario 3, in absence of cyber-attacks; Bottom: residual in absence of cyber-attacks

(b) Top: Noisy measurements in one of the clusters under Scenario 3, in absence of cyber-attacks; Bottom: noisy residual in absence of cyber-attacks

Figure 4.9: Group of measurements belonging to one of the clusters under Scenario 3, in absence of cyber-attacks, in a) noiseless and b) noisy setting.

Stealthy replay attack

In Scenario 4 we consider a stealthy replay attack to further demonstrate detection capabilities of our proposed method:

- Scenario 3: the system is at high loading condition. From $t = 0$ to 20 s, the loading is nominal. At time $t = 20$ s, load at bus 3 increases by 0.5 p.u., and at time $t = 200$ loading returns to nominal value. A stealthy replay attack is launched on measurement $P_{G,8}$: a recorded measurement from Scenario 2 (low loading condition) is used.

Figure 4.10 depicts all measurements from one of the clusters, among which one is under a stealthy replay attack ($P_{G,8}$). In the noiseless scenario, shown in Figure 4.11 a), we compare the attacked measurement $P_{G,8}$ (middle plot) with other measurements belonging to the same cluster (top plot), to obtain the residual in the bottom plot, which only crosses the chosen detection threshold during the attack. In Figure 4.11 b) we consider the same attack scenario in presence of noise. Note that false positive alarms become very likely in this case, using the appropriate threshold designed for the deterministic scenario. If the noise parameters are known, additional statistical methods (e.g. hypothesis testing, etc.) may be employed to distinguish between noise and signal.

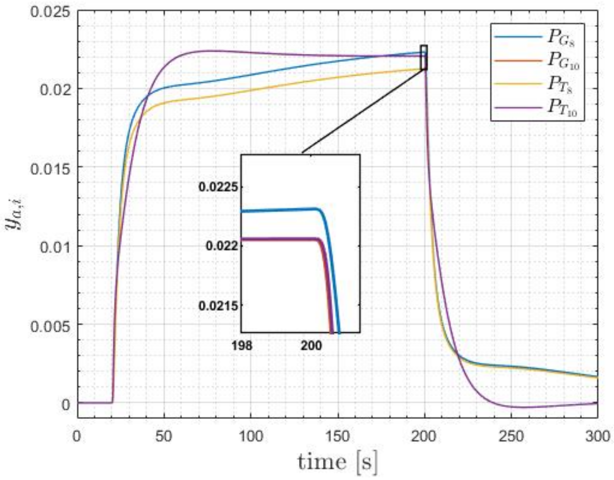
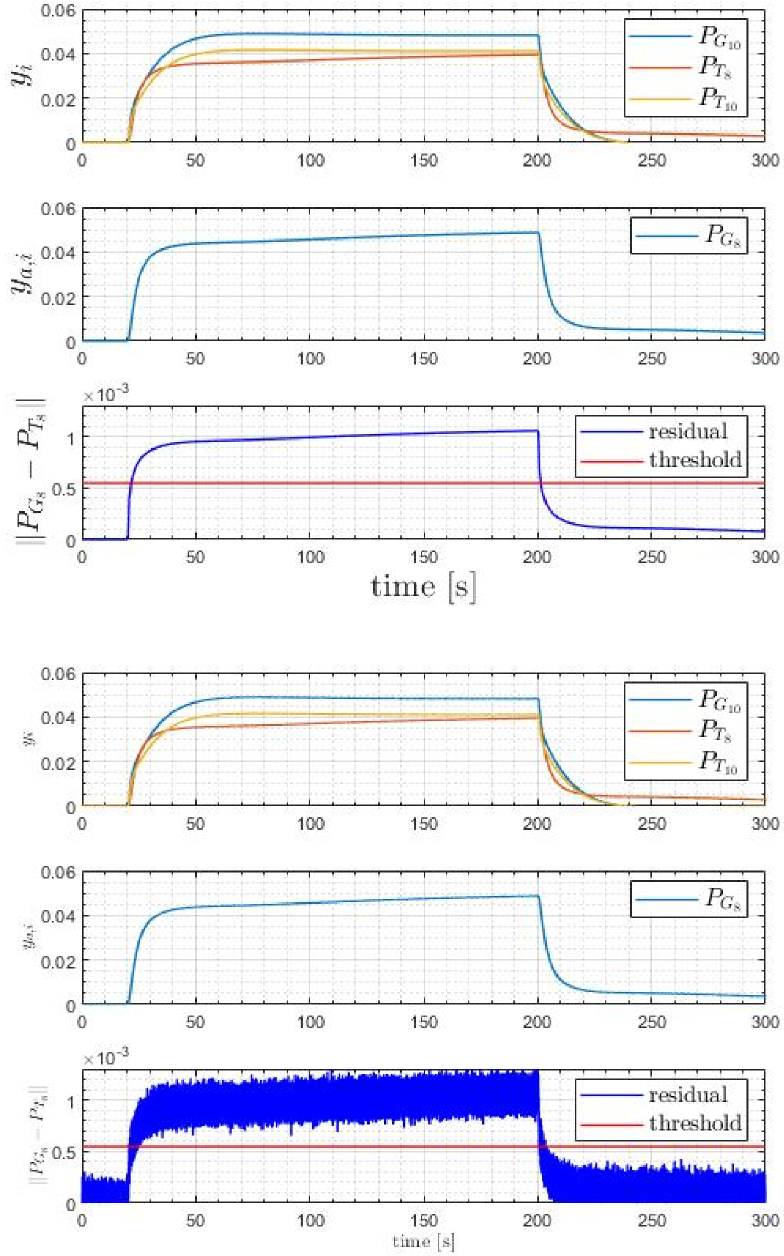


Figure 4.10: Dynamic response of measurements in one of the clusters under Scenario 4; Measurement $P_{G,8}$ is attacked



(a) Top: Other cluster measurements; Middle: Measurement of $P_{G,8}$ under a stealthy replay attack; Bottom: detection residual (blue) and threshold (red)

(b) Top: Other cluster measurements (noisy); Middle: Measurement of $P_{G,8}$ (noisy) under a stealthy replay attack; Bottom: detection residual (noisy) and threshold (red)

Figure 4.11: Group of measurements belonging to one of the clusters under Scenario 3, under scaling attack in a) noiseless and b) noisy setting.

4.7 Summary

Motivated by the threat of stealthy attacks, in this chapter we study an active detection scheme, where the defender modifies parts of the system to discover adversarial behavior. In particular, we adopt a moving-target approach, and show that using a constantly changing detection policy allows detection of stealthy attacks. As a basis of our proposed moving-target detection filter, we use the concept of output clustering introduced in Chapter 3. Clustering of the outputs gives the defender an upper hand, by providing additional information on the system, unknown to the attacker. The performance of our proposed moving-target approach is tested against stealthy attacks on the 5-bus test system and the IEEE RTS 24-bus system.

Chapter 5

Interaction Variable-based Attack Detection in Power Systems

In the previous chapter we demonstrated the effectiveness of the clustering-based moving-target approach in detecting a class of stealthy FDI attacks. The main mechanism of detection is the use of knowledge of the internal structure of the system that the defender can exploit to design the moving-target detection policy. The attacker may have the same knowledge, but does not know when the detection policy is changing. In this chapter, we make the observation that there is another intuitive way to find the internal structure in power systems that can be used for detection of stealthy attacks. In this chapter, we will specifically examine the AGC system, and its aggregation using the Interaction Variables [22].

The rest of this chapter is outlined as follows. The system and attack models are introduced in Sections 5.1 and 5.2, respectively. Our proposed IntVar-based detection and localization methods are derived in Section 5.3 and 5.4, respectively. The effectiveness of the detection method is demonstrated on the 5-bus test system in Section 5.5. Finally, the chapter is concluded in Section 5.6. The results in this chapter are largely based on [71].

5.1 System description

As in Chapter 4, we consider a cyber-physical system, where the physical layer of the system can be described by a set of linear dynamic equations. We also assume that the control algorithm and state estimator, if used, are designed appropriately, so that the closed-loop system is stable and regulated to achieve desired performance objectives. Thus, the stable closed-loop system is described with:

$$\Sigma : \begin{cases} \dot{x}(t) & = \mathcal{A}x(t) + Gd(t) \\ y(t) & = Cx(t) \end{cases} \tag{5.1}$$

where $\mathcal{A}$ is a stable closed-loop system matrix.

5.2 Attack model

Similarly to prior chapters, we consider False Data Injection (FDI) attacks, where an attacker is able to modify the measurement signal $y(t)$ arbitrarily, by injecting a the signal $y_a(t)$:

$$\tilde{y}(t) = y(t) + y_a(t) \tag{5.2}$$

As a consequence, the closed-loop attacked system can be described using:

$$\Sigma_a : \begin{cases} \dot{x}(t) & = \mathcal{A}x(t) + Bu_a(t) + Gd(t) \\ \tilde{y}(t) & = Cx(t) + y_a(t) \end{cases} \tag{5.3}$$

where $\mathcal{A}$ is the closed-loop system matrix.

Attacker's knowledge and resources

In this chapter, we also consider stealthy, i.e. covert, attacks. To launch this kind of attack, the attacker must have *knowledge of the system model*, as well as *disclosure* and *disruption* resources. The depiction of the attack space utilizing these resources can be found in Figure 4.2.

We assume the adversary is attempting an attack on the AGC system, and has the following knowledge and resources:

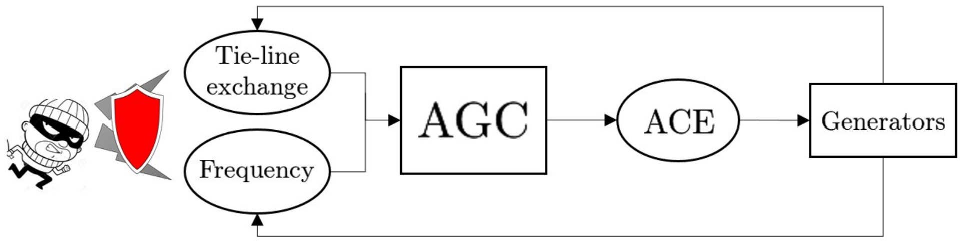


Figure 5.1: Block diagram of attacked AGC system. The attacker manipulates the measurements of the tie-line flows and local frequency in order to compromise the system.

- The attacker has access to the real time sensor measurements utilized in AGC, the tie-line flows and local frequency, and knows the true sensor outputs $y(t)$.
- The attacker has the ability to compromise integrity of the real time sensor measurements utilized in AGC. Specifically, they can replace the true signal $y(t)$ with the arbitrary signal $\tilde{y}(t)$.
- The attacker has knowledge of the system matrices A, B, C, G, and can construct stealthy attacks.

Figure 5.1 depicts the block diagram of the attacked system.

Attack detection problem

As in Chapter 4, we formulate the problem of attack detection for a system defined as above, via a residual test:

$$r = |\tilde{y} - y| > \epsilon \tag{5.4}$$

However, the true measurements y are not known in presence of FDI attacks. In the rest of this chapter, we present a detection filter for attacks on the AGC system. We use the concept of interaction variables, in order to derive an alternative control signal computation based on reliable measurements. If the two methods of control signal computation produce different results, a presence of a cyber-attack in the system can be inferred, and the attack can be localized, i.e. attributed to the affected sensor. Further, the alternative control signal computation can be utilized to continue reliably operating the system in presence of the attack.

5.3 IntVar-based detection filter

The Interaction Variable-based detection filter described in this section was designed to detect cyber-attacks on Automatic Generation Control (AGC). AGC is one of the most critical parts of the operation in today's power systems. Its main function is to automatically control power generation in response to slow, hard-to-predict control area imbalances. Each control area has its own AGC system, with the task of regulating local area frequency to nominal value (60 Hz in USA), and the exchange of power with the neighboring areas to the values agreed upon during economic dispatch. The net power imbalance is represented as the Area Control Error (ACE). ACE is calculated every 2-4 sec in today's operation, and is used to change set-points of generator governors participating in AGC. In order to detect attacks on AGC in power systems, we take advantage of the fact that, in quasi-static operation, the IntVar, much the same as ACE, represents net power imbalance. As such, the two should be equivalent at the rate AGC is implemented.

Following the problem formulation and notation introduced above, for a power system equipped with AGC, the disturbance d will represent the deviation of load from forecast value.

Traditionally, ACE is computed from measurements $y^i = [\omega^i \ F^i]^T$ as a linear combination of frequency deviation from the nominal system frequency (60 Hz in USA) and net tie-line flow deviation from scheduled flows:

$$ACE^k = f(y^k) = \Delta F^k + \frac{\beta^k}{2\pi}\Delta\omega^k \tag{5.5}$$

where $\Delta\omega^k = \omega^k - 60Hz$ and $\Delta F^k = F^k - F^k_{sched}$ represent deviations from scheduled values, and β^k is a frequency bias of area k. In today's AGC, frequency is measured at one location, usually at the location of the Control Center, as it is assumed to be uniform throughout the area.

However, the condition in (5.4) cannot be directly checked to detect cyber-attacks on AGC, as the true measurements y are unknown. Thus, we reformulate the problem as follows. Instead of computing ACE using measurements

y, which might be corrupted by the attacker, we use an alternative set of measurements $\bar{y}$ so that:

$$ACE^k = f(y^k) \equiv g(\bar{y}^k) \tag{5.6}$$

To find this alternative set of measurements $\bar{y}$, we use the definition of Interaction variables given in Chapter 3. More specifically, we extend that definition with the fact that the interaction variable of a lossless system can always be defined as a linear combination of net real power injections of the components in the area [22]:

$$T^i = [0\ \ 0 \ldots t_p^i],\ \ t_p^i \in \mathbb{R}^{1\times n} \tag{5.7}$$

and therefore

$$z^k = t_p^i P^i,\ \forall i \in \mathcal{A}^k \tag{5.8}$$

Thus, we can choose $\bar{y}^k$ to be net power injections at every bus P^i, $i \in \mathcal{A}^k$ and an alternative frequency measurement $\bar{\omega}^k \neq \omega^k$ in area k:

$$\bar{y}^k = \begin{bmatrix} P^i \\ \bar{\omega}^k \end{bmatrix} \in \mathbb{R}^{(n+1)},\ \forall i \in \mathcal{A}^k \tag{5.9}$$

and the alternative computation of ACE:

$$\overline{ACE}^k = g(\bar{y}^k) = z^k - \frac{\beta^k}{2\pi}\bar{\omega}^k \tag{5.10}$$

Finally, we can rewrite the residual r from (5.4) as

$$r = |ACE^k - \overline{ACE}^k| \tag{5.11}$$

and use it to detect cyber-attacks on AGC when $r > \epsilon$.

5.4 Localization of FDI attacks in AGC

Once an attack is detected using the equation (5.11), it would be desirable to know where the attack originated, so that appropriate mitigation procedures can be set in motion. Two possible ways to launch an FDI attack on AGC are either falsifying the frequency measurement at the control center location or the tie-line flow measurements. Therefore, a method is needed to discern between these two possibilities.In normal operation, the collected measurements have to satisfy the power conservation law at every bus. For a lossless system we considered in the rest of this paper, we can write a following equation for each bus j that has a tie-line incident to it:

$$P_j + \mathcal{I}^T P_{jk} + F_j = 0 \tag{5.12}$$

Here we denote with $\mathcal{I}$ the incidence matrix of the area, with P_{jk} power flows to bus j from buses k in the same area, and with F_j the power flow of the tie-line connected to bus j. Since the measurements of P_j and F_j are available to us, and P_{jk}s are functions of them, the equation (5.12) will be violated in case of a FDI attack on tie-line flow measurements. To accommodate for the losses in the system, another threshold can be introduced, and residuals

$$r_j = |P_j + \mathcal{I}^T P_{jk} + F_j| \tag{5.13}$$

can be defined. Then, we can propose a way to discern between the two possible types of attacks, and, in case a tie-line flow measurement is compromised, localize the attack to a specific tie-line flow measurement i:

Consider the control system in (5.3) for which an FDI attack is detected using equation (5.11). Then, the following is true:

- If the residual r_j at bus j satisfies:

$$r_j > \epsilon_1 \tag{5.14}$$

  for some threshold ϵ_1, then the measurement of tie-line flow on the tie-line incident to bus j is under attack.

- If none of the residuals r_j satisfy the above condition, then the measurement of local frequency is under attack.

5.5 Numerical examples on the 5-bus system

In this section, we present the effectiveness of our proposed method on a numerical example on the 5-bus system [42], depicted in Fig 4.3. In this system, buses 1-3 are generator buses, and 4-5 are load buses. The model of the interconnected system was derived as in Section 2. As can be seen in Fig 5.2, the system is divided in two areas, each equipped with its own AGC system, where Area I contains buses 1, 2 and 4, and Area II contains buses 3 and 5.

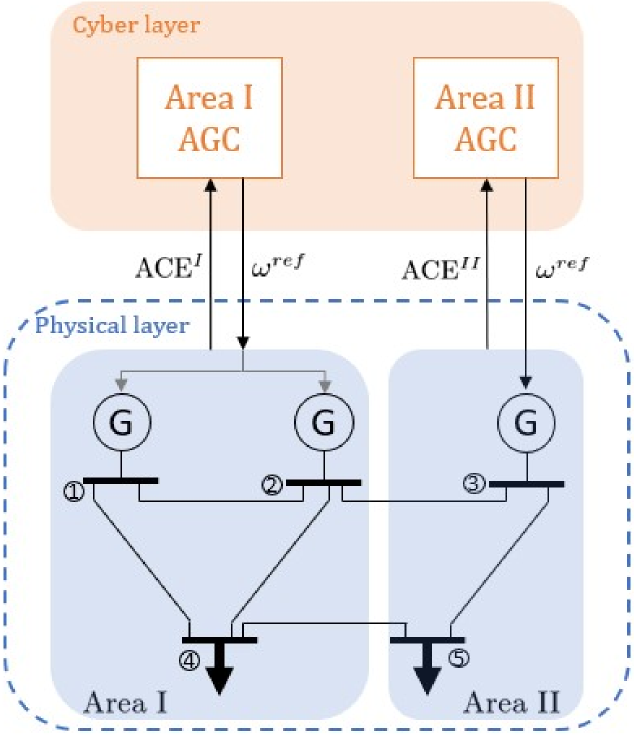


Figure 5.2: Illustration of cyber and physical layers of today's AGC on the 5-bus system

5.5.1 Realistic data generation

In order to generate realistic attack scenarios on AGC, we first simulate the system without any attacks and with real load consumption data used at the load buses. That way, realistic ACE patterns are inserted into the 5-bus test AGC system. We use real time actual load measurements and the load forecast data from two areas in NY-ISO [14]. Since our system model is linear, we use load deviation as inputs to our system, generated by subtracting the forecast values from actual load values. Finally, the load is scaled down to fit the parameters of our small example grid, and injected into the load buses 4 and 5. We simulate the AGC system, driven by these disturbances - deviations in loads, and generate realistic measurements of frequency, tie-line flows, as well as a realistic ACE signal, to which various FDI attacks can be added easily.

5.5.2 Attacked data generation

We demonstrate the efficacy of our proposed method on three different FDI attacks: random attack, ramp attack and scale attack, proposed in [72] and used in [59,73], among others. Random attack aims to add a random positive value $y_a(t) = rand(a, b)$ in the range with $[a, b]$ to $y(t)$ during the attack period T_a:

$$\tilde{y}(t) = \begin{cases} y(t), & \forall t \notin T_a \\ y(t) + rand(a, b), & \forall t \in T_a \end{cases} \tag{5.15}$$

Ramp attack modifies measurements gradually by adding $y_a(t) = \lambda_r * t$ with ramping parameter λ_r in the attack period:

$$\tilde{y}(t) = \begin{cases} y(t), & \forall t \notin T_a \\ y(t) + \lambda_r * t, & \forall t \in T_a \end{cases} \tag{5.16}$$

and scale attack modifies measurements with $y_a(t) = \lambda_s * y(t)$ by scaling up or down with parameter λ_s:

$$\tilde{y}(t) = \begin{cases} y(t), & \forall t \notin T_a \\ y(t) + \lambda_s * y(t), & \forall t \in T_a \end{cases} \tag{5.17}$$

Since no real attacked data is available, attacked data is generated manually by injecting the three types of attacks into tie-line flow measurements, and calculating the ACE signals as in (5.6) based on the attacked measurements.

The FDI attacks are injected to tie-line measurements periodically and every injection lasts for 10 AGC cycles. Three different levels of attacks are generated for each attack type: high-level, medium-level and low-level. High-level attacks refer to large changes on measurements and vice versa. In our generated attack data, on average, a high-level FDI attack changes the ACE value by 2.5%. A medium-level and low-level attack changes the ACE value by 2.0% and 1.5%, respectively.

5.5.3 Performance of the proposed method

In Figure 5.3 we demonstrate the performance of the proposed method in the presence of ramp, scale and random FDI attacks. For the IntVar-based calculation of ACE^I for Area I, the measurements of real power injections on buses 1, 2, and 4 were used, as well as local frequency of generator on bus

2. The detection threshold (red line in Figure X) was chosen based on the parameters of the system, and the same value was used for all three scenarios. For the appropriate choice of threshold, the proposed method of detection is able to detect all attacks with amplitude larger than the chosen threshold. In general, the threshold should be chosen such that physical system dynamics do not trigger false alarms.

5.6 Summary

As introduced in previous chapters, the main mechanism used for detection stealthy cyber-attacks is the use of knowledge of the internal structure of the system that the defender can exploit to design the moving-target detection policy. In this chapter, we make the observation that there is another intuitive way to find the internal structure in power systems that can be used for detection of stealthy attacks. We examine the AGC system, and its aggregation using the Interaction Variables, then develop interaction variable-based detection and localization methods for cyber-attacks on AGC. Finally, we demonstrated the effectiveness of the proposed detection method on the 5-bus test system.

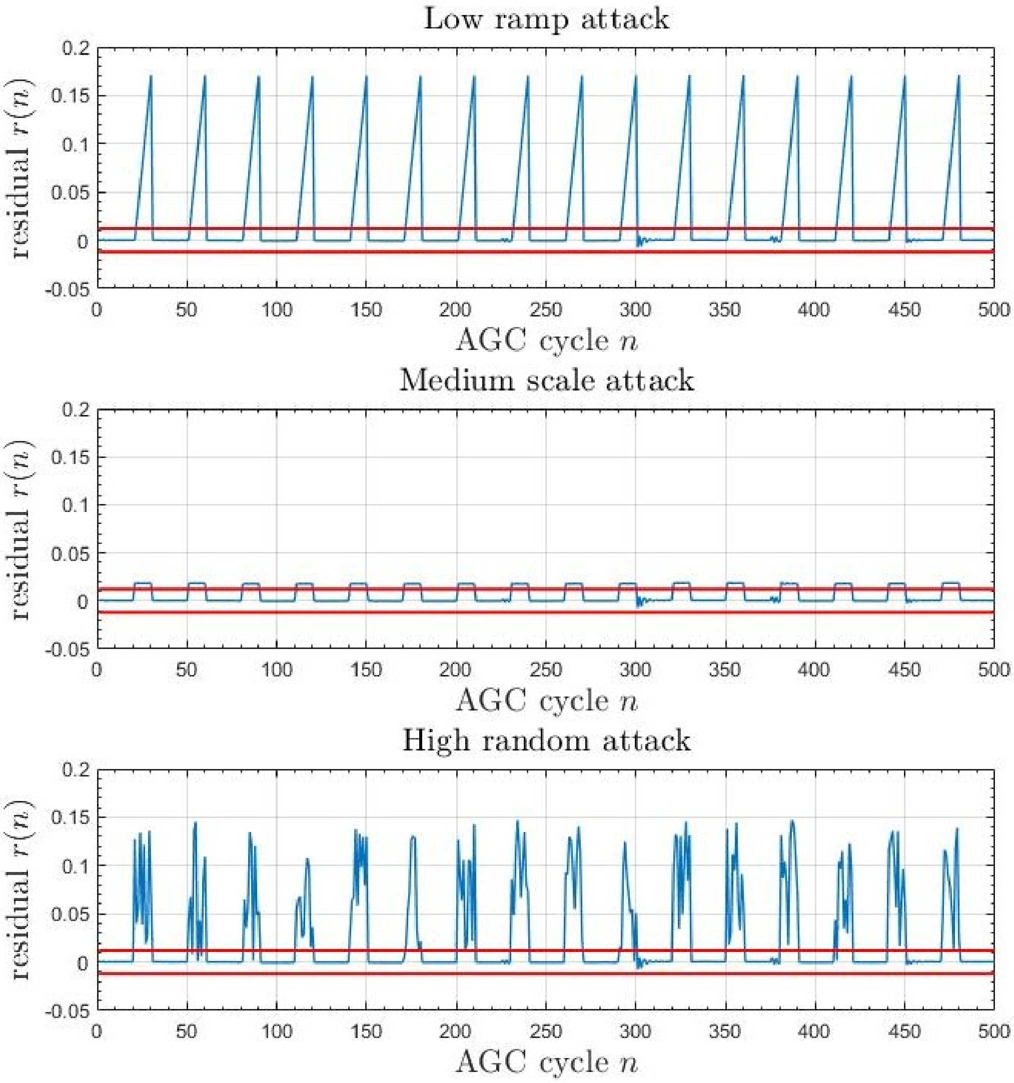


Figure 5.3: Detection results for various attacks: same threshold (red line) used for all three scenarios, all attacks detected. *Top*: low ramp attack, *Middle*: medium scale attack, *Bottom*: high random attack.

Chapter 6

Resilient State Estimation

> *"The term "resilience" means the ability to prepare for and adapt to changing conditions and withstand and recover rapidly from disruptions. Resilience includes the ability to withstand and recover from deliberate attacks, accidents, or naturally occurring threats or incidents."* [76]

Reliable and continued operation of cyber-physical systems requires an accurate state estimate. Today, the reliability of the power grid is largely dependent on employment of redundant components and communication links that make it possible to continue operation during equipment failures and faults that occur naturally [77]. However, such an approach is not adequate in presence of malicious cyber-attackers.

Until recently, IT security tools and lack of connectivity of power control networks to external networks, such as the Internet, were considered sufficient, i.e. "protection through isolation". However, recent successful intrusions via malware, and the increasing connectivity of power grid operational networks to the corporate networks, call for more advanced protection.

The integrity of data, from acquisition to transmission and processing, may be low for several reasons, such as equipment misconfiguration, component or communication link failure, or malicious attacks. In fault detection and identification [78] the objective is to detect if one or more of the components of a system has failed. Traditionally, this is done by comparing the incoming measurements with an expected value of the measurement inferred using an analytical model of the system. This difference signal, also called a residual signal, is then analyzed in order to determine whether a fault has occurred. For example, Bad Data Detector (BDD) is a commonly used filter that detects outliers in power system measurements and excludes them from SE. However, recent research [9] has shown that coordinated FDI attacks may be able to bypass BDD, and negatively impact power system applications, such as economic dispatch (ED) and real-time markets [79,80]. Additionally, in case of any wide-spread attack, naive or intelligent, many measurements may be unavailable or untrusted. In that case, it may not be possible to compute an estimate of the system's state, causing the Energy Management System (EMS) to be suspended.

In the rest of this chapter, we propose a clustering-based method for Resilient SE, that can provide meaningful information on the state of the system in presence of wide-spread coordinated cyber-attacks, leading to improved situational awareness and the ability to mitigate and respond to malicious attacks. While the focus of this work is on wide-spread FDI attacks, our approach is agnostic to the specific form of the attack. We provide the necessary background and state-of-the-art on RSE in Section 6.1. Section 6.2 contains the description of the studied system and mathematical formulation of the problem of Resilient SE. In Section 6.3, we present the Resilient State Estimation method based on output clustering, and demonstrate its efficiency through a numerical example on the IEEE RTS 24-bus power system in Section 6.4. Finally, in Section 6.5 we give some concluding remarks. The results in this chapter are largely based on [81].

6.1 State-of-the-art on resilient state estimation

The problem of state estimation (SE) in presence of cyber-attacks on system measurements has attracted a lot of attention, since a state estimate is crucial to continued operation of the critical infrastructures. In power grids, the inability to produce a state estimate would cause the Energy Management System (EMS) to be suspended, and the operator would have to switch to manual control. The problem of state estimation in an adversarial environment has been studied for many other applications, such as target tracking with compromised sensors, and radar tracking and detection in the presence of jamming attacks.

The impacts of cyber-attacks on power grids have been studied in [50].

Here, the authors leveraged the knowledge of the power system configuration to introduce arbitrary bias to the system without being detected, and studied the effects of the bias on SE. In [79], the impact of malicious attacks on real-time electricity market and the locational marginal price was investigated. Therein, it has been shown that an attacker can make a profit by manipulating the measurements.

There has also been much development in study of optimal attack vectors for SE. In [82], the authors studied the interplay of the power system control center, first developing an optimal attack strategy of the attacker, and then designing a detector for such attack for the defender. A novel attack strategy, the data framing attack, has been proposed in [83]. Here, the authors exploit the BDD by "framing" meters that are providing correct data as sources of bad data in order to exclude them from SE, and ultimately disable it by making the system unobservable.

The resilient SE problem has also been formulated as robust SE, both with noiseless [60,61] and noisy measurements [84–86]. In [87], the authors propose a fusion framework that leverages the intrusion detection from the cyber-layer, to exclude the compromised measurements from SE, thus producing a reliable estimate. However, these methods are fundamentally limited by the observability condition. In other words, in the scenario of a wide-spread coordinated attack, when many measurements may become unavailable, the system will become unobservable, and it will be impossible to produce a state estimate.

6.2 Problem of resilient state estimation

The objective of this chapter is to ensure reliable operation of the power grid, by enabling the Control Center to compute a resilient state estimate in presence of wide-spread coordinated cyber-attacks on system measurements. The structure-preserving dynamic model of the power system, presented in Chapter 2, is used to model the physical system, and is used in its general linear system form to present our proposed approach in the rest of this chapter.

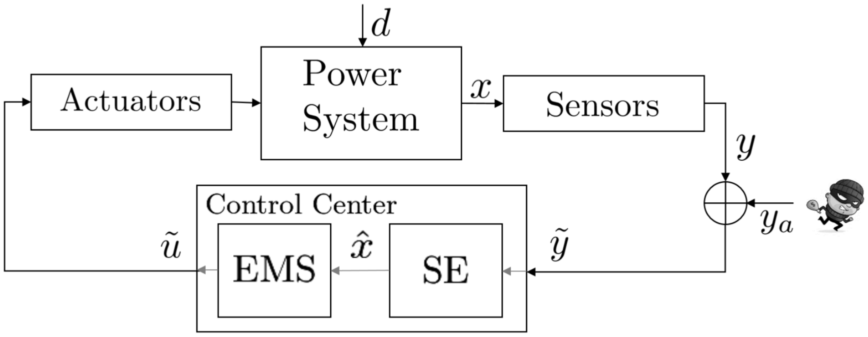


Figure 6.1: Block diagram of the attacked power system. The signal y_a is injected into outputs y to manipulate the system.

6.2.1 Physical system description

Figure 6.1 depicts the block diagram representation of the system we consider. As in previous chapters, we assume the closed-loop dynamics of the physical system are of the form:

$$\dot{x}(t) = \mathcal{A}x(t) + d(t) \tag{6.1}$$

where the states of system are denoted $x \in \mathbb{R}^n$, $d(t)$ denotes a disturbance signal, and the control signal issued by the Control Center (CC) is $\tilde{u} \in \mathbb{R}^p$.

A large network of field sensors is deployed to monitor the operation of the power system in (6.1). A malicious attacker can negatively impact the system by manipulating the measurements, which is represented with the added signal $y_a(t)$ in Figure 6.1:

$$\tilde{y}(t) = y(t) + (e_A^m)^T y_a(t) \tag{6.2}$$

$$y(t) = Cx(t) \tag{6.3}$$

where $\tilde{y} \in \mathbb{R}^m$ are measurements received by the CC, and $y_a \in \mathbb{R}^a$ are a attack signals injected by an attacker. Thus, a potentially manipulated measurement signal $\tilde{y}$ reaches the CC. The CC then processes the received measurements and computes the (potentially incorrect) control signal $\tilde{u}$ to the power system actuators. Finally, the attacked system can be rewritten in closed-loop as:

$$\Sigma_a : \begin{cases} \dot{x}(t) & = \mathcal{A}x(t) + \mathcal{B}y_a(t) + d(t) \\ \tilde{y}(t) & = Cx(t) + (e_A^m)^T y_a(t) \end{cases} \tag{6.4}$$

where $\mathcal{A} = A + BKC$ is the closed-loop system matrix, and $\mathcal{B} = BK$.

6.2.2 State estimation in power systems

State estimation is the core of the on-line analysis functions that constitute EMS. It acts like a filter between the raw measurements received from the system and the application functions that require an accurate and reliable estimate of the current operating state of the system. Besides providing an estimate of the state, SE in power systems typically performs several other important functions, such as topology processing, observability analysis, bad data processing and parameter and structural error processing. In this thesis, however, we will focus only on the problem of computing the current estimate of the system's operating state. Various methods for computing a state estimate have been proposed in the literature. Here, we briefly review Weighted Least Squares (WLS) estimation, commonly used in power system applications, as well as Luenberger observer and Kalman filter, well known and frequently used in general control system applications.

Weighted Least Squares (WLS) estimation

Static state estimation refers to the procedure of obtaining the current state of a system at a given point in time, based on a static relationship between the received measurements y and the system's state x:

$$y = h(x) + e \tag{6.5}$$

where e is the vector of measurement errors. In general, $h(\cdot)$ can be a non-linear function. However, as the problem considered in this thesis is linear, we make the assumption that the function h can be written as $h(x) = Cx$ and

$$y = Cx + e \tag{6.6}$$

State estimation procedure makes use of a set of redundant measurements in order to filter out errors and noise and find an optimal estimate. Measurement error and noise are typically assumed to be zero-mean and independent.

The WLS estimator will minimize the following objective function:

$$\begin{aligned} J(x) &= \sum_{i=1}^{m} (y_i - h_i(x))^2 / Cov(e_i) \\ &= [y - h(x)]^T Cov(e_i)^{-1} [y - h(x)] \end{aligned} \tag{6.7}$$

At the minimum, the optimality conditions of the form $\frac{\partial J(x)}{\partial x} = 0$ will have to be satisfied. WLS State Estimation involves the iterative solution of this equation, starting from an initial guess of the state vector.

Kalman filter

In control theory, the Kalman filter, also known as linear quadratic estimation (LQE), is an algorithm that uses a sequence of measurements observed over time, contaminated by statistical noise and other inaccuracies, to compute an estimate of the system's current state. The algorithm consists of two steps. In the prediction step, the Kalman filter computes an estimate of the current state, along with the embedded uncertainties. In the next time-step, once the next set of measurements is observed, this estimate is updated recursively using a weighted average, with a higher weight assigned to estimates with higher certainty.

The Kalman filter model assumes that the true state x_k at time k evolves according to discrete time dynamics:

$$x_k = F_k x_{k-1} + B_k u_k + w_k \tag{6.8}$$

where F_k is the state transition matrix and w_k is the process noise, assumed to be drawn from a zero mean normal distribution. At time k, a measurement y_k of the true state x_k is made according to:

$$y_k = C_k x_k + v_k \tag{6.9}$$

The Kalman filter is most often conceptualized as two distinct phases: prediction and update phase. The prediction phase uses the state estimate from the previous time-step to compute an estimate of the state at the current time-step. This estimate is also known as the a priori state estimate since it does not include the current observation information. In the update phase, the a priori prediction is refined using the current observation, to produce the a posteriori state estimate. Typically, the two phases alternate.

Luenberger observer

Arguably the most well known method of state estimation in linear control systems is the Luenberger observer. The observer model of the physical system is typically supported by additional terms to ensure that the observer model's state converges to that of the plant. In particular, the output of the observer $\hat{y}$ may be subtracted from the output of the plant y and then multiplied by a gain matrix L. To compute the state estimate $\hat{x}(t)$ of the system in (6.1) from the received measurements, we define the observer:

$$\begin{aligned}\dot{\hat{x}}(t) &= \mathcal{A}\hat{x}(t) + L(\hat{y}(t) - y(t)) \\ &= (\mathcal{A} - LC)\hat{x}(t) + Ly(t) \\ \hat{y}(t) &= C\hat{x}(t)\end{aligned} \tag{6.10}$$

The matrix L is chosen such that the error $e = x - \hat{x}$ between the observer and plant models asymptotically converges to zero. The error dynamics and residual for the system with observer in (6.10) can be written as:

$$\dot{e}(t) = (\mathcal{A} - LC)e(t) \tag{6.11}$$

$$r(t) = y(t) - \hat{y}(t) \tag{6.12}$$

An observer of this form will successfully estimate the state of the system only in the noiseless case. When measurement noise is considered, i.e.:

$$y(t) = Cx(t) + v(t) \tag{6.13}$$

certain assumptions have to be made on the noise, in order to reach a correct state estimate. If the noise is assumed to be white (zero-mean, statistically independent), the observer can be designed so that the residual converges to zero in expectation, i.e. $\mathbb{E}[r(t)] = 0$.

6.2.3 Resilient SE - problem formulation

Note that the three examples of state estimation described above, have the measurement equation in common (equations(6.6),(6.9) and (6.10)). In the rest of this chapter, we propose a resilient SE method that is meant to support the existing state estimation algorithm, and isn't dependent on its particular choice.

Using the observer example, in order to generate a state estimate $\hat{x}$, used for control purposes, it is necessary and sufficient for the pair $(\mathcal{A}, C)$ to be observable. This condition holds during normal operation of power systems, even in presence of sensor failures, due to redundancy in measurements. In other words, when such an equipment failure occurs, the affected measurement is simply removed from state estimation, and the state estimate $\hat{x}$ is computed from the remaining available measurements. A similar procedure can be applied in the event of a cyber-attack. Assuming that attack detection and localization schemes are in place (such as the ones proposed in Chapters 4 and 5, and literature references therein), both in cyber and physical layers, the attacked measurements can be removed from state estimation. This implies that e_A^m is known, and that the measurement matrix C can be decomposed as:

$$C = \begin{bmatrix} C_1 \\ C_A \end{bmatrix} \pi \tag{6.14}$$

where $C_1 \in \mathbb{R}^{(m-a)\times n}$ corresponds to trusted measurements, $C_A \in \mathbb{R}^{a\times n}$ to the attacked measurements, and π is a permutation matrix. Without loss of generality, we assume that measurements are already ordered in this fashion, i.e. $\pi = I$. A state estimate can then be produced if and only if $(\mathcal{A}, C_1)$ is still observable. However, during severe coordinated cyber-attacks, too many measurements may be compromised. Thus, it will not be possible to produce a state estimate.

In this chapter, we address the problem of providing a state estimate in the situation of wide-spread coordinated cyber-attacks. We do so by constructing a matrix $\overline{C}_A$, such that, with an augmented matrix

$$\overline{C} = \begin{bmatrix} C_1 \\ \overline{C}_A \end{bmatrix} \tag{6.15}$$

the pair $(\mathcal{A}, \overline{C})$ is observable, and a state estimate can be computed using the augmented set of measurements

$$\bar{y} = \overline{C}x \tag{6.16}$$

Using this definition, the error dynamics and residual for the system in (6.10) using augmented measurements (6.16) can be written as:

$$\dot{e}(t) = (\mathcal{A} - LC)\hat{x}(t) - (\mathcal{A} - L\overline{C})x(t) - d(t) \tag{6.17}$$

$$r(t) = \hat{y}(t) - \bar{y}(t) \tag{6.18}$$

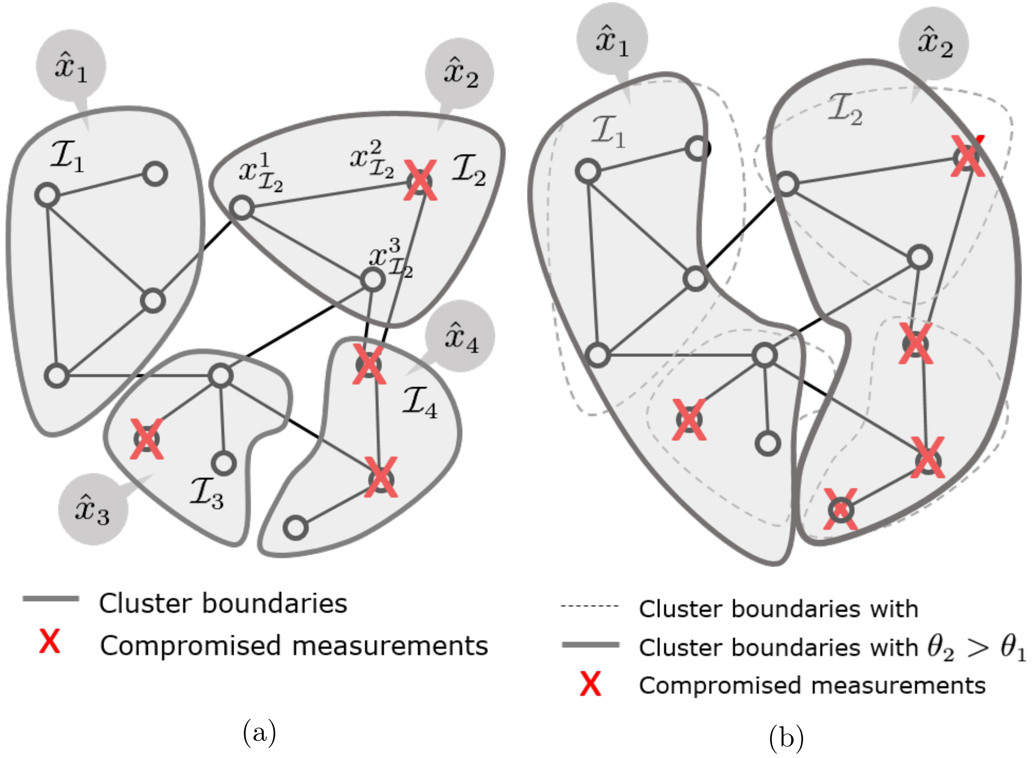


Figure 6.2: **a)** Example of cluster boundaries; $\hat{x}_2$ is used instead of $x^2_{\mathcal{I}_2}$ to produce a state estimate (similarly for other clusters). **b)** Example of cluster boundaries for a larger attack; parameter θ is used to ensure each cluster contains at least one trusted measurement.

The following sections will address the construction of matrix $\overline{C}$ based on clustering of the system outputs. The aim is to first aggregate measurements with similar dynamic responses into clusters, where the aggregate behavior of each cluster is a close approximation of the measurements within it. Then, instead of each of the attacked measurements, the aggregate behavior of the cluster it belongs to can be used as a surrogate during the state estimation process (depicted in Figure 6.2a). As the clustering-based aggregation is an approximation procedure, the resulting state estimate will be less accurate than if all the measurements were available, but it is an important trade-off that must be made in order to retain a necessary level of situational awareness during severe cyber incidents. Therefore, this method is not intended to replace the currently used state estimation methods, but to supplement it during critical events.

6.3 Clustering-based resilient state estimation

In this section, we introduce the clustering procedure on the system Σ, and the construction of matrix $\overline{C}_A$ based on this clustering.

6.3.1 Measurement clustering procedure

Consider the system Σ in normal operation, in absence of cyber-attacks ($y_a(t) \equiv 0$):

$$\Sigma : \begin{cases} \dot{x}(t) & = \mathcal{A}x(t) + d(t) \\ y(t) & = Cx(t) \end{cases} \tag{6.19}$$

To quantify the behavior of measurement signals, and the aforementioned similarity between them, we first define clusters $\mathcal{I}_k$ as disjoint subsets of $\mathbb{L}$, where $\mathbb{L} = \{1, \ldots, l\}$ is the set of measurement indices. More specifically, clusters are subsets of measurements that have a similar trajectories in time domain. Measurements i, j belonging to the cluster $\mathcal{I}_k$ are approximately proportional $a_i y_i(t) \approx a_j y_j(t) \approx \cdots \approx z^{(k)}(t)$, where $a_i, a_j, \ldots$ are constant coefficients. We aim to estimate the full system state based on the combination of received trusted measurements and the hidden system structure contained in the clustered representation of the system. With this intuition in mind, we aim to partition the set $\mathbb{L}$ into clusters $\mathcal{I}_k$ such that

$$p_j g_i(s) = p_i g_j(s), \quad \forall i, j \in \mathcal{I}_k \tag{6.20}$$

where g_i is the i-th element of $g(s) = C(sI_n - \mathcal{A})^{-1}$, a transfer matrix of the system in (6.19). We can rewrite the condition for cluster formation more compactly as

$$(e^n_{\mathcal{I}_k})^T g(s) = p_k^T \bar{g}(s) \tag{6.21}$$

where $\bar{g}(s)$ is a scalar function. This definition provides intuition on the meaning of clustering in our application, but is not practical for designing a procedure that would form such clusters, which would require performing similarity checks on functions. To get around this problem, we will derive a matrix-based condition equivalent to (6.21), based on the notion of observability. To that end, we first derive the observability Gramian of a semistable system (6.19). The observability Gramian is defined as [17]

$$W_o = \int_0^\infty e^{\mathcal{A}^T t} C^T C e^{\mathcal{A} t} dt \tag{6.22}$$

When $\mathcal{A}$ is Hurwitz, the above integral converges, and W_o can also be found as a solution of the Lyapunov equation $\mathcal{A}^T W_o + W_o \mathcal{A} + C^T C = 0$. However, in power systems, the system matrix $\mathcal{A}$ has an inherent structural singularity, as a direct consequence of power conservation law. Due to semistability of the system matrix $\mathcal{A}$, the integral in (6.22) may not converge. Thus, we consider the decomposition of $\mathcal{A} = U \Lambda V^{-1}$, where $U = [u_{max} \quad \bar{U}]$ and $V = [v_{max} \quad \bar{V}]^T$, and u_{max} and v_{max} are the right and left eigenvectors corresponding to the largest eigenvalue ($\lambda_1 = 0$). Let $\bar{\mathcal{A}} = \overline{V}^T \mathcal{A} \overline{U}$ and $\overline{C} = C\overline{U}$, defined as the stable subspace of Σ. Then, the observability Gramian of the semistable system is

$$W_o = \overline{U} \, \overline{W}_o \overline{U}^T \tag{6.23}$$

where $\overline{W}_o$ is the observability Gramian associated with the stable subspace $(\bar{\mathcal{A}}, \overline{C})$ of Σ. In the following theorem we find the condition equivalent to (6.21) using the observability Gramian W_o of the semistable system Σ.

Theorem. *Consider the observability Gramian W_o in* (6.23) *of the semistable system Σ in* (6.19). *Furthermore, let the Cholesky factorization of W_o be given by $W_o = W_L W_L^T$, and $\Phi = W_L$. Then, the condition in* (6.21) *is equivalent to*

$$(e^n_{\mathcal{I}_k})^T \Phi = p_k^T \bar{\phi} \tag{6.24}$$

where $\bar{\phi} \in \mathbb{R}^{1 \times n}$ is a constant vector.

Proof. In order for (6.21) to hold, for each $i, j \in \mathcal{I}_k$ it must hold that $p_j \|g_i(s)\|_{\mathcal{H}_2} = p_i \|g_j(s)\|_{\mathcal{H}_2}$. Similarly, (6.24) is equivalent to $p_j \|\Phi_i\| = p_i \|\Phi_j\|$, where Φ_i is the ith row of the matrix Φ. The $\mathcal{H}_2$-norm of a linear system can be computed as the $\mathcal{L}_2$-norm of its impulse response $h(t)$.

$$\|g(s)\|^2_{\mathcal{H}_2} = \|h(t)\|^2_2 = \operatorname{tr}\left\{ \overline{U} \left[\int_0^\infty e^{\bar{\mathcal{A}}^T t} \overline{C}^T \overline{C} e^{\bar{\mathcal{A}} t} dt \right] \overline{U}^T \right\}$$

For $\|h(t)\|^2_2$ to be finite, the integral above must be finite. Since $\bar{\mathcal{A}}$ and $\overline{C}$ are the stable subspace of Σ, we have $\lim_{t \to \infty} e^{\bar{\mathcal{A}} t} = 0$. Therefore, $\|h(t)\|^2_2$ is finite and equal to:

$$\begin{aligned}\|g(s)\|_{\mathcal{H}_2}^2 = \|h(t)\|_2^2 = \mathrm{tr}\{W_o\} = \mathrm{tr}\{W_L W_L^T\} = \\ = \|W_L\|_F = \|\Phi\|_F\end{aligned}$$

where $\|\cdot\|_F$ is a vector norm applied to each row of Φ. Hence, (6.21) is equivalent to (6.24). □

However, in real systems, the identity in (6.21) is almost never the case. Therefore, we relax the strict equality, and require

$$\|p_j g_i(s) - p_i g_j(s)\|_{\mathcal{H}_2} \leq \varepsilon, \quad \forall i, j \in \mathcal{I}_k \tag{6.25}$$

to hold for each cluster. Equivalently, we can check for linear dependence between rows of matrix Φ:

$$\|p_j \Phi_i - p_i \Phi_j\| \leq \theta \quad \forall i, j \in \mathcal{I}_k \tag{6.26}$$

where $\theta > 0$ and Φ_i is the i-th row of Φ. Here, θ is a parameter that allows us to control the coarseness of clustering. In other words, it allows us to find outputs that have a ”similar”, instead of equal, response, which relaxes the condition (6.21). However, the choice of θ is not trivial, as it introduces a trade-off between accuracy of the approximation and size of clusters. In general, θ should be chosen as a smallest value for which each cluster contains at least one trusted measurement (as depicted in Figure 6.2b).

6.3.2 Construction of matrix $\overline{C}_A$

After the clusters have been defined, we can construct the matrix $\overline{C}_A$ that will be used to augment the set of available trusted measurements so that the system is observable. Then, the system operator can be provided with situational awareness using the resilient state estimate. In the analysis in previous section, we have shown that clusters can be formed such that measurements i, j within the cluster $\mathcal{I}_k$ are approximately proportional, i.e. $a_i y_i(t) \approx a_j y_j(t) \approx \cdots \approx z^{(k)}(t)$. Then, we derived a matrix-based condition to find such clusters. Next, we show that the state estimate can be produced using the augmented matrix $\overline{C}$, by choosing $\overline{C}_A = (e_A^m)^T \Pi^T \Pi C$. The clustering matrix $\Pi \in \mathbb{R}^{K \times n}$ is defined as:

$$\Pi := \mathrm{Diag}\{p_1, p_2, \ldots, p_K\} E \in \mathbb{R}^{K \times n} \tag{6.27}$$

where E is a permutation matrix and p_k are clustering coefficients. The residual $r = \hat{y} - \bar{y}$ defined in (6.18) will converge to 0 if the error system $g_{\hat{y}} - g_{\bar{y}}$ also converges to 0. The transfer matrix associated with $\hat{y}$ is $g_{\hat{y}}(s) = g_y(s) = C(sI - \mathcal{A})^{-1}$, and the transfer matrix associated with $\bar{y}$ is $g_{\bar{y}}(s) = \Pi^T \Pi C \left(sI_n - \mathcal{A}\right)^{-1}$. The following theorem establishes the convergence of the error system.

Theorem. *Consider a semistable linear system in* (6.19) *and the augmented set of measurements $\bar{y}$ in* (6.16)*. Then, the error system $g_e(s) = g_{\hat{y}}(s) - g_{\bar{y}}(s)$ is asymptotically stable, and state x can be estimated using measurements $\bar{y}$.*

Proof. By definition, Π is a unitary matrix. Also, by definition, $v_{max} \in$ colspace(Π^T). Let $\bar{\Pi}$ be an orthogonal complement of Π. Therefore, $I - \Pi^T \Pi = \bar{\Pi}^T \bar{\Pi}$. Consider now the error system g_e:

$$\begin{aligned} g_e(s) &= C(sI - \mathcal{A})^{-1} - \Pi^T \Pi C(sI - \mathcal{A})^{-1} = \\ &= (I_n - \Pi^T \Pi) C(sI - \mathcal{A})^{-1} = \bar{\Pi}^T \bar{\Pi} g(s) \end{aligned} \tag{6.28}$$

We have $\Pi^T \Pi v_{max} = v_{max}$, or equivalently $\bar{\Pi} v_{max} = 0$. This implies that there is pole-zero cancellation in $\bar{\Pi} g(s)$ associated with the zero eigenvalue. Therefore, all poles of $\bar{\Pi} g(s)$ have negative real parts, and the error system g_e is asymptotically stable. □

6.4 Test system and illustrative scenarios

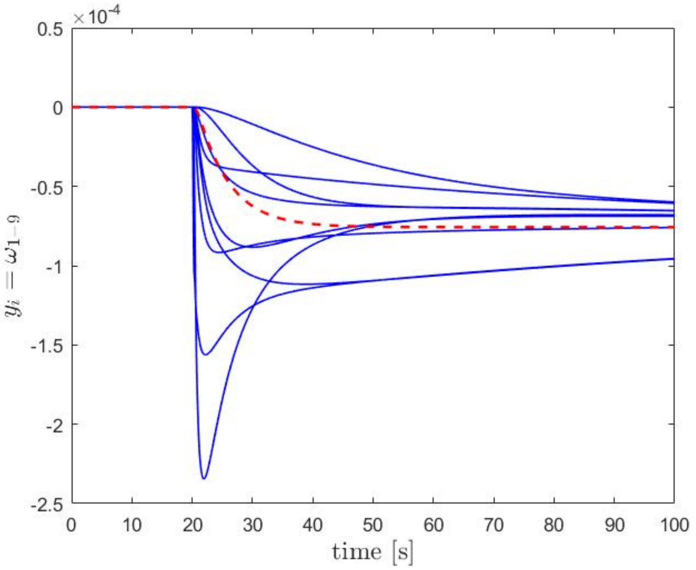
The IEEE RTS 24-bus system [69] consists of 10 generators, equipped with governor control, and 14 loads. The interconnected system model, given in (2.14), where the dimension of x is 68. In Figure 6.3, we analyze the system in the following scenario. From $t = 0$ to 20 s, the loading is nominal. At time $t = 20$ s, load at bus 3 increases by 0.1 p.u., and at time $t = 200$ loading returns to nominal value.

Under this scenario, we plot the trajectories of measurements (solid lines) within two clusters, one containing frequency, and one containing power measurements. In Figures 6.3a and 6.3b, we plot in dotted line the cluster variable, i.e. the surrogate to be used in place of any measurement within the cluster in case of a cyber-attack. In both cases, the cluster variable resembles a centroid of the measurements within it, and can be used for resilient state estimation.

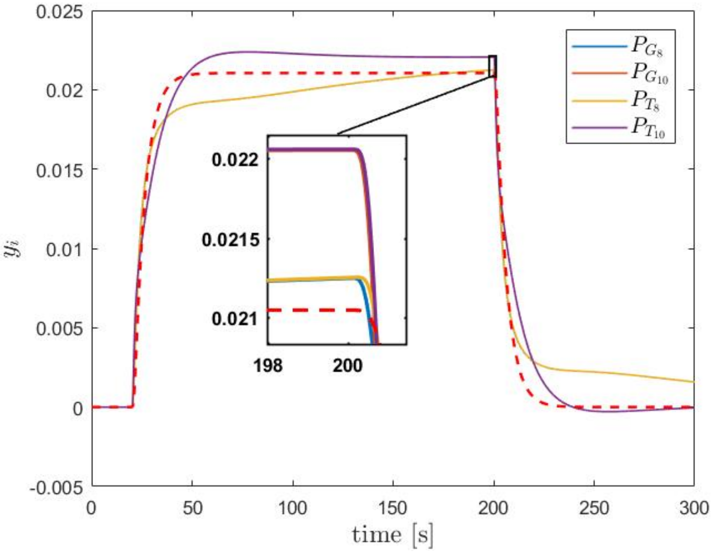
Figures 6.3a and 6.3b present the scenario where measurements were clustered into 21 clusters, which resulted in approximation error of $\sim 7\%$. This choice allows for larger clusters, containing more than one measurement, while maintaining accuracy of approximation and, therefore, suitable for our proposed resilient state estimation method.

6.5 Summary

Reliable and continued operation of cyber-physical systems requires an accurate state estimate. Today, the reliability of the power grid is largely dependent on employment of redundant components and communication links that make it possible to continue operation during equipment failures and faults that occur naturally. However, such an approach is not adequate in presence of malicious cyber-attackers. In this chapter, we propose a clustering-based method for Resilient SE, that can provide meaningful information on the state of the system in presence of wide-spread coordinated cyber-attacks, leading to improved situational awareness and the ability to mitigate and respond to malicious attacks. While the focus of this work is on wide-spread FDI attacks, our approach is agnostic to the specific form of the attack. We demonstrate the efficiency of our proposed algorithms through a numerical example on the IEEE RTS 24-bus power system.



(a) Cluster of frequency measurements



(b) Cluster of power measurements

Figure 6.3: Two examples of clusters in the IEEE RTS 24-bus system. In (a) and (b), solid lines are real measurements, red dotted line is the approximated cluster measurement to be used as surrogate in case of a cyber-attack.

Chapter 7

Summary and Future Work

In this dissertation, we proposed a set of algorithms that enable resilient operation of power systems in presence of intelligent and resourceful attackers. Our methodology is based on the notion of data aggregation as a tool for extracting internal information about the system that may be unknown to the attacker. We argue that the defender must exploit additional degrees of freedom in system design in order to defend against stealthy cyber-attacks. Utilizing the knowledge of the aggregate behavior of different parts of the system, and the ability to manipulate how aggregation is performed, the defender can actively change the detection strategy over time, continuously challenging the attacker. Further, knowledge of the aggregate system behavior can extend situational awareness of the operator during a wide-spread attack, when potentially large number of sensors may be affected. The aggregate system variables can then be used to replace unavailable or untrusted sensor readings. Using the aggregation framework, this thesis focuses on several challenges in power system operation posed by the actions of malicious intruders.

As the first step to resilience and security, we proposed several methods for active attack detection in cyber-physical systems. In the development of these approaches, we were motivated by the fact that passive detectors are not able distinguish between normal and corrupted outputs. Consequently, the defender needs to utilize available degrees of freedom to design resilient systems and controllers, in order to be able to detect otherwise stealthy attacks. We proposed several mechanisms to achieve active detection: a clustering-based and interaction variable (IntVar)-based detection method.

In the first approach, we adopted a moving-target principle, and show that using a constantly changing detection policy allows detection of stealthy attacks. As a basis of our proposed moving-target detection filter, we use the concept of output clustering. Clustering of the outputs gives the defender an upper hand, by providing additional information on the system, unknown to the attacker. The performance of our proposed moving-target approach is tested against stealthy attacks on the 5-bus test system and the IEEE RTS 24-bus system.

In the IntVar-based approach, we made the observation that there is another intuitive way to find the internal structure in power systems that can be used for detection of stealthy attacks. In particular, we examine the AGC system, and its aggregation using the Interaction Variables, then develop interaction variable-based detection and localization methods for cyber-attacks on AGC. We demonstrated the effectiveness of the proposed detection method on the 5-bus test system.

After an attack has been detected, mitigation and self-healing procedures need to be initiated. However, until the intruder has been physically removed from the system, the system operator will have limited knowledge of the system's state and conditions, as many sensors may be unavailable. In that situation, critical processes that provide situational awareness, such as state estimation, need to be enhanced to provide resilience. Reliable and continued operation of cyber-physical systems requires an accurate state estimate. Today, the reliability of the power grid is largely dependent on employment of redundant components and communication links that make it possible to continue operation during equipment failures and faults that occur naturally. However, such an approach is not adequate in presence of malicious cyber-attackers. In this thesis, we proposed a clustering-based method for Resilient SE, that can provide meaningful information on the state of the system in presence of wide-spread coordinated cyber-attacks, leading to improved situational awareness and the ability to mitigate and respond to malicious attacks. While the focus of this work is on wide-spread FDI attacks, our approach is agnostic to the specific form of the attack. We demonstrated the efficiency of our proposed algorithms through a numerical example on the IEEE RTS 24-bus power system.

7.1 Future work

There are several research directions extending from the material presented in this dissertation that could be explored in the future:

Impacts of cyber-attacks on power system stability

The impacts of cyber-attacks on electric power systems have not been well studied on realistic large-scale power systems. Demonstrating that such attacks can cause disruptions in grid operation, and analysis of its impact is needed. This can inform security operators whether a particular attack should be a concern, identify those most harmful forms of cyber-attacks, and provide insight on how to effectively mitigate cyber-attacks.

Detection accuracy improvement through statistical analysis

In this thesis, we consider a deterministic description of the system, in absence of measurement and process noise. However, application of our proposed detection methods on noisy systems may result in an increased rate of false positives. Further, an appropriate choice of the detection threshold may be difficult in presence of severe noise. In the future, any knowledge of the statistical properties of the process or measurement noise can be used to improve the false positive rate. For example, hypothesis testing can be used in conjunction with our proposed methods to distinguish between the noisy signal and the cyber-attack.

Incorporation of data-driven methods

In this thesis, we base the design of our cyber-security methods on the knowledge of the system model. However, certain assumptions have to be made in order to obtain any model. In reality some assumptions may not hold, rendering the system model invalid. Given that, it is difficult, if not impossible, to guarantee performance of any model-based approach in the full range of system conditions. In future work, data-driven methods can be used to enhance the performance of our proposed model-based methods. For example, any unmodeled physical dynamics can be modeled with a statistical model, which could be learned from data. One possible difficulty in adopting this hybrid approach in power systems may be availability of real system data. However, it is a viable option for other cyber-physical systems, such as self-driving cars and robotic systems, where all of the system data is available to the designer.

Appendix A

Model data for the 5-bus system

For the 5-bus system, generator parameters are listed in Table A.1,and the transmission line parameters are listed in Table A.2. System matrices A and G for the system model in (4.5) are given below. For the load at bus 4, actual load and load forecast are shown in Figure A.1, and load deviation around the forecast in Figure A.2. Similarly, for the load at bus 5, actual load and load forecast are shown in Figure A.3, and load deviation around the forecast in Figure A.4.

Gen #	Bus #	J	D	e_T	T_u	K_t	r	T_g
1	1	10	5	1696	.2	10744	19	.25
2	2	8	4	1696	.2	10744	19	.25
3	3	5	4	1696	.2	10744	19	.25

Table A.1: Generator parameters of the 5-bus test system

From	To	
Bus #	Bus #	B
1	2	10
1	4	10
2	4	10
3	5	10
2	3	0.5
4	5	0.5

Table A.2: Transmission line parameters for the 5-bus test system

$$A = \begin{bmatrix} -0.5 & 0 & 0 & 0 & 0 & -0.1 & 0 & 0 & 0 & 0 & 0.1 & 0 & 0 & 5.90e{-}05 & 0 & 0 \\ 0 & -0.5 & 0 & 0 & 0 & 0 & -0.125 & 0 & 0 & 0 & 0 & 0.125 & 0 & 0 & 7.37e{-}05 & 0 \\ 0 & 0 & -0.8 & 0 & 0 & 0 & 0 & -0.2 & 0 & 0 & 0 & 0 & 0.2 & 0 & 0 & 1.18e{-}04 \\ 0 & 0 & 0 & -10 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -10 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 20 & -10 & 0 & -10 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -10 & 20.5 & -0.5 & -10 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -0.5 & 10.5 & 0 & -10 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -10 & -10 & 0 & 20.5 & -0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -10 & -0.5 & 10.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -5 & 0 & 0 & 5.37e{+}04 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -5 & 0 & 0 & 5.37e{+}04 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -5 & 0 & 0 & 5.37e{+}04 \\ -4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -76 & 0 & 0 \\ 0 & -4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -76 & 0 \\ 0 & 0 & -4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -76 \end{bmatrix}$$

$$G = \begin{bmatrix} 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T$$

Appendix B

Model data for the 24-bus system

For the 24-bus system, generator parameters are listed in Table A.3. The transmission line parameters are listed in Table A.4.

Gen #	Bus #	J	D	e_T	T_u	K_t	r	T_g
1	1	10.92	11	1696	0.2	10744	19	0.25
2	2	5.34	11	1696	0.2	10744	19	0.25
3	7	10.92	11	1696	0.2	10744	19	0.25
4	13	12.99	11	1696	0.2	10744	19	0.25
5	15	24.72	11	1696	0.2	10744	19	0.25
6	16	6.61	11	1696	0.2	10744	19	0.25
7	18	5	11	1696	0.2	10744	19	0.25
8	21	47.1	10	1696	0.2	10744	19	0.25
9	22	47.1	10	1696	0.2	10744	19	0.25
10	23	3.71	10	1696	0.2	10744	19	0.25

Table A.3: Generator parameters of the 24-bus test system

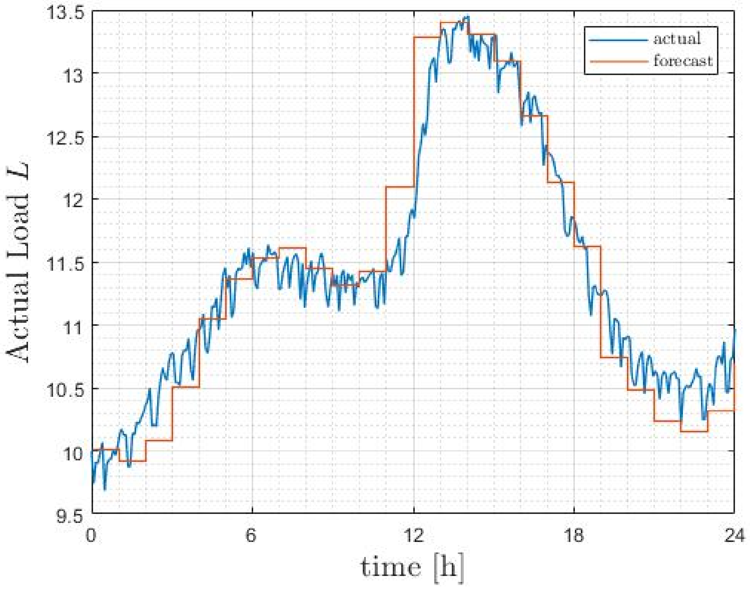


Figure A.1: Forecast vs. actual load at bus 4 over 24h period

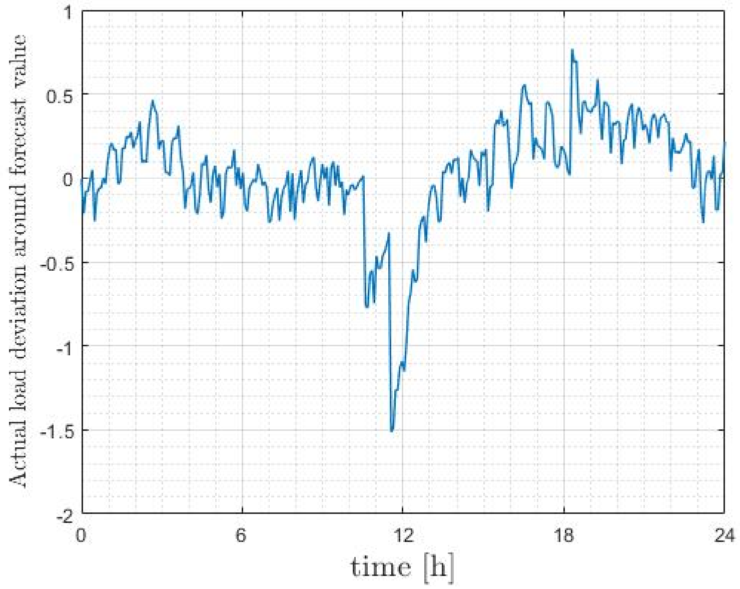


Figure A.2: Deviation of load at bus 4 from forecast value

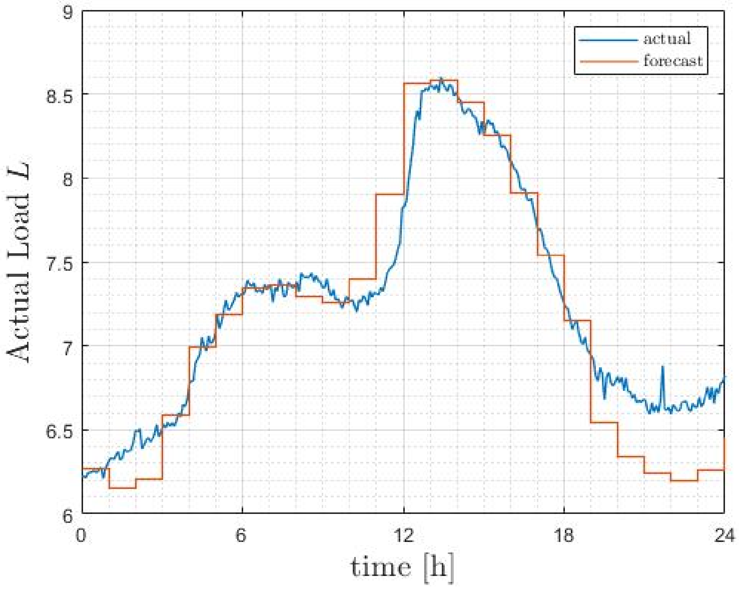


Figure A.3: Forecast vs. actual load at bus 5 over 24h period

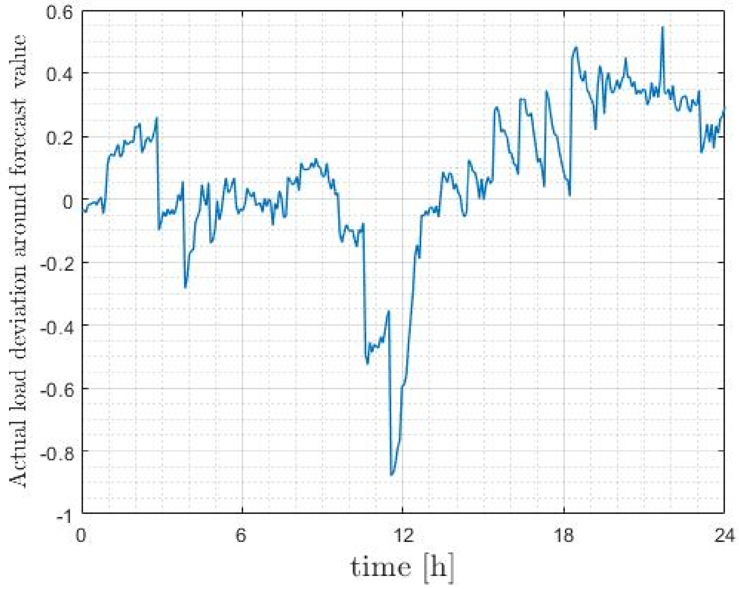


Figure A.4: Deviation of load at bus 5 from forecast value

From Bus #	To Bus #	B
1	2	71.94
1	3	4.73
1	5	11.83
2	4	7.89
2	6	5.21
3	9	8.40
3	24	11.92
4	9	9.64
5	10	11.33
6	10	16.53
7	8	16.29
8	9	6.06
8	10	6.06
9	11	11.92
9	12	11.92
10	11	11.92
10	12	11.92
11	13	21.01
11	14	23.92
12	13	21.01
12	23	10.35
13	23	11.56
14	16	25.71
15	16	57.80
15	21	20.41
15	21	20.41
15	24	19.27
16	17	38.61
16	19	43.29
17	18	69.44
17	22	9.50
18	21	38.61
18	21	38.61
19	20	25.25
19	20	25.25
20	23	46.30
20	23	46.30
21	22	14.75

Table A.4: Transmission line parameters for the 24-bus test system

Bibliography

[1] Alvaro Cardenas, Saurabh Amin, Bruno Sinopoli, Annarita Giani, Adrian Perrig, Shankar Sastry, et al. Challenges for securing cyber physical systems. In *Workshop on future directions in cyber-physical systems security*, volume 5. Citeseer, 2009.

[2] Carol Hawk and Akhlesh Kaushiva. Cybersecurity and the smarter grid. *The Electricity Journal*, 27(8):84–95, 2014.

[3] Richard J Campbell. Cybersecurity issues for the bulk power system, 2015.

[4] Anuradha M Annaswamy, Ahmad R Malekpour, and Stefanos Baros. Emerging research topics in control for smart infrastructures. *Annual Reviews in Control*, 42:259–270, 2016.

[5] Nicolas Falliere, Liam O Murchu, and Eric Chien. W32. stuxnet dossier. *White paper, Symantec Corp., Security Response*, 5(6):29, 2011.

[6] Jill Slay and Michael Miller. Lessons learned from the maroochy water breach. In *International Conference on Critical Infrastructure Protection*, pages 73–82. Springer, 2007.

[7] Defense Use Case. Analysis of the cyber attack on the ukrainian power grid. 2016.

[8] Blake Sobczak. 'Cyber event' disrupted U.S. grid networks - DOE. *E& E News*, April 30, 2019.

[9] Le Xie, Yilin Mo, and Bruno Sinopoli. Integrity data attacks in power market operations. *IEEE Transactions on Smart Grid*, 2(4), 2011.

[10] Jason Andress. *The basics of information security: understanding the fundamentals of InfoSec in theory and practice.* Syngress, 2014.

[11] Yilin Mo and Bruno Sinopoli. Secure control against replay attacks. In *2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 911–918. IEEE, 2009.

[12] André Teixeira, Daniel Pérez, Henrik Sandberg, and Karl Henrik Johansson. Attack models and scenarios for networked control systems. In *Proceedings of the 1st international conference on High Confidence Networked Systems*, pages 55–64. ACM, 2012.

[13] Marija D Ilic. From hierarchical to open access electric power systems. *Proceedings of the IEEE*, 95(5):1060–1084, 2007.

[14] NY-ISO. Real-time actual load data reports, 2018.

[15] Xiaoming Mou, Weixing Li, and Zhimin Li. A preliminary study on the thevenin equivalent impedance for power systems monitoring. In *2011 4th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT)*, pages 730–733. IEEE, 2011.

[16] Jiangmeng Zhang and Daniel Chen. On the application of phasor measurement units to power system stability monitoring and analysis. In *2012 IEEE Power and Energy Conference at Illinois*, pages 1–6. IEEE, 2012.

[17] Chi-Tsong Chen. *Linear System Theory and Design.* Oxford University Press, Inc., New York, NY, USA, 3rd edition, 1998.

[18] Koji Tsumura, Stefanos Baros, Kunihisa Okano, and Anuradha M Annaswamy. Design and stability of optimal frequency control in power networks: A passivity-based approach. In *2018 European Control Conference (ECC)*, pages 2581–2586. IEEE, 2018.

[19] Stefanos Baros and Marija D Ili. On transient stability and voltage regulation through mimo feedback linearizing control of generator and energy storage. In *2015 American Control Conference (ACC)*, pages 4320–4326. IEEE, 2015.

[20] Nasser Jaleeli and Louis S VanSlyck. Nerc's new control performance standards. *IEEE Transactions on Power Systems*, 14(3):1092–1099, 1999.

[21] Howard F Illian. Frequency control performance measurement and requirements. Technical report, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 2010.

[22] Marija D. Ilic and John Zaborszky. *Dynamics and Control of Large Electric Power Systems.* John Wiley& Sons, Inc., 2000.

[23] Vijay Vittal and AR Bergen. Power systems analysis. *Prentice Hall*, pages 1–2, 1999.

[24] Dylan J Shiltz, Stefanos Baros, Miloš Cvetković, and Anuradha M Annaswamy. Integration of automatic generation control and demand response via a dynamic regulation market mechanism. *IEEE Transactions on Control Systems Technology*, 27(2):631–646, 2017.

[25] Gran N Ericsson. Toward a framework for managing information security for an electric power utility—cigré experiences. *IEEE transactions on power delivery*, 22(3):1461–1469, 2007.

[26] Keith Stouffer, Joe Falco, and Karen Scarfone. Guide to industrial control systems (ics) security. *NIST special publication*, 800(82):16–16, 2011.

[27] Peter W Sauer and Mangalore Anantha Pai. *Power system dynamics and stability*, volume 101. Prentice hall Upper Saddle River, NJ, 1998.

[28] Arthur R. Bergen and David J. Hill. Structure preserving model for power system stability analysis. *IEEE Trans. Power Appar. Syst.*, 1981.

[29] Marija D. Ilic, Le Xie, Usman A. Khan, and José M. F. Moura. Modeling of future cyber-physical energy systems for distributed sensing and control. *IEEE Transactions on Systems, Man and Cybernetics: Part A*, 2010.

[30] Stefanos Baros and Marija Ilić. intelligent balancing authorities (ibas) for transient stabilization of large power systems. In *2014 IEEE PES General Meeting— Conference & Exposition*, pages 1–5. IEEE, 2014.

[31] Anuradha Annaswamy and Stefanos Baros. *A Dynamic Framework for Electricity Markets*, pages 129–153. Springer New York, New York, NY, 2018.

[32] Petar V Kokotovic, Robert E O'Malley Jr, and Peddapullaiah Sannuti. Singular perturbations and order reduction in control theory—an overview. *Automatica*, 12(2):123–132, 1976.

[33] Joe H Chow, G Peponides, PV Kokotovic, B Avramovic, and JR Winkelman. *Time-scale modeling of dynamic networks with applications to power systems*, volume 46. Springer, 1982.

[34] S Geeves. A modal-coherency technique for deriving dynamic equivalents. *IEEE transactions on power systems*, 3(1):44–51, 1988.

[35] Athanasios C Antoulas, Danny C Sorensen, and Serkan Gugercin. A survey of model reduction methods for large-scale systems. Technical report, 2000.

[36] Lars Pernebo and Leonard Silverman. Model reduction via balanced state space representations. *IEEE Transactions on Automatic Control*, 27(2):382–387, 1982.

[37] Peter Feldmann and Roland W Freund. Efficient linear circuit analysis by padé approximation via the lanczos process. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 14(5):639–649, 1995.

[38] Petar V Kokotovic, Bozidar Avramovic, Joe H Chow, and James R Winkelman. Coherency based decomposition and aggregation. *Automatica*, 18(1):47–56, 1982.

[39] MA Pai and RP Adgaonkar. Identification of coherent generators using weighted eigenvectors. In *IEEE Transactions on Power Apparatus and Systems*, volume 98, pages 1140–1140, 1979.

[40] Ganesh N Ramaswamy, Luis Rouco, Ollivier Fillatre, George C Verghese, Patrick Panciatici, Bernard C Lesieutre, and David Peltier. Synchronic modal equivalencing (sme) for structure-preserving dynamic equivalents. *IEEE Transactions on Power Systems*, 11(1):19–29, 1996.

[41] Takayuki Ishizaki, Kenji Kashima, Antoine Girard, Jun ichi Imura, Luonan Chen, and Kazuyuki Aihara. Clustered model reduction of positive directed networks. *Automatica*, 59:238 – 247, 2015.

[42] Xiaojun Zhang Liu. *Structural modeling and hierarchical control of large-scale electric power systems.* PhD thesis, MIT, 1994.

[43] Ana Jevtic and Marija Ilic. A dynamic strategy for cyber-attack detection in large-scale power systems via output clustering. *American Control Conference*, 2020.

[44] Richard J. Campbell. Cybersecurity issues for the bulk power system. Technical report, Congressional Research Service Report, 2015.

[45] Yilin Mo, Tiffany Hyun-Jin Kim, Kenneth Brancik, Dona Dickinson, Heejo Lee, Adrian Perrig, and Bruno Sinopoli. Cyber–physical security of a smart grid infrastructure. *Proceedings of the IEEE*, 100(1):195–209, 2012.

[46] Himanshu Khurana, Mark Hadley, Ning Lu, and Deborah A Frincke. Smart-grid security issues. *IEEE Security & Privacy*, 8(1):81–85, 2010.

[47] O. Kosut, Liyan Jia, R. J. Thomas, and Lang Tong. Limiting false data attacks on power system state estimation. In *2010 44th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6, 2010.

[48] Yao Liu, Peng Ning, and Michael K. Reiter. False data injection attacks against state estimation in electric power grids. In *Proceedings of the 16th ACM Conference on Computer and Communications Security*, pages 21–32, New York, NY, USA, 2009. ACM.

[49] Fabio Pasqualetti, Florian Dörfler, and Francesco Bullo. Attack detection and identification in cyber-physical systems. *IEEE transactions on automatic control*, 58(11):2715–2729, 2013.

[50] Yao Liu, Peng Ning, and Michael K Reiter. False data injection attacks against state estimation in electric power grids. *ACM Transactions on Information and System Security (TISSEC)*, 14(1):1–33, 2011.

[51] André Teixeira, Saurabh Amin, Henrik Sandberg, Karl H Johansson, and Shankar S Sastry. Cyber security analysis of state estimators in electric power systems. In *49th IEEE conference on decision and control (CDC)*, pages 5991–5998. IEEE, 2010.

[52] Aranya Chakrabortty, Joe H Chow, and Armando Salazar. A measurement-based framework for dynamic equivalencing of large power systems using wide-area phasor measurements. *IEEE Transactions on Smart Grid*, 2(1):68–81, 2010.

[53] Kwang Y Lee and Mohamed A El-Sharkawi. *Modern heuristic optimization techniques: theory and applications to power systems*, volume 39. John Wiley & Sons, 2008.

[54] AC Zambroni de Souza, JC Stacchini de Souza, and AM Leite da Silva. On-line voltage stability monitoring. *IEEE Transactions on Power Systems*, 15(4):1300–1305, 2000.

[55] Deepa Kundur, Xianyong Feng, Shan Liu, Takis Zourntos, and Karen L Butler-Purry. Towards a framework for cyber attack impact analysis of the electric smart grid. pages 244–249. IEEE, 2010.

[56] Fabio Pasqualetti, Antonio Bicchi, and Francesco Bullo. A graph-theoretical characterization of power network vulnerabilities. In *Proceedings of the 2011 American Control Conference*, pages 3918–3923. IEEE, 2011.

[57] André Teixeira, Henrik Sandberg, and Karl H Johansson. Networked control systems under cyber attacks with applications to power networks. In *Proceedings of the 2010 American Control Conference*, pages 3690–3696. IEEE, 2010.

[58] Yuan Chen, Soummya Kar, and José MF Moura. Dynamic attack detection in cyber-physical systems with side initial state information. *IEEE Transactions on Automatic Control*, 62(9):4618–4624, 2016.

[59] Siddharth Sridhar and Manimaran Govindarasu. Model-based attack detection and mitigation for automatic generation control. *IEEE Transactions on Smart Grid*, 5(2):580–591, 2014.

[60] Y. Shoukry and P. Tabuada. Event-triggered state observers for sparse sensor noise/attacks. *IEEE Transactions on Automatic Control*, pages 2079–2091, 2016.

[61] H. Fawzi, P. Tabuada, and S. Diggavi. Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Transactions on Automatic Control*, pages 1454–1467, 2014.

[62] Tong Huang, Bharadwaj Satchidanandan, PR Kumar, and Le Xie. An online detection framework for cyber attacks on automatic generation control. *IEEE Transactions on Power Systems*, 33(6):6816–6827, 2018.

[63] Bharadwaj Satchidanandan and Panganamala R Kumar. Dynamic watermarking: Active defense of networked cyber–physical systems. *Proceedings of the IEEE*, 105(2):219–240, 2017.

[64] André Teixeira, Iman Shames, Henrik Sandberg, and Karl H Johansson. Revealing stealthy attacks in control systems. In *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1806–1813. IEEE, 2012.

[65] Fei Miao, Quanyan Zhu, Miroslav Pajic, and George J Pappas. Coding sensor outputs for injection attacks detection. In *53rd IEEE Conference on Decision and Control*, pages 5776–5781. IEEE, 2014.

[66] Taouba Rhouma, Jean-Yves Keller, Karim Chabir, Dominique Sauter, and Mohamed Naceur Abdelkrim. Coding control signals and switching lqg controller for secure fault-tolerant control against stealthy false data injection. In *2016 3rd Conference on Control and Fault-Tolerant Systems (SysTol)*, pages 750–755. IEEE, 2016.

[67] Ye Yuan and Yilin Mo. Security in cyber-physical systems: Controller design against known-plaintext attack. In *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 5814–5819. IEEE, 2015.

[68] André Teixeira, Iman Shames, Henrik Sandberg, and Karl Henrik Johansson. A secure control framework for resource-limited adversaries. *Automatica*, 51:135–148, 2015.

[69] Christos Ordoudis, Pierre Pinson, Juan Miguel Morales González, and Marco Zugno. An updated version of the ieee rts 24-bus system for

electricity market and power system operation studies. Technical report, Technical University of Denmark (DTU), 2016.

[70] Rui Tan, Hoang Hai Nguyen, Eddy YS Foo, David KY Yau, Zbigniew Kalbarczyk, Ravishankar K Iyer, and Hoay Beng Gooi. Modeling and mitigating impact of false data injection attacks on automatic generation control. *IEEE Transactions on Information Forensics and Security*, 12(7):1609–1624, 2017.

[71] Ana Jevtic, Fengli Zhang, Qinghua Li, and Marija Ilic. Physics-and learning-based detection and localization of false data injections in automatic generation control. *IFAC CPES Symposium*, 51(28):702–707, 2018.

[72] Yu-Lun Huang, Alvaro A Cárdenas, Saurabh Amin, Zong-Syun Lin, Hsin-Yi Tsai, and Shankar Sastry. Understanding the physical and economic consequences of attacks on control systems. *International Journal of Critical Infrastructure Protection*, 2(3):73–83, 2009.

[73] Siddharth Sridhar and G Manimaran. Data integrity attack and its impacts on voltage control loop in power grid. In *Power and Energy Society General Meeting, 2011 IEEE*, pages 1–6. IEEE, 2011.

[74] Stefanos Baros, Chin-Yao Chang, Gabriel E Colon-Reyes, and Andrey Bernstein. Online data-enabled predictive control. *arXiv preprint arXiv:2003.03866*, 2020.

[75] Stefanos Baros and Marija D Ilić. Distributed torque control of deloaded wind dfigs for wind farm power output regulation. *IEEE Transactions on Power Systems*, 32(6):4590–4599, 2017.

[76] White House. Presidential policy directive–critical infrastructure security and resilience. *Press Release, February*, 12, 2013.

[77] Stefanos Baros, Dylan Shiltz, Prateek Jaipuria, Alefiya Hussain, and Anuradha M Annaswamy. Towards resilient cyber-physical energy systems. 2017.

[78] M-A Massoumnia, George C Verghese, and Alan S Willsky. Failure detection and identification. *IEEE transactions on automatic control*, 34(3):316–321, 1989.

[79] Liyan Jia, Robert J Thomas, and Lang Tong. Impacts of malicious data on real-time price of electricity market operations. In *2012 45th Hawaii International Conference on System Sciences*, pages 1907–1914. IEEE, 2012.

[80] André Teixeira, Henrik Sandberg, György Dán, and Karl H Johansson. Optimal power flow: Closing the loop over corrupted data. In *2012 American Control Conference (ACC)*, pages 3534–3540. IEEE, 2012.

[81] Ana Jevtic and Marija Ilic. Resilient state estimation in presence of severe coordinated cyber-attacks on large-scale power systems. *IEEE Power & Energy Society General Meeting*, 2020.

[82] Oliver Kosut, Liyan Jia, Robert J Thomas, and Lang Tong. Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures. In *2010 first IEEE international conference on smart grid communications*, pages 220–225. IEEE, 2010.

[83] Jinsub Kim, Lang Tong, and Robert J Thomas. Data framing attack on state estimation. *IEEE Journal on selected areas in communications*, 32(7):1460–1470, 2014.

[84] Yasser Shoukry, Alberto Puggelli, Pierluigi Nuzzo, Alberto L Sangiovanni-Vincentelli, Sanjit A Seshia, and Paulo Tabuada. Sound and complete state estimation for linear dynamical systems under sensor attacks using satisfiability modulo theory solving. In *2015 American Control Conference (ACC)*, pages 3818–3823. IEEE, 2015.

[85] Miroslav Pajic, James Weimer, Nicola Bezzo, Paulo Tabuada, Oleg Sokolsky, Insup Lee, and George J Pappas. Robustness of attack-resilient state estimators. In *ICCPS'14: ACM/IEEE 5th International Conference on Cyber-Physical Systems (with CPS Week 2014)*, pages 163–174. IEEE Computer Society, 2014.

[86] Shahrokh Farahmand, Georgios B Giannakis, and Daniele Angelosante. Doubly robust smoothing of dynamical processes via outlier sparsity constraints. *IEEE Transactions on Signal Processing*, 59(10):4529–4543, 2011.

[87] Saman Zonouz, Katherine M Rogers, Robin Berthier, Rakesh B Bobba, William H Sanders, and Thomas J Overbye. Scpse: Security-oriented

cyber-physical state estimation for power grid critical infrastructures. *IEEE Transactions on Smart Grid*, 3(4):1790–1799, 2012.