2014

New Approaches to Smart Grid Security with SCADA Systems

Bixiang Tang
Louisiana State University and Agricultural and Mechanical College

NEW APPROACHES TO SMART GRID SECURITY WITH SCADA SYSTEMS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The School of Electrical Engineering and Computer Sciences

by
Bixiang Tang
B.S., Jiangsu University, 2006
M.S., Jiangsu University, 2009
M.S., Louisiana State University, 2011
August 2014

Dedicated to my parents, professors and friends

Acknowledgements

I would like to express my sincere appreciation to my advisor Dr.Gu Guoxiang for his valuable academic suggestions and patient guidance throughout the research and preparation of this dissertation. His expertise and technical advice deeply influenced me and my work recorded herein. Without his valuable suggestions and constructive direction, this dissertation would not have been completed.

Thanks also goes to my cherished parents who always trust and support me through the years. Without their support, I would not be able to study here and chase my dream. I deeply thank them.

Finally, I would like to thank the faculties, students in Division of Electrical and Computer Engineering, School of Electrical Engineering and Computer Science (EECS) at Louisiana State University for all the help I received.

Table of Contents

List of Figures

Abstract

The use of information technology in electric power grid introduces the vulnerability problem looming the future smart grid. The supervisory control and data acquisition (SCADA) is the first defense, which itself is undermined by potential malicious attacks. This dissertation studies two particular security threats facing the smart grid and SCADA systems: the unobservable attack and the replay attack. The former is well known in fault detection of the power grid and has received renewed interest in the past a few years, while the latter is motivated by the Stuxnet worm allegedly used against the nuclear facilities in Iran. For unobservable attacks, this dissertation adopts the dynamic state estimation approach and treats each bus of the power grid as a dynamic agent. A consensus estimation strategy is proposed to estimate the dynamic states of the power grid, based on which unobservable attacks can be effectively detected. Detection of replay attacks is harder. Two different approaches are proposed in this dissertation. The first is the whitening filter approach that converts the detection of the replay attack into an equivalent white noise detection through whitening a feedback signal. However this approach is less effective, if the replay attack does not change much the whiteness of the filtered feedback signal. Hence a second approach termed as spectrum estimation is proposed. It is shown that the spectrum of the feedback signal in presence of the replay attack can be very different from the case when the replay attack is absent. This approach improves the detection results of the former one. Both are illustrated and examined by the simulation studies.

1 Introduction

Smart grid is a future power grid for which the supervisory control and data acquisition (SCADA) system will play a crucial role in monitoring and controlling the power grid. In fact SCADA systems are already widely used in many different industries, including electric power systems. According to [3], a SCADA system used for power grid provides $100 \sim 200$ voltage/current measurement samples per second, enabling real time monitoring and control for power grid. Because of its effectiveness, the SCADA system is now an essential component of the power grid, and an indispensable part of the future smart grid. SCADA now becomes an enabling technology vital to the smart grid.

Notwithstanding advantages of the SCADA system, it also brings some drawbacks. Due to many sites over large distance and use of information technology, especially wireless communications and networking, SCADA systems give rise to the security problem enticing the vulnerability of the smart grid which becomes more and more serious, endangering the safety of our national infrastructures, including the power grids. This dissertation is focused on this looming security problem for the future safety of the smart grid, and endeavors to provide some viable solution approaches to tackle this difficult problem.

In this chapter we will give a brief introduction of smart grid, emphasizing the advantages and also potential problems facing the smart grid. The security issues in the smart grid will be described subsequently. Based on the background information on smart grid and security issues, we will present our dissertation research from signal/system models to existing methods and to our proposed solutions.

This dissertation is organized as follows. The first chapter will provide an introduction to the smart grid covering both of its advantages and the challenges it faces. Since the security is the theme of this dissertation, we will give more detailed description on the looming security problem facing the smart grid. Our dissertation will be focused on two particular types of malicious attacks against the smart grid with one termed as *unobservable* attacks, and the other called replay attacks. Both have received great attention in the research community [4, 2, 5, 6, 7, 8]. Unfortunately the existing detection methods are not very effective and each has its drawbacks. For this reason, we undertake our research on these two different attacks.

In Chapter 2, we propose a dynamic state estimation approach to detection of the unobservable attacks. The novelty of our proposed approach lies in the consensus estimation by treating each bus as a dynamic agent. Although it requires dynamic modeling of the power grid, the estimation results are far superior to the static state estimation currently employed. More importantly unobservable attacks are not unobservable anymore, and they can be effectively detected demonstrated by both our theoretical results and simulation studies.

The replay attack is a more serous threat to the smart grid. Two chapters of this dissertation are devoted to detection of the replay attack. Chapter 3 is focused on the whitening filter approach, while Chapter 4 is focused on the spectral estimation approach. Replay attacks are motivated by the Stuxnet worm that is allegedly brought down the Iranian nuclear program. However it also opens a door to attacks against industrial infrastructures such as power grid. Mo and Sinopoli [9, 2] are the first to study the replay attack via injecting a Gauss white noise to the control input signal. However their approach is limited to Linear Quadratic Gaussian (LQG) feedback control systems. In addition the injected noise deteriorates the control performance. Although several papers followed to address the performance deterioration issue, injection of the noise to the feedback system remains. We propose to make use of the communication noises that exist in the SCADA system induced by the network channels without injecting noises to the system, and demonstrate the effectiveness of

our proposed two approaches. Our dissertation is concluded in Chapter 4 that highlights the contribution of this dissertation, and outlines the future direction in security research for smart grid.

1.1 Introduction to Smart Grid

In this section, issues in the power grid and advantages of smart grid will be elaborated based on the available literature.

1.1.1 Issues facing the power grid

The power system has been evolved for more than one hundred years. With its development and extension, at present in the United States, the power system has more than 14,000 transmission substations and 4,500 large distribution substations, which make it the most complex electric infrastructure in the world. However, as time pass by, the electric infrastructure is aging, outmoded, underfunded and over stressed today. Right now, we are using the 19th century system from the days of Edison and the 20th century equipments by Westinghouse to keep up with a 21st century economy [10]. The power industry is facing great challenges and compelled to solve many important issues which are discussed next.

- Efficiency Issue: Our power grid has served us for many years. Approximately 70 percent of the transformers and transmission lines are 25 years old, and 60 percent of the circuit breakers are 30 or more years old [11]. The resistance in the transmission lines increases as the power grid is ageing. Consequently, more power is lost in the grid in form of heat. Since the power grid keeps expansion, and has become more complex and larger than before, it becomes increasingly difficult to access the status and information of the power system. To make sure that there is enough power to satisfy all the customers' energy demand, utility companies always have to generate excess power to have adequate margin in meeting the power needs. Therefore, the excessive power is thrown away in the grid. As a result, all the customers have to pay for the waste. In

addition it is harder to optimize the power delivery due to the increased complexity and wide spread of the power grid. Right now, the entire power infrastructure in the U.S. is managed by about 130 control centers. These centers are isolated and operated blindly to each other. So, they cannot effectively cooperate with each other, which results in low efficiency for power delivery [10]. For example, it is possible for some regions to have high power demand, but there are no enough power plants to generate enough power to meet the power demand. While in some other regions, the power companies generate excess power to meet the customers' peak energy demand in that area which is often wasted. As a result, it leads to the lack of power in some regions while wasting power in some other regions. Hence the issue of low efficiency of the power delivery needs to be tackled. Moreover the booming of the renewable energy and its application (of hybrid vehicles) provide new energy sources. There is no reason that the wind power in Texas cannot be balanced and mixed with hydropower in Washington State. However the existing power grids are not well prepared for these new power sources. How to effectively use these new energy sources becomes more urgent than before.

- High Cost Issue: The cost for producing electricity in the daytime is much more expensive than its cost in night. In fact the peak electricity demand can be 100 times more costly to produce [10]. Since power companies cannot control the customers' usage of electricity power, most customers are using power in the daytime with high cost. Thus, it costs a lot to produce the electricity power. Also, to satisfy the maximum power requirement in the peak time, power plant has to keep hot standby, which means that the power plant has to consume coal or fossil fuel to keep generator running at the synchronous speed even when there is no high power demand. If more power plants are built today, then the cost of hot standby will increase significantly. Because of the shortage of fossil fuel and coal, it is urgent to consider new ways to reduce the high cost in producing electricity.

- Reliability Issue: With the power grid extension and ongoing interconnections, the complexity of the power system has increased worldwide. At the same time, there are still many aging transmission lines and devices on duty. As such the power system becomes more difficult to control than before, and thus more susceptible to failures, even the catastrophic failures. For example, in a short span of two months in 2003, there were several blackouts around the world and affected a number of customers [12, 13, 14]; On August 14, 2003, in Northeast United States and Canada, the blackouts affected approximately 50 million people. It took over a day to recover power to New York City and other affected areas. It is considered as one of the worst blackouts in the history of these countries; On August 28, 2003, in London, the blackouts affected commutes during the rush hour and caused an approximately 50-minute loss of power supply to about 20% of the London demand (734MW); On September 23, 2003, in Sweden and Denmark, the blackout affected about 5 million people. The power supply was restored after 5 hours to most of the customers; On September 28, 2003, in Italy, the blackout affected about 57 million people. The power restored to major cities after 5-9 hours. It is considered the worst blackout in Europe. So far as it is reported, the worst blackout was in India happened just on July 31, 2012, spreading to more than half the country. It left more than 600 million people in northern and eastern India out of power. This is considered as the worst on the earth. The reason for the blackout lies in the increased complexity of the power system and unstable aging devices. It increases the difficulty of system-wide coordination of back up protection and also causes more disturbances due to the high sensitivity of these aging devices. Although the blackouts are still the small probability events, they are always associated with huge expenses to power utilities and customers. Thus now is the time for us to pay more attention to the reliability issue. As just described above, power system has been developed for more than a hundred years. But there is no significant change in the past one hundred

years. Our electric grid is still using electro-mechanics from the 1960s and 70s rather than microprocessors in the digital age today. As the power system keeps expansion and grows to more complex, some problems which seem very minor in the past become extremely urgent today. Hence it is necessary to make an evolution to the power grid which can encounter these problems. And it will be a vast, interconnected, intelligent system that is monitored and controlled end to end-all the way down to billions of individual devices. That is the smart grid that is going to be the next generation power grid.

1.1.2 Advantages of smart grid

This subsection introduces the smart grid and answers the basic question such as what smart grid is and what benefits it can offer to us [12, 13, 14, 10, 11].

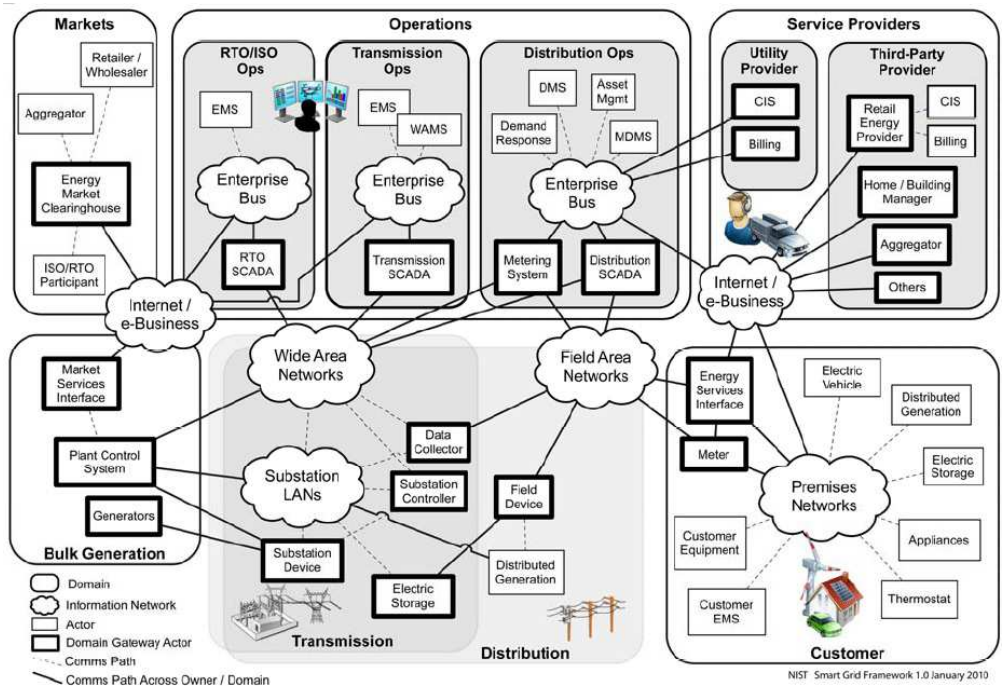

Figure 1.1: Conceptual model for smart grid

Roughly speaking, smart grid is a new generation power grid that uses optimized control algorithm, smart meters, networking and communications infrastructure that are embedded

in residences, companies, factories and throughout the power distribution grid. It is an intelligent and interactive power grid which provides customers and the grid the ability to monitor and regulate energy consumption comprehensively in real time. It is an evolution of the power grid. People will benefit more from its application.

The smart grid offers the following advantages.

- More efficiency and low cost: With the installation of smart meters and the interconnection between customers and utility companies, utility companies can frequently gather data from customers by smart meters. Also, customers can easily and timely get their usage information. It makes the power grid a two-way interactive grid. With the interactivity, power companies are able to estimate the energy demand in the power grid more precisely and regulate the power supply in real time. So, there is no need to generate excessive power any more, which reduces the power waste in the grid and maximizes the power supply efficiency while still guards against power grid overload.

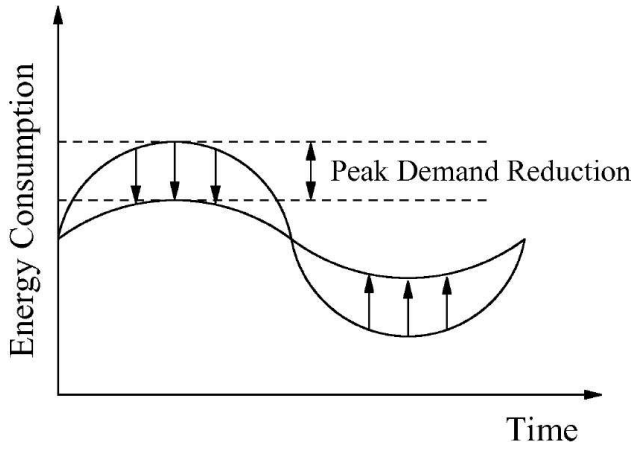

Figure 1.2: Schematic illustration of power usage change

Recall that in the peak loads time excessive energy is very expensive due to the fact that large amount usage of the power is from thermal power plants which use expensive coal and natural gas as the fuel. Also, with the application of smart grid, time-based pricing are going to be feasible. Customers will be provided the fluctuated utility price

which changes according to the fuel consumption in real time. So they can adjust their power usage according to the real time price. Therefore, it will change the customers' power usage habits. As shown in Fig. 1.2, these power usage habit changes will reduce the peak demand and shift part of power usages to the low price time period, which reduces the total cost of power generation and save money for customers.

- More reliable, more green: In the smart grid, with the application of advanced control algorithms, and fast system state feedback, the power grid becomes more reliable. It can handle a wide variety of power sources and storage systems while also maintaining stability. Keep in mind that the renewable energy such as solar and wind are always fluctuating. In the past, it was difficult to coordinate renewable energy with the power grid. But in smart grid, due to its fast regulating ability, the renewable energy can be easily connected to the power grid and coordinated with other distributed power plants to increase the power supply redundancy. Thus it encourages more renewable energy to be used in the power generation, which reduces the greenhouse gas emission and makes power generation more environment friendly. Meanwhile with the increasing use of hybrid vehicles and electric vehicles, smart grid is able to coordinate these vehicles to serve as batteries and backup power for the grid, which makes the power grid more robust in presence of the fluctuate demand impact from customers. With the interactive ability, the control center can coordinate the power usage of customers so as to reduce the adverse impact to the power grid. For example, in factories, many machines contain large numbers of motors. The start current of a motor is $5 \sim 7$ times as normal. If these machines start at the same time, the start current will have a huge impact to the grid. To reduce this kind of impacts, factories have to invest a lot of money on upgrading the infrastructure. But in the smart grid, control center can intelligently control the start sequence of machines, which will reduce the total start current so as to minimize the impact to the grid. As a result it will also save the cost of infrastructure investment.

And with the fast response ability and application of advanced control algorithm, the smart grid can react quickly to any of the power fluctuation in grid and coordinate the power generation and distribution, and protection of customer activities. Thus, the stability margin of the smart grid will be increased, which increases the ability to resist disturbance and reduces the chance of black out in the grid. Thus, the smart grid is more reliable, robust, highly efficient and having low cost.

1.2 Security Issues in Smart Grid

It is important to point out that as a coin has two sides, smart grid also brings some new problems. The most imminent one is the potential security issue emerging in the smart grid. As shown in Fig. 1, with the wide use of smart meters, network and communications infrastructure that are embedded in residences, companies, factories and throughout the power distribution grid, the supervisory control and data acquisition (SCADA) system becomes more complex and interconnected. The connections between smart meters and control networks as well as Internet make the smart grid more vulnerable to penetration. The increased size and complexity of the power grid networks provide plenty potential intrusion nodes. Malicious devices might be inadvertently infiltrated inside the trusted networks by personnel. Devices used both inside and outside of the trusted networks can be infected with malware when used outside, and infiltrate that malware when used inside the trusted networks. Also, the components of the IT infrastructure with susceptibilities or wrong configurations may lead to back doors and holes in the network. Networking devices in the trusted network can be taken control for bypassing proper protocol. Moreover, adversaries can take advantage of the susceptibilities of the devices and setup back doors for future access to the center control network which is the core of the smart grid. There are wide variety motivations for adversaries to launch attacks on the smart grid. Some of attacks are for economic reasons like reducing the utility fee bills. Some of the attacks might be just the pranks. While some of the attacks are for very serious terrorism reasons. Terrorists might attack the power grid

to threaten people by taking control of electricity and other life-critical resources. For example, terrorists can take control of the smart meter to send fake data to the control center so as to mislead the control center to make a wrong action to the power grid. These wrong actions may mess up the electricity delivery and cause huge impact to the power grid, which reduce its stability and reliability or even cause serious consequences. Some of the relays may frequently close and open when they are not required to do that. Generators might be burned during the frequently trip off and synchronous with the power grid. Nuclear plants might be threatened not by the bomb but by the huge oscillation in the grid. Terrorists may shut down the electricity in some areas for their next step of the terrorism action. Or they may make blackout in the power grid, which may influence large number of areas like industries, traffic, communication or hospitals. These attacks will cause serious impacts to people's life and homeland security. Also people's privacy information might be robbed by attackers for future crime purpose. So the reasons for adversaries to attack the smart grid are various, which increase the vulnerability of the smart grid. Thus security of the smart grid is an urgent research problem facing our nation.

There are various kinds of adversary actions in the smart grid today. Once the adversaries gain access to the power control network, they can perform a wide range of attacks. And these attacks will lead to both cyber and physical aftermaths. For the cyber aftermaths, they can be specified as follows [9, 15, 16, 17, 18, 19, 20, 21, 22, 23]:

- Malware spreading and controlling devices: attackers can write some malware programs and spread them to control center and smart meters. Once the device or system is infected by the malware, adversaries can use the malware to change or modify some functions in the device or system such as taking control of the device or sending false information.

- Vulnerabilities in common protocols: in the smart grid, communication between components will be based on existing protocols which inherit the vulnerabilities. Common

protocols may include TCP/IP and remote procedure call (PRC). Attackers can access the control system network by investigating these vulnerabilities in protocols.

- Access through database links: the database on the control system network records the power system activities, which also mirror logs into the business network. The improperly configured modern database architectures will be attacked by skilled attackers. They can access to the database on the business network which gives a path to the control system network.

- Modbus security issues: The SCADA system uses the Modbus protocol as the communication protocol. As the Modbus is a simple client-server protocol, it is originally designed for low-speed serial communications in process control network. Thus, it was not designed for high-security-critical environments. Therefore, several attacks can be launched. These include broadcast message spoofing, baseline response replay, direct slave control, Modbus network scanning, passive reconnaissance and rogue interloper.[9]

- Eavesdropping attacks: adversaries can gather sensitive information by monitoring the network communication between smarts components. They will use the information for future crime activities such as stealing power usage, disclosure of the controlling structure of smart grids.

- Injecting false information on price and meter data: by compromising the smart meters and communication equipment, adversaries can inject false information to the control center, such as false price and meter data. This fake information can result in negative price, power shortage or many other significant damages to the public. In addition, it will produce huge financial impacts on electricity markets [5].

For the physical aftermaths, they are specified as follow [7, 24, 25, 26, 27, 28]:

- Interception of SCADA frames: attackers can intercept SCADA Distributed Network Protocol 3.0 (DNP3) frames and gather unencrypted plain text frames by using a protocol analysis tool for sniffing network traffic. This intercepted data includes control and configuration information. It can be used in future on another SCADA system or intelligent equipment device, so as to shut down or disrupt service. Malware targeting industrial control system: attackers can inject worms into control systems and reconfigure control system settings. A well-known example is the computer worms called Stuxnet in 2010 [29]. It is the first known worms which specifically target at SCADA systems and had been infect thousands of computers worldwide.

- DoS/DDoS attacks on networks and servers: attackers can launch DoS/DDoS attacks against various power grid components such as smart meters, communication infrastructures, and utility business servers. Once they successfully attack these components, the electricity will not be able to control in the target region and power supply might be shut down as the result of the attack.

- Sending fake commands to smart meters in a region: adversaries can send fake command or data to devices in a target region so as to make them do wrong action. They can send disconnect command to some relays in the target region to stop the power delivery in that area. Or they can make invalid switching of electric devices to result in unsafe connections. These may cause the target place burn on fire. Thus, the unsecured communication may threaten people's life.

- Aurora attack: This kind of attacks came out and caught engineers' eye in 2007. The intent of the Aurora attack is to intentionally open a breaker and close it out of synchronism to cause damage to connected power system devices such as generators, motors and transformers. When an out-of-synchronism close is initiated, the resulting high electrical current and torque will translate to the twist stress on the mechanical

shaft of rotating equipment. This kind of twist stress will reduce the life of the rotating equipment and even possibly destroy it.[30]

The above discussion shows the importance of the smart grid security. These security issues need to be resolved before the smart grid can be operated successfully. My research will focus on the cyber-physical security which will be outlined in the next section.

1.3 Dissertation Research

Per our discussion in the previous section, there exist various potential adversary actions in the smart grid, which reveal the vulnerability of the smart grid. Possible attack actions can be separated into two groups and thus pose two major challenges. One is to take control of the target region computer for future crime actions. These attack actions are based on the cyber system, which can be solved by recent cyber security approaches. The other is to modify the data or injecting malicious data into the control system, which can bypass the cyber security. Hence the existing security approaches are inadequate to address the second challenge. It is important to observe that in the smart grid, the core of a monitoring and control system lies in the critical task of state estimation. Plenty of distributed measurements over the power grid are used to keep tracking of the current grid state. The accuracy of the measurement data is essential for maintaining stability and preventing disruption. Because many of the measurement data come from various meters distributed in the power grid and there are always some bad data measurements due to the fault of meters, fault detection techniques are often employed in early research. Power system researchers have developed many techniques for processing the data in order to eliminate the bad data measurement [31]. These techniques first detect if there are bad measurement data, and then identify and remove them.

In early time, researchers considered only false data. However the advent of smart grid has brought in potential attacks that inject malicious data into measurement data. It is

entirely possible for attackers to compromise meters and to introduce false data. Later on it is realized that the existing false data detection techniques can also detect the malicious measurements injected by adversaries. This is because all the existing data detection techniques rely on the same assumption that the square of the difference between observed measurements and their corresponding estimates will significantly increase if there are bad data measurements [32]. Unfortunately this assumption is not always true. It is possible to bypass the existing detection techniques by attackers if they know the configuration of the power grid. Attackers can systematically generate bad data measurements so as to bypass the bad data measurement detection techniques without increasing the square of the error between observed measurements and their corresponding estimates. Such attacks are called *unobservable*. For this reason there is an urgent need to develop new methods for detection of the unobservable malicious data injected by attackers. This is one of main focuses of our dissertation research.

Another devastating and deceiving attack is the replay attack. Stuxnet worm is a prime example. While being used initially to counter the Iran nuclear program allegedly [33] the Stuxnet worm opens the door to malicious attackers. So long as attackers have remote access to sensing and actuation devices, they are able to modify the software or reprogram the devices to launch coordinated attacks against the system infrastructure without being detected virtually by the underlying SCADA system until it is too late. Specifically replay attacks assume that the sensing data are secretely recorded by the attacker using software, which are then replayed back to the SCADA system while conducting the attack on the physical system. The deception created by replay is often seen in movies and spy fictions. It is now possible to deploy the replay attack through Stuxnet worm without being virtually detected by the SCADA – a nightmare to the SCADA. A solution approach as proposed in [2] injects a known independently identically distributed (i.i.d.) zero-mean Gaussian noise into the control input serving as the authentication signal. Assuming LQG control systems,

a χ^2 detector is used to detect the presence of the replay attack. It is shown in [2] that in presence of the replay attack, the normalized error covariance of the innovation signal of the Kalman filter deviates from the identity with a higher variance dependent on the variance of the injected noise or the authentication signal. As its variance increases, the detection rate improves but the control performance suffers. There exists a tradeoff between the detection rate and loss of the control performance in terms of the variance of the authentication signal. A method is proposed in [34] for designing the covariance of the authentication signal to minimize the performance loss while guaranteeing a certain probability of detection rate. The proposed method is demonstrated with an application to security for micro-grids. A different approach is taken in [1] by switching the LQG controller between the one with no added noise and the one with added noise. Results from noncooperative stochastic game are used to minimize the worst-case control and detection cost.

Replay attacks represent another potentially lethal threat to the smart grid. Although the method proposed in [2] and studied in [34, 1] provide effective detection algorithms, injecting noise deteriorates serious by the performance of the smart grid, because the noise variance has to be large enough in order to be effective for detecting the replay attack. So the research for replay attacks is far from over, and is the focus of this dissertation research. In the next two subsections, we provide more detailed description for these two major attacks threatening the future smart grid.

1.3.1 Signal and system models

This subsection presents signal and system models for the unobservable attacks. The replay attack is too complex to be included, which is delayed to Chapter 3.

In the power system, electricity is generated from power plant and transmitted to the load centers by transmission lines, the load centers then distribute the electricity to customers by the distribution system enabled by power lines. Control center monitors and controls the power system working in order. Monitoring the power system status is very important in

keeping the system reliability. To make sure that the power system continues to work well even in presence of faults, power engineers use meters to monitor system components and transmit the meter data to the control center. Fig. 1.3 shows an IEEE 14-bus test system that is one of the commonly used power system models. In this system, each bus and transmission line is equipped with Current Transformers (CTs) and Power/Voltage Transformers (PTs). These CTs and PTs feedback signals to meters so that meters can provide current and voltage signals to the control center. In the control center, the server can gather all the data from these meters and use them to calculate the power system status.

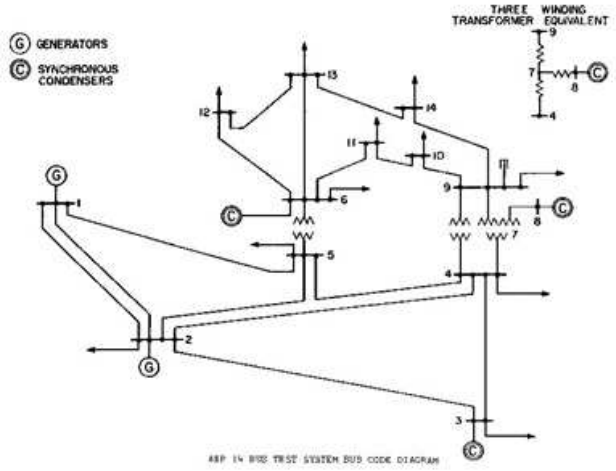

Figure 1.3: IEEE 14-bus test system

Denote $\mathcal{N}(\mu, \Sigma)$ as normal distribution with mean μ and covariance Σ. We assume linearized measurement model of the power grid described by [29]

$$z = Hx + v, \qquad v \sim \mathcal{N}(0, R) \tag{1.1}$$

where z of size m is the power system state that the control center calculated, x of size n is the data vector from meters, and H of dimension $m \times n$ is the measurement matrix. The system state includes the power flow, power load or any states that the control center wants by computing the data from meters. The data vector includes current, voltage, power angle and any other data that meters can directly provide. The system matrix depends on the topology

of the power grid, the parameters of the transmission lines, and the distribution of CTs and PTs. The vector v represents the measurement noise, assumed to be normal distributed with mean zero and diagonal covariance matrix R. Commonly $m > n$ is assumed in order to have redundancy in the measurement so that it has the ability to protect against the bad data which may come from failure or breakdown of meters. The existing estimation techniques rely on three statistical estimation criteria applied to state estimation: the maximum likelihood criterion, the weighted least-square criterion, and the minimum variance criterion [35]. These criteria lead to an identical estimator with the optimal solution given by

$$\hat{x} = (H^T R^{-1} H)^{-1} H^T R^{-1} z, \qquad R = \text{diag}(\sigma_1^2, \sigma_1^2, \cdots, \sigma_m^2). \tag{1.2}$$

Due to various reasons such as meter failures and malicious attacks, bad measurements can be resulted in. Various detection methods have been developed to ensure the success of state estimation [36, 35, 37, 38, 23, 39, 40, 41]. Clearly normal meter measurements always provide acceptable accuracy for estimation of the state variables which are close to their actual values, while abnormal meters may shift the estimated state variables away from the true value. To prevent the estimation from bad measurements, measurement residual $z - H\hat{x}$ is always calculated, and its Euclidean-norm $\|z - H\hat{x}\|$ is used to detect the presence of bad measurements. For instance $\|z - H\hat{x}\|$ can be compared with a threshold τ, and if $\|z - H\hat{x}\| > \tau$, then a bad measurement exists [31]. Many researchers have studied the problem of bad measurements detection and identification in power system [42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52]. However, all of them use the same inequality $\|z - H\hat{x}\| \lesseqgtr \tau$ to detect the existence of bad measurements. It is important to point out that a smart attacker can systematically bypass this detection method.

To attack the system, adversaries often inject wrong data to the power system instead of the true measurements, leading to the signal model

$$z_a = z + a = Hx + v + a \tag{1.3}$$

where z_a is the data collected by control center, a is the vector of the injected attack data added to the original measurements. Denote $\hat{x}_a$ and $\hat{x}$ the estimates of x using the malicious measurement z_a and the original measurement z, respectively. We have $\hat{x}_a = \hat{x} + c$ where c is a nonzero vector of length n and represents estimation error injected by the adversary. It is shown in [31] that if the adversary can find a $c \in \mathbb{R}^n$ such that

$$a = Hc, \qquad z_a = Hx + v + a = H(x + c) + v, \tag{1.4}$$

then the measurement residual of z_a is equal to that of z, and the attack will not change the residuals of the measurement. Indeed there holds

$$\|z_a - H\hat{x}_a\| = \|z + a - H(\hat{x} + c)\| = \|z - H\hat{x}\|. \tag{1.5}$$

As a result, the attack is able to bypass the bad measurement detection. For this reason, such attacks are termed "unobservable". By ignoring the measurement noise, the SCADA will believe that $x + c$ is the true data from meters. It is possible to protect the measurement meters and their signal transmission routes in order to prevent adversaries from changing the value from meters. However, in the real world, due to the high cost and difficulties of management, only a small set of meters can be protected. Therefore it is possible for adversaries to launch attacks from the unprotected devices without altering the measurement or estimation residual. On the other hand, attackers may not know the actually system information, and they cannot find the proper c vector to make $a = Hc$. In this situation, attacks are very similar to bad measurements.

1.3.2 Existing methods

For the observable attacks, many existing detection methods can be used, including generalized likelihood ratio test (GLRT) [4], largest normalized residue test (LNRT) [53], $J(\hat{x})$ detector [51]. The LNRT and $J(\hat{x})$ detectors are two classical bad data detectors. They were often used to test bad measurements in the past when the sample size is small. The GLRT

performs better than the previous two methods when the sample size is large. However, the best and powerful detector dose not exist. There are always pros and cons for any detector. For instance the computation and memory complexities of GLRT may increase due to the high sparsity condition of the attack vector a [4]. There thus exist tradeoffs between the detection performance and complexity including the sparsity level.

For the unobservable attacks, it is much more challenge to detect than the observable attack. Recently the graph-theoretic method is suggested to prevent the unobservable attack [4]. This method is aimed to prevent the smallest unobservable attack by finding the smallest set of meters which causes the system unobservable if they are removed. By protecting these meters, attacker will not be able to attack the grid sneakily. However even if these meters are known to us, it is still difficult to protect all of them due to the high cost and management difficulties. Recall that the power grid is enormously big and complex, while the number of meters being protected is limited [29]. Thus the detection methods are still under investigation and more advanced methods being developed nowadays.

1.3.3 Proposed solutions

Since the static state estimators are widely used in power systems and play a very important role in power grid operation, all the existing detection methods for malicious attacks are based on the static state estimation, by using the following measurement model

$$z = Hx + a + v$$

It leads to the difficulties in detecting unobservable attacks. As the power grid expands, static state estimation becomes increasingly a limiting factor. Facing the increasing generations and loads, the power grid becomes extremely large and complex. It is very challenging for static state estimators to capture the power system dynamic behaviors, which also limits the methods of detecting attacks.

Recently there is a new trend in the research community in studying the dynamic state estimation techniques, enabled by the SCADA system that is capable of measuring and transmitting the power grid information with sampling frequency of $100 \sim 200$ Hz [54]. For this reason the time varying nature and dynamic behaviors can be modeled quite accurately [55]. By using dynamic state estimation, the power system state can also be estimated and predicted more accurately, because the system state estimation is now related not only to the present measurement but also based on the past system states. As a result dynamic state estimation algorithms constitute a significant and integral part of the power system state estimation techniques with a potential to impact the operation of power system control and real time monitoring.

The dynamic state estimation comes with several advantages [56, 57]:

- It helps to identify and reject bad data so as to improve the estimator performance.
- In case of pseudo measurements, it readily provides high quality values and avoid ill conditioning.
- It can also be used for data validation as the states are predicted one more time against the measurement samples.
- In addition it can help identify abnormal changes in the system, the topological errors, and other anomalies.

With these advantages, dynamic state estimation algorithms play more and more important role in modern day system analysis [58, 59]. Today the SCADA systems have evolved from early telemetry systems that used tone-based modulation techniques to transfer analog and digital values at low data rates over telephone lines and radio links [3]. As the development of sensor technology and the increasing processing ability of micro control unit; the data accumulation ability of SCADA systems has been significantly improved with the faster

data measurement rate which is up to 4 samples per cycle [54], i.e., 200 Hz sampling rate. The modern SCADA systems are able to provide real time updates from thousands of remote terminal units (RTUs) which are often spread over large geographical areas, using a range of secure communications media, to multiple 'users' that may also be remotely located [3]. Due to the accuracy and efficiency provided by the SCADA system today, the dynamic state estimation becomes more and more feasible and practical in its application to power grids. For this reason, the system dynamic state information and realization matrices are identifiable. A new approach based on dynamic state information is proposed in this dissertation to detect the unobservable attack, and a new detection algorithm will be developed, where the output residual error will significantly increase if an unobservable attack exists. The effectiveness of this approach will be demonstrated in the next chapter.

However a more serious threat to the future smart grid is the replay attack in which adversaries record a sequence of sensor measurements and replay the sequence afterwards [9]. The well known replay attack is launched by Stuxnet malware. It is a computer worm virus that was discovered in June 2010. It was allegedly designed to attack Siemens Step 7 software which is widely used for program logic controller (PLC) programming. It allegedly attacked the Iran's uranium enrichment plant at Natanz when the plant was undergoing finishing touches in mid to late 2009. The Stuxnet allegedly caused 984 which is one fifth of the centrifuges damaged. As Iran ran isolated valves to allow engineers to take away failing centrifuges from its system without disrupting the enrichment process, Stuxnet shut off some of these valves to increase pressure across the entire set of centrifuges so as to spun them wildly out of control and damage them. Meanwhile, Stuxnet also secretly recorded the normal operations status readings when the plant runs under normal condition, and then played those readings back to the plant operators when the systems was failing. It is like many Hollywood movies that bad guys play the pre-recorded security tape when they rob the bank. Thus it would appear to the operator that everything was running smoothly while the

centrifuges was already damaged. This replay attack prevents a safety system from taking some actions to prevent abnormal operation.

The PLC is used in most of the industry infrastructures, especially in the power grid. The Stuxnet story tells us that the power devices are very vulnerable to the replay attack. Recent investigation and report by Symantec Corporation, a computer security software maker in Silicon Valley, show that they snared the Stuxnet in a global malware collection system. It is thus a fact that the Stuxnet worm already began to pop up around the globe. The worm appeared not only in Iran, but also appeared in India, Indonesia and other countries. No one knows whether their systems have been infected this worm and when it will start attacking their systems. Thus it is a very urgent task to detect and prevent this new type of attacks [60, 33].

In order to detect the replay attack, a simple method has been proposed in [9] for LQG (linear quadratic Gauss) feedback control systems. The basic idea is to inject an additional Gauss random noises to the control signal. Because of the use of the Kalman filter in the LQG control system, the output estimation error is white and has the smallest error variance. The injected random noise serves as a time stamp. Because it is generated by the designer of the control system, it does not deteriorate the output estimation error. However when the replay attack is present, then the injecting noises cannot be canceled, thereby increasing the error variance for the output estimation. Several other papers follow with modifications on the injected noise so to minimize the negative impact to the control system performance. However no matter how to modify the injected noise, it is inevitable to deteriorate the control system performance, and may also sacrifice the detection rate against the replay attack. This problem will be investigated in this dissertation, and we propose two new approaches without injecting Gauss noises at the control input. Instead the existing communication noises induced by communication channels will be utilized to develop our new detection

algorithms. The advantages of our new approaches lie in the fact that they can detect replay attack without sacrificing the control system performance.

2 A Dynamic Estimation Approach

The state estimator is a fundamental component of modern electric power systems. The state of an electric power system operating in steady-state is defined as the vector of voltage magnitudes and phase angles (voltage phasors) at all network buses. In this manner, knowledge of the voltage phasor at each bus and the impedance of each branch (also known as network topology) completely characterizes the state of the power system. The system data exchange (SDX) service of the North American Electric Reliability Corporation (NERC), introduced in 2003, is capable of providing hourly topology information; while SCADA systems are capable of providing local power data in real-time [54].Therefore it is now feasible to estimate the dynamic state of a power system.

Static-state estimators, introduced in [61], use power meter readings from SCADA systems to generate an estimate of the state of the power system when operating in steady-state. However, a power system rarely operates in steady-state since generation units and loads are continually added and/or removed. This is especially true for modern systems where renewable distributed generation sources contribute to unpredictable power fluctuations [62]. Under normal conditions a power system actually operates in a quasi-steady-state where the state changes slowly. This means that it is possible to monitor a power system by taking measurements over short intervals of time to estimate the state. Indeed, the smart grid infrastructure facilitates the collection of these measurements. However, effective algorithms to process this large amount of measurements are still being investigated. Some recent papers that use graph theory as a tool for solving different aspects of the power flow problem are

[63] which investigates the real-time identification of multiple external line outages; and [64] which focuses on determining groups of transmission lines that if removed would cause a severe blackout.

The objective of this chapter is to develop a distributed dynamic state estimator based on the MAS framework (for an overview of MAS see [65]). By employing a distributed estimator, power measurements obtained from the SCADA system can be used to obtain state estimates of the power system when operating in quasi-steady state. More importantly distributed state estimators can be employed to detect the unobservable attacks

2.1 Background Material

This section provides the background material to be used in later sections.

Network Graph and Its Associated Matrices

Consider a power system consisting of N buses denoted by the set of nodes or vertexes $\mathcal{V} = \{v_i\}_{i=1}^{N}$ and T transmission lines described by the set of edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. The network topology of a power system is then specified by the weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. An edge starting at node i and ending at node j is denoted by $(v_i, v_j) \in \mathcal{E}$. The node index set is denoted by $\mathcal{N} = \{1, \cdots, N\}$. The neighborhood of node i is denoted by the set $\mathcal{N}_i = \{j \mid (v_j, v_i) \in \mathcal{E}\}$. A path on the graph is an ordered set of distinct nodes $\{v_{i_1}, \cdots, v_{i_K}\}$ such that $(v_{i_{j-1}}, v_{i_j}) \in \mathcal{E}$. If there is a path in $\mathcal{G}$ from node v_i to node v_j, then v_j is said to be reachable from v_i, denoted as $v_i \rightarrow v_j$. The set of descendants of node v_k is denoted as $\mathcal{S}_k = \{v_j \in \mathcal{V} : \exists \text{ a path } v_k \rightarrow v_j\}$. The digraph is called connected if there exists a node v_k such that $v_j \in \mathcal{S}_k$ for $j = 1, \ldots, N,\ j \neq k$. Such a node v_k is called a *connected node*.

Let $\mathcal{A} = [\ a_{ij}\ ] \in \mathbb{R}^{N \times N}$ be a weighted adjacency matrix. The value of $a_{ij} \geq 0$ represents the coupling strength of edge (v_j, v_i). Denote the degree matrix for $\mathcal{A}$ by $\mathcal{D} = \text{diag}\{\deg_1, \cdots, \deg_N\}$ with $\deg_i = \sum_{j \in \mathcal{N}_i} a_{ij}$ and the Laplacian matrix as $\mathcal{L} = \mathcal{D} - \mathcal{A}$. Let $1_N \in \mathbb{R}^N$ be a vector of 1's. It is clear that $\mathcal{L}1_N = 0$ and thus it has at least one zero eigen-

value. It is also known that $\text{Re}\{\lambda_i(\mathcal{L})\} \geq 0 \; \forall \; i$. In fact the only eigenvalues of the Laplacian matrix on the imaginary axis are zero in light of the Gershgorin circle theorem. In addition, zero is a simple eigenvalue of $\mathcal{L}$, if and only if $\mathcal{G}$ is a connected digraph. We would like to call attention to the fact that similar conditions on the eigenvalues of $\mathcal{L}$ can be obtained through other properties of $\mathcal{G}$. For example, in [66], it is stated that $\mathcal{L}$ has one zero eigenvalue if and only if $\mathcal{G}$ has a spanning tree.

Positive and Bounded Real Lemmas

Consider an internally stable transfer matrix

$$H(z) = C_h(zI - A_h)^{-1}B_h \tag{2.1}$$

where A_h is a Schur stability matrix. Define its $\mathcal{H}_\infty$ norm via

$$\|H\|_\infty = \sup_{|z|>1} \overline{\sigma}[H(z)].$$

The transfer matrix $H(z)$ is said to be bounded real, if $\|H\|_\infty < 1$. The following lemma provides necessary and sufficient conditions for $H(z)$ to be bounded real [67].

Lemma 1. *Consider the internally stable transfer matrix $H(z)$ in* (2.1)*. The following are equivalent:*

(a) $\|H\|_\infty < 1$*, i.e., $H(z)$ is bounded real;*

(b) *There exists a stabilizing solution $X_h \geq 0$ to the algebraic Riccati equation* (ARE)

$$X_h = A_h'X_hA_h + A_h'X_hB_h(I - B_h'X_hB_h)^{-1}B_h'X_hA_h + C_h'C_h \tag{2.2}$$

satisfying $I - B_h'X_hB_h > 0$;

(c) *There exists a stabilizing solution $Y_h \geq 0$ to* ARE

$$Y_h = A_hY_hA_h' + A_hY_hC_h'(I - C_hY_hC_h')^{-1}C_hY_hA_h' + B_hB_h' \tag{2.3}$$

satisfying $I - C_hY_hC_h' > 0$.

It is important to point out that the requirement on the stabilizing solution can be removed. It is a fact [67] that if a solution $X_i \geq 0$ to ARE (2.2) exists satisfying $I - B_h' X_i B_h > 0$ or a solution $Y_i \geq 0$ to ARE (2.3) exists satisfying $I - C_h Y_i C_h' > 0$, then its corresponding stabilizing solution satisfying the corresponding inequality also exists.

Positive real transfer matrices are more complex than bounded real ones. They are square transfer matrices; they allow simple poles on the unit circle with the rest inside the unit circle; they cannot be strictly proper. However for convenience we will consider only stable positive real transfer matrices. For this reason we begin with a square transfer matrix

$$S(z) = D_s + C_s(zI - A_s)^{-1}B_s \tag{2.4}$$

where A_s is a Schur stability matrix and $R_s = D_s + D_s' > 0$. Such an $S(z)$ is said to be strict positive real, if

$$S(z) + S(z)^* > 0 \quad \forall\ |z| \geq 1.$$

The following result provides equivalent conditions for square $S(z)$ to be positive real [68, 69].

Lemma 2. *Consider the internally stable square transfer matrix $S(z)$ in* (2.4)*. The following are equivalent:*

(α) $S(z) + S(z)^* > 0 \quad \forall\ |z| \geq 1.$, *i.e., $S(z)$ is strict positive real;*

(β) *There exists a stabilizing solution $X_s \geq 0$ to* ARE

$$X_s = A' X_s A + (C_s' - A_s' X_s B_s)(R_s - B_s' X_s B_s)^{-1}(C_s - B_s' X_s A_s) \tag{2.5}$$

satisfying $R_s - B_s' X_s B_s > 0$.

(γ) *There exists a stabilizing solution $Y_s \geq 0$ to* ARE

$$Y_s = A_s Y_s A_s' + (B_s - A_s Y_s C_s')(R_s - C_s Y_s C_s')^{-1}(B_s' - C_s Y_s A_s') \tag{2.6}$$

satisfying $R_s - C_s Y_s C_s' > 0$.

However asymptotic stability for $S(z)$ is not required for $S(z)$ to be PR. The next result is quoted from [70].

Corollary 1. *For $S(z)$ in* (2.4) *with minimal realization and all eigenvalues of A_s being simple and on the unit circle, the following statements are equivalent to $S(z)$ being* PR*:*

(i) *There exists $X_s > 0$ such that*

$$X_s - A_s' X_s A_s = 0, \quad C_s' - A_s' X_s B_s = 0,$$

and $D_s + D_s' - B_s' X_s B_s \geq 0$;

(ii) *There exists $Y_s > 0$ such that*

$$Y_s - A_s Y_s A_s' = 0, \quad B_s' - C_s Y_s A_s' = 0,$$

and $D_s + D_s' - C_s Y_s C_s' \geq 0$.

If the inequality in (i) *and* (ii) *are strict, then it is SPR.*

2.2 Modeling and Estimation

Real Power Flow Model

For a power system operating in steady-state, with negligible admittance at every branch, the real power injected at bus i [71] is expressed as

$$P_i = V_i \sum_{j \in \mathcal{N}_i} V_j b_{ij} \sin(\theta_i - \theta_j). \tag{2.7}$$

where b_{ij} is the admittance along line (i, j), V_i is the voltage magnitude of bus i, θ_i is the voltage phase angle of bus i, and $\mathcal{N}_i$ is the set of neighboring buses connected to bus i. For a power system with $N+1$ buses indexed from 0 to N (under the assumption that we know the voltage level at each bus and that the phase angle of one bus is chosen as a reference), the state of the power system will consist of a vector of N phase angles.

According to our previous discussion, modern power systems usually consist of generation units and loads that are continually added and/or removed. Consequently, we assume that the power system is actually operating in a quasi-steady-state, where the voltage phase angles at each bus are changing slowly and randomly. Therefore, we propose to model the phase

angle vector in the following form

$$\theta(k) = \begin{bmatrix} \theta_1(k) & \theta_2(k) & \cdots & \theta_N(k) \end{bmatrix}',$$

assuming that we have renumbered the buses to set bus 0 as the reference. To model the dynamic behavior of the voltage phase angles, we propose to represent $\theta_i(k)$ for $i = 1, \cdots, N$, as the output of an *agent* or dynamic system with the following discrete-time dynamics

$$\begin{aligned} x_i(k+1) &= A_i x_i(k) + B_i v_i(k), \\ \theta_i(k) &= C_i x_i(k), \end{aligned} \tag{2.8}$$

where $x_i \in \mathbb{R}^{n_i}$, $A_i \in \mathbb{R}^{n_i \times n_i}$, $B_i \in \mathbb{R}^{n_i}$, and $C_i' \in \mathbb{R}^{n_i}$, assuming strictly causal model. How to obtain the matrices (A_i, B_i, C_i) is postponed to the next section. The sequence $\{v_i(k)\}$ is a white noise process with $\mathrm{E}\{v_i(k)\} = 0$ and autocovariance sequence (ACS)

$$R_{v_i}(k) := \mathrm{E}\{v_i(t)v_i(t-k)\} = \delta_K(k),$$

where $\mathrm{E}\{\cdot\}$ represents the expectation operator, and $\delta_K(k)$ is the Kronecker delta function defined by

$$\delta_K(k) = \begin{cases} 1, & k = 0, \\ 0, & k \neq 0. \end{cases}$$

The model (2.8) assumes that $\theta_i(k)$ is a wide-sense-stationary (WSS) process which may not be the case for voltage phase angles. However it provides a useful first approximation for the dynamic state estimation problem. More sophisticated models may incorporate non-stationary random processes that will be studied in the future. It is important to note that θ_i is the output of the i^{th} agent; and that in general each agent will have a different state-space model and state vector dimension. In the MAS literature, systems of this type are called *heterogeneous MASs*. If all agents have the same state-space models, then they are referred to as *homogeneous* MASs. We expect that a heterogeneous MAS will provide a better model than a homogeneous MAS for the dynamics of the phase angles.

Another approximation we will make is to assume that the difference $\theta_i(k) - \theta_j(k)$ is small for all $i, j \in \mathcal{N}$. This is usually termed a small-signal model or linear DC power flow model. With this assumption, and under the quasi-steady-state approximation, (2.7) is converted to

$$P_i(k) = V_i \sum_{j \in \mathcal{N}_i} V_j b_{ij} \left[\theta_i(k) - \theta_j(k)\right].$$

We assume that all sensor devices can measure noise corrupted power measurements, but only one sensor is able to measure its own phase. Therefore the measurement equation is expressed as

$$y_i(k) = d_i P_i(k) + g_i \theta_i(k) + \eta_i(k) \tag{2.9}$$

with $\eta_i(k)$ additive white Gauss noise (AWGN) of variance $\sigma^2_{\eta_i}$ and $g_i \neq 0$ for only one i. By rewriting (2.9) as

$$y_i(k) = d_i V_i \sum_{j \in \mathcal{N}_i} V_j b_{ij} \left[\theta_i(k) - \theta_j(k)\right] + g_i \theta_i(k) + \eta_i(k)$$

it is clear that we can only measure the differences between the phase angle at a bus, except for one bus where $g_i \neq 0$. To simplify notation, let us define $a_{ij} = V_i V_j b_{ij}$ so that the above equation can be rewritten as

$$y_i(k) = d_i \sum_{j \in \mathcal{N}_i} a_{ij} \left[\theta_i(k) - \theta_j(k)\right] + g_i \theta_i(k) + \eta_i(k). \tag{2.10}$$

It is clear that $b_{ij} = b_{ji}$, since it represents the admittance across the same branch, thereby $a_{ij} = a_{ji}$ for all (i, j). In the rest of this chapter, $y(k)$ is denoted as a column vector consisting of $\{y_i(k)\}_{i=1}^N$. That is, $y(k)$ represents the collective measurements. Similarly, we denote

$$x(k) = \begin{bmatrix} x_1(k) \\ \vdots \\ x_N(k) \end{bmatrix}, \quad w(k) = \begin{bmatrix} w_1(k) \\ \vdots \\ w_N(k) \end{bmatrix},$$

as the collective state vectors and the collective noise processes with $w_i(k) = \begin{bmatrix} v_i(k) & \eta_i(k)/\sigma_{\eta_i} \end{bmatrix}'$. It is now straightforward to see that the collective measurement equations can be expressed as

$$x(k+1) = Ax(k) + Bw(k), \tag{2.11}$$

$$y(k) = (D\mathcal{L} + G)Cx(k) + Ew(k), \tag{2.12}$$

where $\mathcal{L}$ is the Laplacian matrix, and

$$A = \text{diag}\left(A_1, \ldots, A_N\right), \; B = \text{diag}\left(\tilde{B}_1, \ldots, \tilde{B}_N\right),$$

$$C = \text{diag}\left(C_1, \ldots, C_N\right), \; E = \text{diag}\left(E_1, \ldots, E_N\right),$$

$$D = \text{diag}\left(d_1, \ldots, d_N\right), \; G = \text{diag}\left(g_1, \ldots, g_N\right)$$

with $\tilde{B}_i = \begin{bmatrix} B_i & 0 \end{bmatrix}$ and $E_i = \begin{bmatrix} 0 & \sigma_{\eta_i} \end{bmatrix}$. From (2.12), it is clear that the power network topology induces a graph on the measurement equation.

Estimation

We propose the following modified version of the distributed estimator introduced in [72]

$$\hat{x}_i(k+1) = A_i\hat{x}_i(k) - K_i\left(\hat{y}_i(k) - y_i(k)\right)$$

$$\hat{\theta}_i(k) = C_i\hat{x}_i(k), \tag{2.13}$$

where $\{K_i\}$ are state estimation gains. We will show later that this proposed estimator requires $g_i \neq 0$ for only one i. Physically, this means that we require that only one device be able to measure its own phase angle, but the power measurements $\{P_i\}$ are available for all i.

To design the estimator gain, we shall study the error difference signal between the estimated and actual agent state. Set $e_i(k) = \hat{x}_i(k) - x_i(k)$ as the error signal and

$$e(k) = \begin{bmatrix} e_1(k) & \cdots & e_N(k) \end{bmatrix}'$$

that is the collective error signal. Denote $\mathcal{M} = D\mathcal{L} + G$. After some algebra, we obtain the collective error dynamics

$$e(k+1) = (A - K\mathcal{M}C)\, e(k) - (B - KDE)w(k) \tag{2.14}$$

where $D = \text{diag}\,(d_1, \ldots, d_N)$, $G = \text{diag}\,(g_1, \ldots, g_N)$, and $K = \text{diag}\,(K_1, \ldots, K_N)$. Estimation is aimed at designing the gain matrix K that stabilizes the error dynamics and minimizes the estimation error variance. Because the Laplacian matrix is not diagonal, $\{e_i(k)\}$ are coupled and thus minimization of the estimation error variance is difficult to achieve considering that the grid size N can be prohibitively large. As an alternative we will seek to stabilize the error dynamics and minimize the error dynamics locally and distributively. The following two lemmas are useful.

Lemma 3. *Let $\mathcal{L}$ be the Laplacian associated with $\mathcal{G}$. There exist diagonal matrices $D > 0$ and $G \geq 0$, with rank of G equal to 1, such that*

$$(D\mathcal{L} + G) + (D\mathcal{L} + G)' > 0 \tag{2.15}$$

if and only if $\mathcal{G}$ is connected, and

$$\text{rank}\left\{\begin{bmatrix} \mathcal{L} & e_i \\ -e_i' & 0 \end{bmatrix}\right\} = N + 1.$$

for at least one $i \in \mathcal{N}$.

For a proof of this result, see the proof of Lemma 1 in [73]. The next result is established for the multivariable systems, although its application to dynamic state estimation for smart grid involves only single-input/single-output agents.

Lemma 4. *Let $T_i(z) = J_i + C_i\,(zI - A_i)^{-1}\,K_i$ be a square transfer matrix with no eigenvalues of A_i strictly outside the unit, those on the unit circle being simple, and $J_i > 0$ a positive diagonal matrix. Then there exists an estimation gain K_i such that $(A_i - K_iC_i)$ is a Schur stability matrix and $T_i(z)$ is strict positive real.*

Proof. Without loss of generality we assume that

$$A_i = \mathrm{diag}(A_{i_1}, A_{i2})$$

with A_{i_1} having all simple eigenvalues on the unit circle and A_{i_2} being a Schur stability matrix. Partition C_i and K_i compatibly as

$$C_i = \begin{bmatrix} C_{i_1} & C_{i_2} \end{bmatrix}, \quad K_i = \begin{bmatrix} K_{i_1} \\ K_{i_2} \end{bmatrix}.$$

Hence $T_i(z) = T_{i_1}(z) + T_{i_2}(z)$ with

$$T_{i_j}(z) = J_{i_j} + C_{i_j}\left(zI - A_{i_j}\right)^{-1} K_{i_j}$$

for $j = 1, 2$ and $J_i = J_{i_1} + J_{i_2}$. We will show that $J_{i_1} = J_i$, $J_{i_2} = 0$, and $K_{i_2} = 0$ provide a possible candidate, and we thus need only to design K_{i_1} to render $T_i(z)$ SPR. Since A_{i_1} has only simple eigenvalues on the unit circle, it has the form

$$A_{i_1} = \mathrm{diag}\left(1, -1, \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix}, \cdots\right).$$

For this reason, $T_{i_1}(z)$ being SPR, if and only if

$$A_{i_1} A_{i_1}' = I, \quad K_{i_1} = A_{i_1} C_{i_1}',$$

and $2J_{i_1} - C_{i_1} C_{i_1}' > 0$ in light of Corollary 1. For $T_{i_2}(s)$ to be positive real, it is equivalent to

$$\begin{aligned} \hat{T}_{i_2}(z) &:= (2J_{i_2})^{-1/2} T_{i_2}(z) (2J_{i_2})^{-1/2} \\ &= \frac{1}{2} I + \hat{C}_{i_2}(zI - A_{i_2})^{-1} \hat{K}_{i_2} \end{aligned}$$

being positive real, where $\hat{K}_{i_2} = K_{i_2}(2J_{i_2})^{-1/2}$ and $\hat{C}_{i_2} = (2J_{i_2})^{-1/2} C_{i_2}$. In accordance with Lemma 2, $\hat{T}_{i_2}(z)$ is positive real, if and only if the ARE

$$Z_i = A_{i_2} Z_i A_{i_2}' + (\hat{K}_{i_2} - A_{i_2} Z_{i_2} \hat{C}_{i_2}')(I - \hat{C}_{i_2} Z_{i_2} \hat{C}_{i_2}')^{-1} (\hat{K}_{i_2} - A_{i_2} Z_{i_2} \hat{C}_{i_2}')'$$

has a solution $Z_i \geq 0$ satisfying $I - \hat{C}_{i_2} Z_i \hat{C}'_{i_2} > 0$. Now setting $\hat{K}_{i_2} = 2A_{i_2} Z_{i_2} \hat{C}'_{i_2}$, the above ARE is equivalent to a new ARE

$$Z_i = A_{i_2} Z_i A'_{i_2} + A_{i_2} Z_i \hat{C}'_{i_2} (I - \hat{C}_{i_2} Z_i \hat{C}'_{i_2})^{-1} \hat{C}_{i_2} Z_i A'_{i_2}$$

having a solution $Z_i \geq 0$ satisfying $I - \hat{C}_{i_2} Z_i \hat{C}'_{i_2} > 0$. In light of Lemma 1 and its subsequent discussion, $\hat{G}_i(z) = \hat{C}_{i_2}(zI - A_{i_2})^{-1} B_{i_2}$ has $\mathcal{H}_\infty$ norm $\|\hat{P}_i\|_\infty < 1$, if and only if

$$Y_i = A_{i_2} Y_i A'_{i_2} + B_i B'_i + A_{i_2} Y_i \hat{C}'_{i_2} (I - \hat{C}_{i_2} Y_i \hat{C}'_{i_2})^{-1} \hat{C}_{i_2} Y_i A'_{i_2}$$

has a solution $Y_i \geq 0$ satisfying $I - \hat{C}_{i_2} Y_i \hat{C}'_{i_2} > 0$. Because the above is equivalent to ARE for Z_i by taking $B_i = 0$, $\|\hat{G}_i\|_\infty < 1$ always holds when $B_i = 0$. In fact $Z_i = 0$ is the stabilizing solution by the Schur stability of A_{i_2}, leading to $\hat{K}_{i_2} = 0$. This concludes the proof. □

With these results in place, we now provide a method to construct the estimator gain K in the following theorem.

Theorem 1. *There exists a stabilizing estimation gain K to* (2.14)*, if the underlying smart grid topology, $\mathcal{G}$, is connected.*

Proof. Assume $\mathcal{G}$ is connected. Let $\mathcal{M} = D\mathcal{L} + G$ with D and G as in (2.14). Since $g_i \neq 0$ for only one i, G is a rank one diagonal matrix and by Lemma 4, $\mathcal{M} + \mathcal{M}' > 0$. Note that the error dynamics (2.14) are stable if

$$\det(zI - A + K\mathcal{M}C) \neq 0 \quad \forall\, |z| \geq 1, \tag{2.16}$$

i.e., $A - K\mathcal{M}C$ has no eigenvalues outside and on the unit circle. Since $\mathcal{M} + \mathcal{M}' > 0$, there exists $J = \mathrm{diag}(J_1, \ldots, J_N)$ with $J_i > 0$ diagonal and invertible such that

$$\mathcal{M} + \mathcal{M}' - \mathcal{M}(J + J')\mathcal{M}' > 0. \tag{2.17}$$

Note that (2.17) is a linear matrix inequality (LMI). So J can easily be computed with an LMI solver. In addition the existence of such a J is clear: by taking $J \to 0$, the above strict

inequality holds. Hence sufficiently small J satisfies the strict inequality (2.17). It is now noted that (2.17) is equivalent to

$$(I - \mathcal{M}J)\mathcal{M}' + \mathcal{M}(I - J'\mathcal{M}') > 0.$$

Multiplying the above by $(I - J'\mathcal{M}')^{-1}$ from right and $(I - \mathcal{M}J)$ from left yields

$$\mathcal{M}'(I - J'\mathcal{M}')^{-1} + (I - \mathcal{M}J)^{-1}\mathcal{M} > 0.$$

Since J can be taken sufficiently small, $(I - \mathcal{M}J)^{-1}$ exists. Define

$$\Pi = (I - \mathcal{M}J)^{-1}\mathcal{M} \iff \mathcal{M} = \Pi(I + J\Pi)^{-1}$$

where "$\Longleftrightarrow$" stands for the equivalence relation. Then inequality (2.17) holds if and only if

$$\Pi + \Pi' > 0.$$

Upon substitution of $\mathcal{M} = \Pi(I + J\Pi)^{-1}$, the stability condition (2.16) is equivalent to

$$\lambda(z) := \det(zI - A + K\Pi(I + J\Pi)^{-1}C) \neq 0 \quad \forall\, |z| \geq 1. \tag{2.18}$$

Denote $T_i(z) = J_i + C_i(zI - A_i)^{-1}K_i$ and

$$T(z) = \mathrm{diag}(T_1(z), \ldots, T_N(z)) = J + C(zI - A)^{-1}K.$$

In light of Lemma 3, K_i exists for each i such that $T_i(z)$ is positive real. Combined with $\Pi + \Pi' > 0$, it implies that [73]

$$\det[I + \Pi T(z)] \neq 0 \quad \forall\, |z| \geq 1. \tag{2.19}$$

By the root locus argument, the above is equivalent to

$$\det[T(z)^{-1} + \Pi] \neq 0 \quad \forall\, |z| \geq 1.$$

Substitution of the realization of $T(z)$ yields

$$\det[T(z)^{-1} + \Pi] = \det[J^{-1} + \Pi - J^{-1}C(zI - A + KJ^{-1}C)^{-1}KJ^{-1}].$$

Hence the inequality (2.19) is in turn equivalent to

$$\det[I + J\Pi - C(zI - A + KJ^{-1}C)^{-1}KJ^{-1}] \neq 0$$

$\forall |z| \geq 1$. That is, the inequality (2.19) holds if and only if

$$\det[I - (I + J\Pi)^{-1}C(zI - A + KJ^{-1}C)^{-1}KJ^{-1}] \neq 0$$

for all $|z| \geq 1$. On the other hand the stability condition (2.18) is equivalent to

$$\begin{aligned}
\lambda(z) &= \det[zI - A + KJ^{-1}J\Pi(I + J\Pi)^{-1}C] \\
&= \det[zI - A + KJ^{-1}(I - (I + J\Pi)^{-1})C] \\
&= \det[zI - A + KJ^{-1}C - KJ^{-1}(I + J\Pi)^{-1})C] \\
&\neq 0 \quad \forall\ |z| \geq 1.
\end{aligned}$$

Again by the root locus argument, the above is equivalent to

$$\begin{aligned}
\tilde{\lambda}(z) &:= \det[I - KJ^{-1}(I + J\Pi)^{-1}C(zI - A + KJ^{-1}C)^{-1}] \\
&= \det[I - (I + J\Pi)^{-1}C(zI - A + KJ^{-1}C)^{-1}KJ^{-1}] \\
&\neq 0 \quad \forall\ |z| \geq 1
\end{aligned}$$

where we used the fact that $\det(I + AB) = \det(I + BA)$. The above is identical to inequality (2.19), and thus there indeed exist $\{K_i\}_{i=1}^{N}$ such that (2.16) is true. □

Under the stability assumption for $(A - K\mathcal{M}C)$, the steady-state estimation error covariance $\Sigma = \mathrm{E}\{e(k)e(k)'\}$ for (2.14) satisfies the Lyapunov equation

$$\Sigma = (A - K\mathcal{M}C)\Sigma(A - K\mathcal{M}C)' + (KDE - B)(KDE - B)'. \tag{2.20}$$

In addition the steady-state estimation error covariance for θ can be obtained as

$$\Sigma_\theta = C\Sigma C' + \Sigma_\eta \tag{2.21}$$

with Σ_η as the covariance of the measurement noise vector.

Remark 1. It is important to note that our proposed estimator is distributed, in the sense that each individual estimation gain K_i is designed based on the solution of an ARE of order n_i. This is certainly desirable since the collective state vector may be of a very large dimension, making controller design computationally prohibitive. Also, due to Lemma 4, we require only one g_i to be nonzero. Recall that we require diagonal $G \geq 0$ with rank of G equal to 1. Loosely speaking, we require only one agent estimator to be placed on a connected bus of the power system.

Determination of Phase Angle Dynamics

We now consider identification of the realization matrices (A_i, B_i, C_i) based on phase angle measurement data. Recall that these matrices model the behavior of the bus voltage phase angles when the power system is operating in a quasi-steady state. According to the dynamical equation (2.8), $\theta_i(k)$ is a sequence of continuous random variables with a power spectral density (PSD) $\Phi_{\theta_i}(\omega)$. The PSD is defined as the discrete time Fourier transform (DTFT) of the ACS $R_{\theta_i}(k)$,

$$\Phi_{\theta_i}(\omega) = \sum_{k=-\infty}^{\infty} R_{\theta_i}(k) e^{-j\omega k} \tag{2.22}$$

and admits the spectral factorization in form of

$$\Phi_{\theta_i}(\omega) = \Theta_i(e^{j\omega})\Theta_i(e^{j\omega})^* \tag{2.23}$$

where $\Theta_i(z)$ is stable. If $\Phi_{\theta_i}(\omega) > 0$ for all $\omega \in \mathbb{R}$, then $\Theta_i(z)$ also has all its zeros strictly inside the unit circle. Computation of approximate $\Theta_i(z)$ based on $\{\widehat{\theta}_i(k)\}$, measurements of $\{\theta_i(k)\}$, has been well studied in the literature. We refer readers to the book [74]. It is emphasized that many algorithms can be carried out to estimating an approximate PSD of $\theta_i(k)$. A simple way is based on

$$\Phi_{\theta_i}^{(n)}(\omega) = \mathrm{E}\left\{ \frac{1}{n} \left(\sum_{k=0}^{n-1} \widehat{\theta}_i(k) e^{-j\omega k} \right) \left(\sum_{\tau=0}^{n-1} \widehat{\theta}_i(\tau) e^{-j\omega \tau} \right)^* \right\}.$$

In practice expectation can be replaced by average of the measured data over more than one time horizon of interval length n. Since the above can be computed via the use of the fast Fourier transform (FFT) algorithm, its numerical efficiency and reliability are well-known. Various estimation algorithms from [75] can be used to identify $\Theta_i(z)$. Once $\Theta_i(z)$ is obtained for $i \in \mathcal{N}$, its realization matrices $\{A_i, B_i, C_i\}$ can be easily determined.

A serious problem for identification of $\{\Theta_i(z)\}$ based on data $\{\widehat{\theta}_i(k)\}$ is that these measurements are not directly available. Indeed the measurement model in (2.10) implies that $P(k) = \mathcal{L}\theta(k)$ where $\mathcal{L}$ is the Laplacian matrix associated with coupling coefficients $\{a_{ij}\}$, and

$$P(k) = \begin{bmatrix} P_1(k) \\ \vdots \\ P_N(k) \end{bmatrix}, \quad \theta(k) = \begin{bmatrix} \theta_1(k) \\ \vdots \\ \theta_N(k) \end{bmatrix}$$

are collective bus powers and phase angles, respectively. Since the Laplacian matrix $\mathcal{L}$ has a zero eigenvalue, power measurements cannot be used to estimate the phase angles at each time instance. In fact even if we collect the noise-free measurements of $\{P_i(k)\}$ over many time samples and assume that the phase angles are time invariant over the given time interval, the phase angles are still not uniquely determined. Hence an alternative has to be sought.

Recall our dynamic state estimation algorithm proposed in (2.13). The fact that measurements of one of the phase angles are available, prompts us to use measurement equation (2.9):

$$y_i(k) = d_i P_i(k) + g_i \theta_i(k) + \eta_i(k)$$

over some time interval. In the measurement equation, $d_i > 0$ for each i but $g_i > 0$ for only one $i \in \mathcal{N}$, and $\{\eta_i(k)\}$ are measurement noises. That is, we need measurements of one

phasor with the rest relative measurements. Denote collective measurements and noises as

$$y(k) = \begin{bmatrix} y_1(k) \\ \vdots \\ y_N(k) \end{bmatrix}, \quad \eta(k) = \begin{bmatrix} \eta_1(k) \\ \vdots \\ \eta_N(k) \end{bmatrix},$$

respectively. Then the measurement equations in (2.9) for $1 \leq i \leq N$ can be packed into the following collective representation:

$$y(k) = (D\mathcal{L} + G)\theta(k) + \eta(k)$$

at each given time sample. Lemma 3 indicates that diagonal $D > 0$ and diagonal $G \geq 0$ with rank 1 exist such that

$$(D\mathcal{L} + G) + (D\mathcal{L} + G)' > 0.$$

Hence $\mathcal{M} = (D\mathcal{L} + G)$ is a full rank matrix, and

$$\widehat{\theta}(k) = (D\mathcal{L} + G)^{-1} y(k)$$

is the maximum likelihood (ML) estimate of $\theta(k)$, provided that $\eta(k)$ is a vector-valued temporally white process with mean zero and covariance Σ_η, which holds in practice. Clearly $\widehat{\theta}(k)$ is an unbiased estimate of $\theta(k)$ with error covariance given by

$$\mathrm{E}\left\{[\theta(k) - \widehat{\theta}(k)][\theta(k) - \widehat{\theta}(k)]'\right\} = \mathcal{M}^{-1}\Sigma_\eta \mathcal{M}'^{-1}.$$

Denote $\Sigma_{v_\theta} = \mathcal{M}^{-1}\Sigma_\eta \mathcal{M}'^{-1}$. Then $\widehat{\theta}(k) = \theta(k) + v_\theta(k)$ with $v_\theta(k)$ a white process of mean zero and covariance Σ_{v_θ}.

It is important to note that the voltage phase angles may not be decoupled, i.e., the cross product terms may not be negligible. This will affect the estimate of the PSD and in turn negatively affect the model that generates the phase angles. However we believe that the couplings among different phase angles are weak in normal operation of the power grid. In the least we can assume that the phase angles are approximately decoupled over each time interval. The cross coupling takes place only at some discrete time samples. For this reason, our proposed distributed estimation algorithm has its merit.

2.3 Secure Smart Grid

This section will develop a novel detection method for unobservable attacks widely known for state estimation in power systems. We will demonstrate that our results on distributed dynamic state estimation obtained in the previous section can be used to develop a new detection algorithm that is especially effective for unobservable attacks.

2.3.1 A novel detection algorithm

Prior to presenting our new detection algorithm, a conventional anomaly detector, discussed in Chapter 1, will be presented first. Such a conventional approach can detect only observable attacks.

Conventional Approach

Recall the power measurement model that can be written as

$$y(k) = (D\mathcal{L} + G)Cx(k) + Ew(k) \tag{2.24}$$

$$= \mathcal{M}\theta(k) + Ew(k), \tag{2.25}$$

$$\theta(k) = Cx(k). \tag{2.26}$$

When there is an attack vector a, the measurement equation becomes

$$y_a(k) = \mathcal{M}\theta(k) + Ew(k) + a, \tag{2.27}$$

For an observable attack, it changes the measurement residual defined as

$$\begin{aligned} \|y_a(k) - \hat{y}_a(k)\| &= \|\mathcal{M}\theta(k) + Ew(k) + a - \mathcal{M}\hat{\theta}\| \\ &= \|y(k) - \hat{y}(k) + a\|. \end{aligned}$$

Hence the residual increases due to presence of the attack vector a, which is indeed the case when the adversary has no knowledge of the topology of the grid. However when adversaries have knowledge of the topology of the grid, they can find $\theta_a \in \mathbb{R}^n$, such that the attack

$a = \mathcal{M}\theta_a$, and it becomes unobservable, because such an attack vector a does not alter the residual as shown next:

$$\begin{aligned}\|y_a(k) - \hat{y}_a(k)\| &= \|\mathcal{M}\theta(k) + Ew(k) + \mathcal{M}\theta_a(k) - \mathcal{M}(\hat{\theta}(k) + \theta_a(k))\| \\ &= \|y(k) - \hat{y}(k)\|.\end{aligned}$$

That is, the attack a becomes a part of the parameter vector θ. As a result when the attack is unobservable, it can bypass the residual test, which renders the conventional approach unsuitable for detecting the unobservable attacks.

A New Approach

As explained earlier, conventional ways for detect the malicious attack are based on system measurement data as well as statistical information of the measurement model. When attacks or anomalies exist, and the measurements and the statistical model do not match each other, then attacks or anomalies can be detected. But if the attack has knowledge of the grid topology, i.e., the measurement model, then the attacker will be able to bypass the residue test by constructing the attack vector in the range space of the measurement model. So it becomes a part of the parameter vector and is thus termed as *unobservable*. Many research papers are devoted to tackle such a difficult detection problem [29, 4, 31, 28, 76, 77]. Fortunately the dynamic state estimation as presented in the previous section can be used to derive a new detector, and the consensus approach can be implemented to detect unobservable attack.

According to the previous discussion on real power flow model and state estimation, the discrete-time static measurement equations in (2.12) are the same as

$$y(k) = \mathcal{M}\theta(k) + Ew(k), \quad \theta(k) = Cx(k), \tag{2.28}$$

$$x(k+1) = Ax(k) + Bw(k), \quad \widehat{\theta(k)} = \mathcal{M}^{-1}y(k), \tag{2.29}$$

where $\widehat{\theta(k)}$ is the phasor angle vector calculated directly from the power measurement vector $y(k)$. The static estimate $\widehat{\theta(k)}$ has N components denoted as

$$\widehat{\theta(k)} = \begin{bmatrix} \widehat{\theta_1(k)} \\ \vdots \\ \widehat{\theta_N(k)} \end{bmatrix}.$$

For the dynamic state estimator in (2.13), estimation equations can be written as

$$\hat{y}(k) = \mathcal{M}\hat{\theta}(k), \tag{2.30}$$

$$\hat{\theta}(k) = C\hat{x}(k) = \mathcal{M}^{-1}\hat{y}(k), \tag{2.31}$$

$$\hat{x}(k+1) = A\hat{x}(k) - DK[\hat{y}(k) - y(k)] - GK[\hat{\theta}(k) - \theta(k)] + KDEw(k), \tag{2.32}$$

$$= (A - K\mathcal{M}C)\hat{x}(k) + K\mathcal{M}Cx(k) + KDEw(k), \tag{2.33}$$

Recall the output residual defined as

$$e_y(k) := y(k) - \hat{y}(k). \tag{2.34}$$

Since the output residual can not tell which bus is under attack, we modify the above equation by multiplying it with $\mathcal{M}^{-1}$ from left. This results in the output angle residual as

$$e_\theta = \mathcal{M}^{-1}e_y(k) = \widehat{\theta(k)} - \hat{\theta}(k), \tag{2.35}$$

Thus, by monitoring the angle residual, it is possible to determine which bus is under attack. Under the normal operating condition, i.e., in absence of attack and fault of the power grid, the residuals are very small due to

$$\|\widehat{\theta(k)} - \hat{\theta}(k)\| = \|[\theta(k) - \hat{\theta}(k)] + \mathcal{M}^{-1}\eta\|, \tag{2.36}$$

$$\mathrm{E}\{[\widehat{\theta(k)} - \hat{\theta}(k)][\widehat{\theta(k)} - \hat{\theta}(k)]'\} = C\Sigma C' + \Sigma_\eta + \mathcal{M}^{-1}\Sigma_\eta\mathcal{M}'^{-1} \tag{2.37}$$

in light of (2.21). If there is an attack, the residual will significantly increased. Specifically the state space model under attack can be described as

$$y_a(k) = \mathcal{M}\theta(k) + Ew(k) + a, \tag{2.38}$$

$$x(k+1) = Ax(k) + Bw(k), \tag{2.39}$$

where a is the attack vector. For observable attacks, the residual is

$$\|\widehat{\theta(k)} - \hat{\theta}(k)\| = \|[\theta(k) - \hat{\theta}(k)] + \mathcal{M}^{-1}(\eta + a)\|, \tag{2.40}$$

$$\mathrm{E}\{[\widehat{\theta(k)} - \hat{\theta}(k)][\widehat{\theta(k)} - \hat{\theta}(k)]'\} = C\Sigma C' + \Sigma_\eta + \mathcal{M}^{-1}(\Sigma_\eta + \Sigma_a)\mathcal{M}'^{-1}. \tag{2.41}$$

In the case of unobservable attacks, adversaries can find $\theta_a \in \mathbb{R}^n$ such that $a = \mathcal{M}\theta_a$. Then the residual is given by

$$\|\widehat{\theta(k)} - \hat{\theta}(k)\| = \|[\theta(k) - \hat{\theta}(k)] + \mathcal{M}^{-1}\eta + \theta_a\|, \tag{2.42}$$

$$\mathrm{E}\{[\widehat{\theta(k)} - \hat{\theta}(k)][\widehat{\theta(k)} - \hat{\theta}(k)]'\} = C\Sigma C' + \Sigma_\eta + \mathcal{M}^{-1}\Sigma_\eta\mathcal{M}'^{-1} + \Sigma_a. \tag{2.43}$$

The above shows that no matter what the attack is (observable or unobservable), the residual always includes the attack component. Therefore the residual increases if the attack is present. For this reason, our new approach is capable of detecting the attack by examining the distribution of the output residual sequence, and compare the residual $\|\widehat{\theta(k)} - \hat{\theta}(k)\|$ with a threshold τ. An attack exists, if

$$\|\widehat{\theta(k)} - \hat{\theta}(k)\| > \tau. \tag{2.44}$$

2.3.2 Simulation studies

To illustrate our proposed detection approach, we carried out a simulation study for the IEEE 14-bus test system. The configuration data of the test systems is obtained from the data sheet of IEEE 14-bus [78]. To reduce the calculation complexity, all the generators are removed. In this system, we set bus-1 as the slack bus with its power angle being always

equal to 0^o. Under the normal operation, a power system operates in a quasi-steady-state. The bus loads are constant over short time intervals. The customer activities are modeled as a wide-sense stationary (WSS) random process, denoted as $v(k)$. As convention, we assume that $v(k) \sim \mathcal{N}(0, \sigma_v^2)$ is an independent process. Each bus is assumed to be described by the following random walk Gauss-Markov equation

$$x_i(k+1) = x_i(k) + v_i(k) \tag{2.45}$$

where the initial condition $x_i(0) = x_{i0}$ a Gauss random variable, independent of $\{v_i(k)\}$ for all i. Hence $x_i(k)$ is also a normal process. The variance of the process noise is set as $\sigma_v = 0.01$ in order to model slow variation of the phasor angle due to power consumers, considering that the sampling rate for a typical SCADA system in a power grid is about 100 $\sim$ 200 Hz [3, 54]. The following figure shows a typical power angle θ_i on bus-8 under the normal condition. In this case, on bus-8 a typical power angle θ_i fluctuates around -20^o:

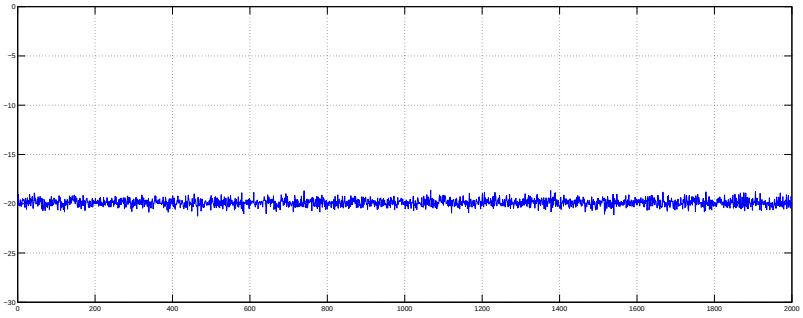

Figure 2.1: Typical power angle plot

For the measurement equation, the signal to noise ratio (SNR) is set to be 10dB. That is, $\sigma_\eta = \sigma_y/\sqrt{10}$ with σ_y^2 as the variance of the measurement signal. By convention, $\eta(k) \sim \mathcal{N}(0, {\sigma_\eta}^2)$ is an independent Gaussian noise process. In the simulation, a total of 2000 samples is taken. Both the conventional approach and the consensus approach are used under both

observable and unobservable attacks. The next figure shows the residual responses when the attack is absent in the system.

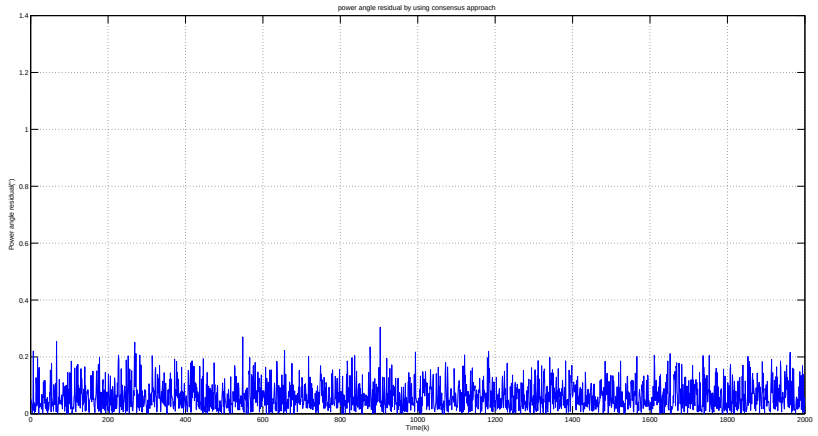

Figure 2.2: Angle residual (consensus approach, no attack)

The method used in estimation is the consensus approach, based on local and distributed measurements, and random walk Markov Gauss dynamic state space models. It is seen that the estimation error is quite small. We also simulated the conventional approach to estimating the power phaser angles, without using dynamic state space models, and the replay attack is absent. In fact the same data as from the previous figures is used. The error residual is very close to that used in the consensus approach. These two results seem to admit similar estimation performance.

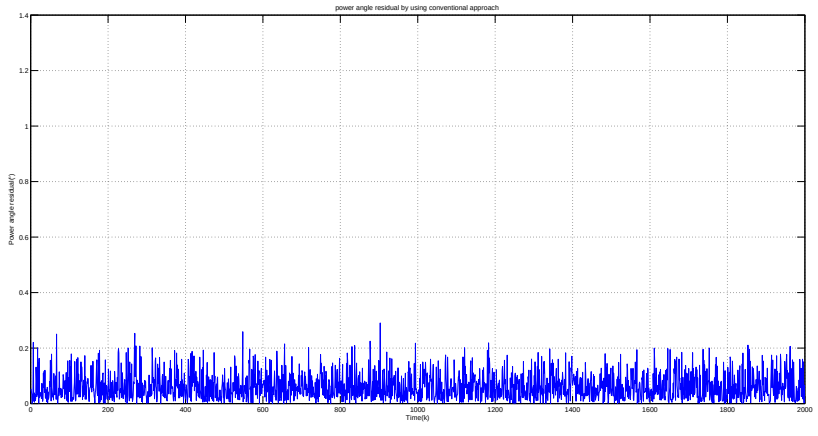

Figure 2.3: Angle residual (conventional approach, no attack)

Next we consider malicious attacks in the form of observable attacks. We assume that the attack vector a as a constant vector by simply setting its ith component $a_i = 1$, and the attack starts at the 500th time sample. The residual responses are shown in the next figure. The sudden jump in the error response plot indicates the presence of the attack.

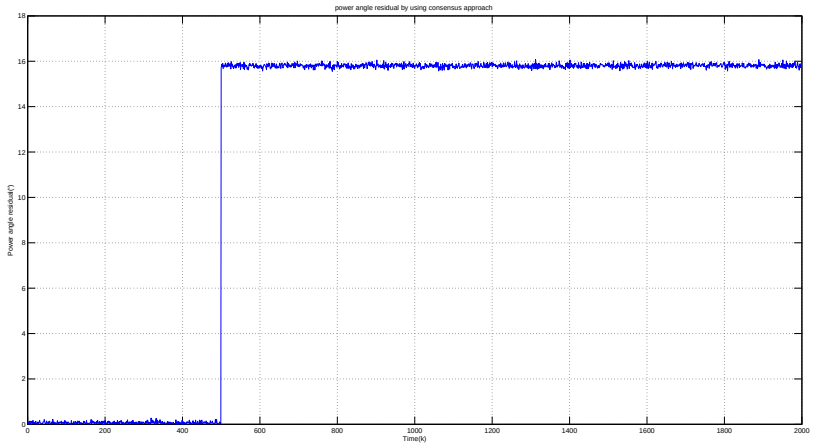

Figure 2.4: Angle residual (consensus approach, observable attack)

We comment that the reason for the delectability lies in the fact that the dynamic state space model is used and consensus estimation approach is adopted, and thus the estimated angle cannot change abruptly as the actual angle change due to the attack. Hence the estimation error increases right away. It needs to be pointed out that the conventional approach can also detect the observable attack as illustrated by the angle residual plot next.

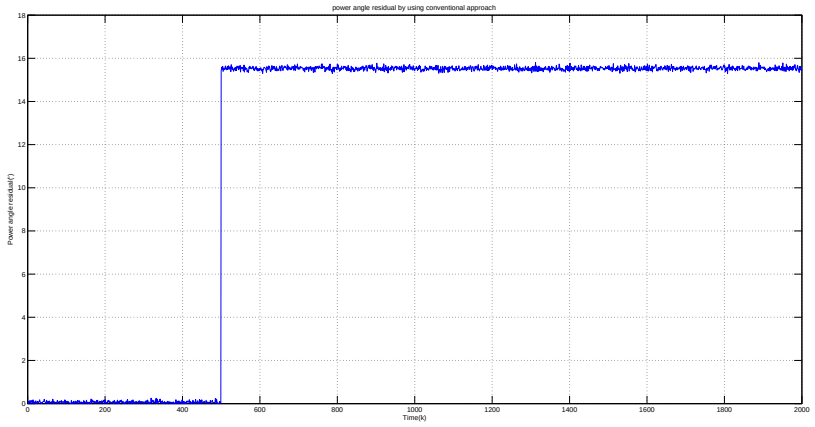

Figure 2.5: Angle residual (conventional approach, observable attack)

The more interesting case is when unobservable attacks exist. We suppose that there is only one attack on the bus-8 starting at the 500th sample. The attack vector can be set as $a = \mathcal{M}\theta_a$ with only $\theta_{a_8} = 1$, resulting in an unobservable attack. Using the consensus approach, the residual responses are shown in the next figure.

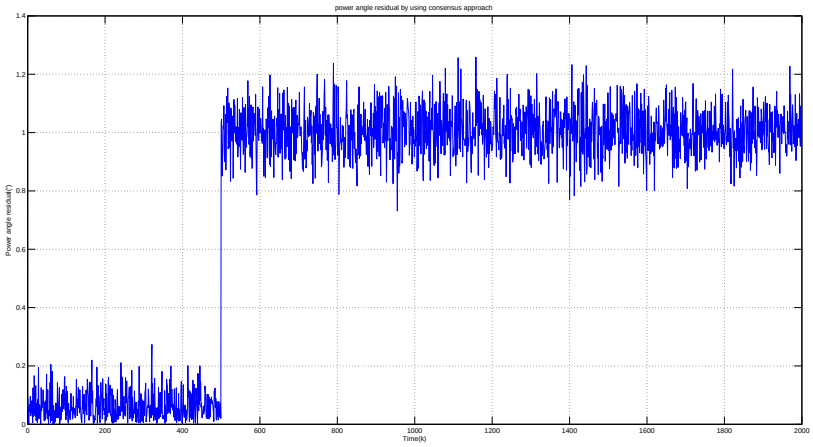

Figure 2.6: Angle residual (consensus approach, unobservable attack)

Even though the attack on the bus-8 has a small amplitude, the residual shows clearly a significant increase in the estimation error for the phasor angle of the bus-8. The underlying reason for the detectability is the same as the case of observable attack. The next figure shows the averaged error residual with the same window size.

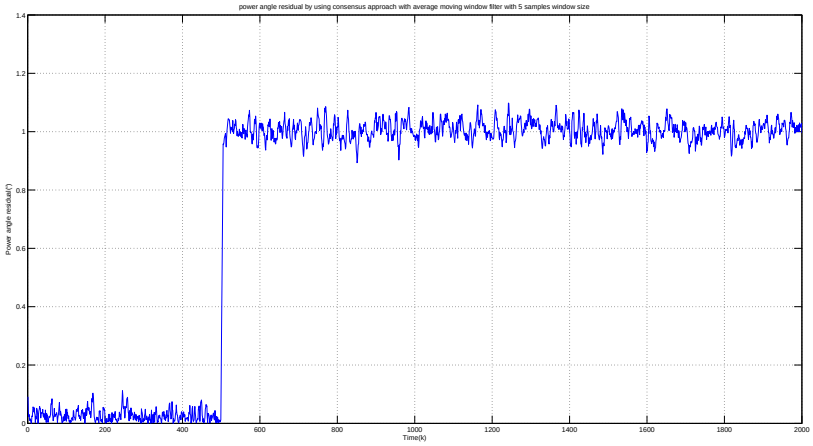

Figure 2.7: Averaged angle residual (consensus approach, unobservable attack)

However when the conventional approach is taken, the situation is rather different as illustrated next by the angle residual.

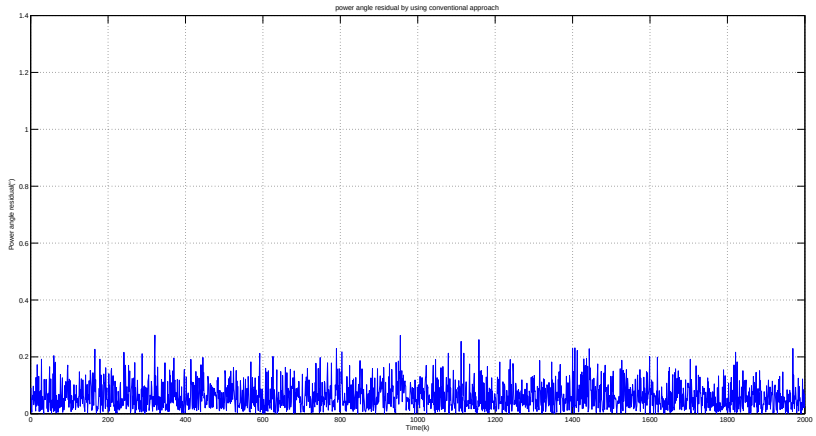

Figure 2.8: Angle residual (conventional approach, unobservable attack)

It shows that under the unobservable attack, the residual obtained from conversional approach has no significant change. The next figure shows the estimated angle for both the conventional (red) and consensus (blue) approaches.

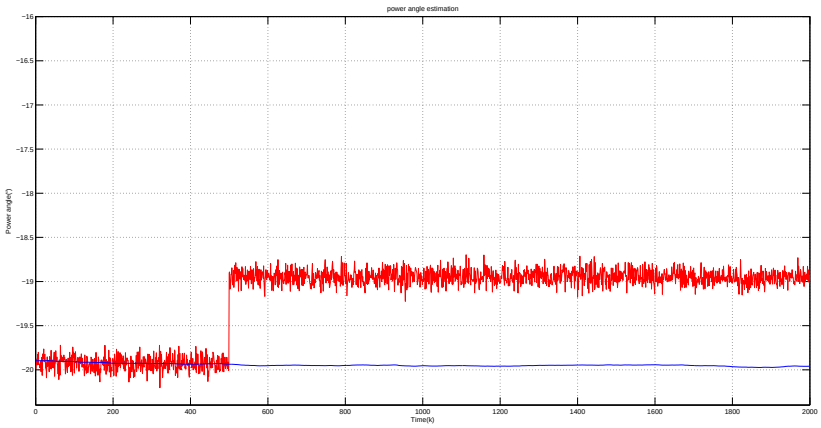

Figure 2.9: Estimated angle (conventional and consensus approach, unobservable attack)

The above simulation studies demonstrated that our proposed consensus approach is capable of detecting both the observable and unobservable attack, while the conventional approach fails.

We also provide our simulation study for the detection rate when the unobservable attacks are present with different attack amplitudes. The following figure shows the detection rate when the false alarm rate is kept at 5% using the consensus approach.

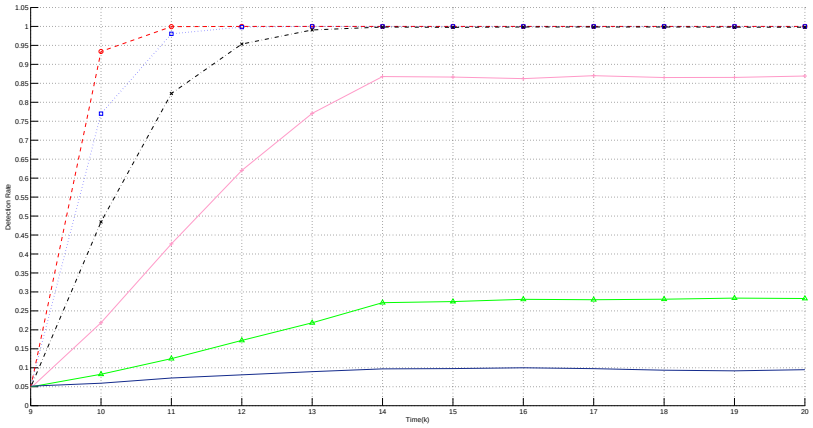

Figure 2.10: Detection rate under different attack magnitudes

In the above figure, we set the window size equals to 5, and the attack magnitudes vary. The vertical axis represents the detection rate while the horizontal axis giving the time samples. The red dash line with $\circ$ shows the detection rate when the attack amplitude is 5% of the actual power angle; The blue dot line with $\square$ shows the detection rate when the attack amplitude is 4% of the actual power angle; The detection rate curves for 3%, 2%, 1%, and 0.5% of the attack amplitude shown in black dot dash line with $\times$, pink solid line with $+$, green solid line with Δ, and purple solid line, respectively.

The detection rate curves in Figure 2.10 are results of the average of 10,000 ensemble experiments. To reduce the calculation time, we set the time sample size of each trial to 20. The unobservable attack start at time sample 10. Hence, each curve start at time sample 9 and with false alarm rate equal to 5%. Note that the detection rate decreases as the attack magnitude reduces, considering that as the attack amplitude decreases, it will be submerged into the noise, and thus becomes undetectable.

The simulation study in Figure 2.10 is carried out with a window size fixed to be 5. In practice we also want to detect the attack as quickly as possible after the attack is launched. Thus the window size can not be too large. For this reason, we now fix the attack magnitude to 2% and compare the detection rate of different window size when the false alarm rate is still fixed at 5% with the attack starting at time sample 10.

Fig. 2.11 shows the detection rate curve under 5% false alarm rate when the window sizes vary. In the above figure, the blue solid line with $\circ$) corresponds to the window size of 7, the red dash line with $\square$ to the window size of 6, the black dot dash line with $\times$ to the window size of 5, the pink dot line with $\diamond$ to the window size of 4, the green solid line with Δ to the window size of 3, the purple solid line $\star$ to the window size of 2, and the blue solid line to the window size of 1.

It is worth to noticing that reducing window size increases the sensitivity of the detector. However it also reduces the steady-state detection rate. On the other hand increasing the window size improves the steady-state detection rate, but the latency is also increased. So there is a tradeoff between the window size and detection rate versus the sensitivity. The choice of window size depends on the power system protection settings.

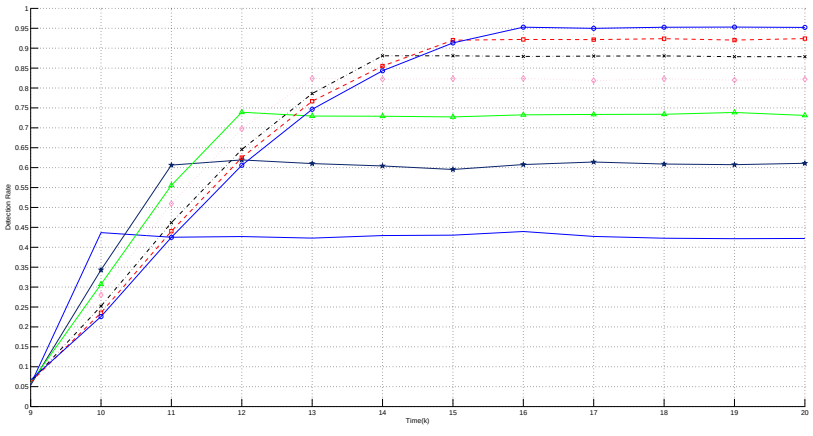

Figure 2.11: Detection rates for different window sizes

2.4 Conclusion

We have proposed a viable distributed dynamic state estimator based on the multi-agent system (MAS) framework, assuming the availability of SCADA systems. A useful result from [73] is employed to model the phasers as WSS processes, and to develop a viable dynamic state estimator. We have assumed that the power system is described by a linear DC power flow model and is operating in a quasi-steady state. Conditions on the existence of a a distributed estimator were derived and a solution was proposed to construct the distributed estimation gains. Stability of the distributed state estimator was shown to hold. Although we assumed that the phase angles are approximately decoupled, the results presented in this chapter provides an initial attempt to tackle the dynamic estimation problem for the power grid, and can be valuable to future research in this important research problem area. More importantly with the dynamic state estimator and DC power flow model, we proposed a consensus approach to detecting malicious attacks. While both the conventional and our proposed approaches are effective for observable attacks, it is not the case for the unobservable attacks. In fact the conventional approach fails utterly. On the other hand our proposed approach is capable of detecting the unobservable attacks, verified by our simulation studies.

3 Whitening Filter Approach

Recently replay attack has become a serious threat to the SCADA system. Adversaries record a sequence of sensor measurements and replay the sequence afterwards while conducting attacks on the system [2, 34, 79]. A well known replay attack is launched by the Stuxnet malware. It is a computer worm virus that was discovered in June 2010 designed to attack Siemens Step 7 software allegedly which is a widely used software for PLC programming. It allegedly attacked the Iran's uranium enrichment plant at Natanza in mid to late 2009 and caused 984 which is one fifth of the centrifuges damaged. Stuxnet virus secretly recorded the normal operations status readings when the system runs under normal condition, and then played those readings back to the system operators when the systems failed. It is like many Hollywood movies that bad guys play the pre-recorded security tape when they rob the bank. Thus, it would appear to the operator that everything was running smoothly while the system was already damaged. That prevents a safety system from doing some action to prevent abnormal operation. As the PLC is used in most of the industry infrastructures, especially in power grid, the replay attacks entice vulnerability of the power devices. Recently investigation and report by a computer security software maker Symantec Corporation show that they catched the Stuxnet in a global malware collection system. The Stuxnet has already begun to spread around the globe. Not only in Iran, but also in India, Indonesia and other countries have spotted this kind of worm . No one knows whether their systems have been infected of such a worm, and when it starts attacking their systems. This is thus an urgent call to all of us to develop methods to protect against replay attack [60, 33].

This chapter will study the detection problem of the replay attack, and is endeavored to contribute to this emerging research area. A whitening filter approach is proposed and shown to be effective. We begin our chapter with the literature survey, followed by presentation of the background material. Our proposed detection method based on whitening filters will be detailed together with the computation aspect. The chapter will be concluded with simulation studies.

3.1 Literature Survey

Mo and Sinopoli are the first [2] to consider the replay attack, one year before the Stuxnet was reported in the news media. Their proposed method is quite simple. Its basic idea is to inject a Gauss random noise in the control input, and test the estimation residue for the output estimation error of the Kalman filter. They assume that the control system is a discrete time linear time invariant (LTI) Gaussian with an infinite horizon Linear Quadratic Gaussian (LQG) controller in which a Kalman filter is employed to estimate the system's state. A χ^2 failure detector is used to detect the presence of the replay attack. Because the added white noise is known to the controller, it is canceled in the Kalman filter if the replay attack is absent. However when the replay attack is present, then the added white noise at the control input cannot be canceled, resulting in higher output estimation error. For this reason the added white Gauss noise serves a time stamp.

To be specific, the LTI system state dynamics are described by

$$x_{k+1} = Ax_k + Bu_k + w_k,$$

$$y_k = Cx_k + v_k,$$

with $x_k \in \mathbb{R}^n$ is the state variable, u_k is the control input, and $y_k \in \mathbb{R}^m$ is the output measurement at time k. The initial condition of the state vector $x_0 \sim \mathcal{N}(\bar{x}_0, \Sigma)$. The vector signals $w_k \in \mathbb{R}^n$ is the process noise with Gaussian distribution $w_k \sim \mathcal{N}(0, Q)$, and v_k is the measurement noise with Gaussian distribution $v_k \sim \mathcal{N}(0, R)$.

The LQG controller consists of the Kalman filter and state feedback gain. Assumed that the LQG system is in the steady-sate. Denote $\hat{x}_{k|i}$ as the optimal estimate for x_k conditioned on all the measurements up to time index i. The Kalman filter computes the optimal state estimate recursively according to

$$\begin{aligned}\hat{x}_{k|k} &= (I - KC)\hat{x}_{k|k-1} + Ky_k, \quad \hat{x}_{k+1|k} = A\hat{x}_{k|k} \\ K &= PC'(R + CPC')^{-1}, \\ P &= APA' - APC'(R + CPC')^{-1} + Q,\end{aligned}$$

initialized with $\hat{x}_{0|-1} = \bar{x}_0$. By using the state estimate $\hat{x}_{k|k}$, the LQG control input has the state feedback form, given by

$$u_k = u_k^\star = L\hat{x}_{k|k}, \quad L = -(U + B'SB)^{-1}B'SA,$$

where S is the stabilizing solution to the following algebraic Riccati equation (ARE):

$$S = A'SA + W - A'SB(B'SB + U)^{-1}B'SA.$$

For LQG control, the one-step prediction output error $y_k - C\hat{x}_{k|k-1}$ is white, which is an important property of the Kalman filter. Its covariance is given by

$$\mathrm{E}\{(y_k - C\hat{x}_{k|k-1})(y_k - C\hat{x}_{k|k-1})'\} = R + CPC' =: \mathscr{P}.$$

Hence the normalized output error variance satisfies the following equation:

$$\mathrm{E}\{\|\mathscr{P}^{-1/2}(y_k - C\hat{x}_{k|k-1})\|^2\} = \mathrm{E}\{(y_k - C\hat{x}_{k|k-1})'\mathscr{P}^{-1}(y_k - C\hat{x}_{k|k-1})\} = m.$$

It is important to note that the normalized output error

$$(y_k - C\hat{x}_{k|k-1})'\mathscr{P}^{-1}(y_k - C\hat{x}_{k|k-1})$$

has a χ^2 distribution, and χ^2 detector is often employed to detect any anomalies of the feedback control system. However it does not help for detection of the replay attack.

Supposed that the attacker knows all sensor readings and can record and replay them. Denote the modified reading as $\tilde{y}_k$ with $\tilde{y}_k = y_{k-T}$ for some large integer $T > 0$. It is proven in [2] that under the replay attack the residues $\tilde{y}_k - C\hat{x}_{k|k-1}$ will converges to the same distribution as of $y_k - C\hat{x}_{k|k-1}$, i.e.,

$$\mathrm{E}[(\tilde{y}_k - C\hat{x}_{k|k-1})'\mathscr{P}^{-1}(\tilde{y}_k - C\hat{x}_{k|k-1})] = E[(y_k - C\hat{x}_{k|k-1})'\mathscr{P}^{-1}(y_k - C\hat{x}_{k|k-1})] = m.$$

Thus the conventional χ^2 detector is useless in detecting the replay attack.

To remedy the above problem, Mo and Sinopoli [2] propose to redesign the controller by adding a Gauss white noise signal to the control input, leading to

$$u_k = u_k^* + \Delta u_k, \quad \Delta u_k \sim \mathcal{N}(0, \mathscr{L}),$$

where u_k^* is the original optimal LQG control signal, and u_k^* and Δu_k are independent of each other. In this case the residues $\{\tilde{y}_k - C\hat{x}_{k|k-1}\}$ do not converges to the same distribution as $y_k - C\hat{x}_{k|k-1}$. To be specific, in absence of attack,

$$\mathrm{E}[(y_k - C\hat{x}_{k|k-1})'\mathscr{P}^{-1}(y_k - C\hat{x}_{k|k-1})] = m,$$

Under the replay attack, there holds

$$\mathrm{E}[(\tilde{y}_k - C\hat{x}_{k|k-1})'\mathscr{P}^{-1}(\tilde{y}_k - C\hat{x}_{k|k-1})] = m + 2\mathrm{Tr}(C'\mathscr{P}^{-1}C\mathscr{U}),$$

where $\mathrm{Tr}(\cdot)$ denotes trace, and $\mathscr{U}$ satisfies the following Lyapunov equation

$$\mathscr{U} - B\mathscr{L}B' = \mathscr{A}\mathscr{U}\mathscr{A}'.$$

with $\mathscr{A} \triangleq (A + BL)(I - KC)$. As a result the replay attack is can be detected using the χ^2 detector at each time index k in the following form

$$g_k = \sum_{i=k-\mathscr{T}+1}^{k} (y_k - C\hat{x}_{k|k-1})'\mathscr{P}^{-1}(y_k - C\hat{x}_{k|k-1}) \lessgtr \text{threshold}$$

where $\mathscr{T}$ is the window size to estimate the expectation, which means that g_k is χ^2 distributed with $\mathscr{T}$ degrees of freedom.

The threshold depends on the requirement of detection rate versus false alarm rate. In addition the greater the covariance $\mathscr{L}$ of Δu is, the higher the detection rate and also the greater loss of system performance will be. As shown in [2] for a particular example, this method achieves the detection rate achieves 35% with the covariance $\mathscr{L} = 0.6$, but at the same time the LQG performance is lost 91% with respect to the optimal LQG cost. There thus exists a tradeoff between the detection rate and control performance.

To reduce the loss of performance, [1] applies a game theoretic method to detection of the replay attack. Based on the method of Mo and Sinopoli, [1] also considers the LTI plant using the Kalman filter, optimal LQG controller, and χ^2 detector. The difference lies in employment of a finite horizon, zero-sum, non-stationary stochastic game approach to reduce the control and detection cost. The authors of [1] use switching between two different controllers according to the system dynamics. One is the secure controller (capable of detecting the replay attack but with suboptimal cost) designed by Mo and Sinopoli with the control input signal $u_k = u_k^* + \Delta u_k$. The other one is the optimal controller (incapable of detecting detect replay attack) with no additional noise signal Δu_k being added to the control input signal. An optimal control policy is developed based on the game theory framework by switching between the two controllers shown in Fig. 3.1. The simulation results show thatthe system performance loss is reduced half while detection rate is only reduced 5 percent.

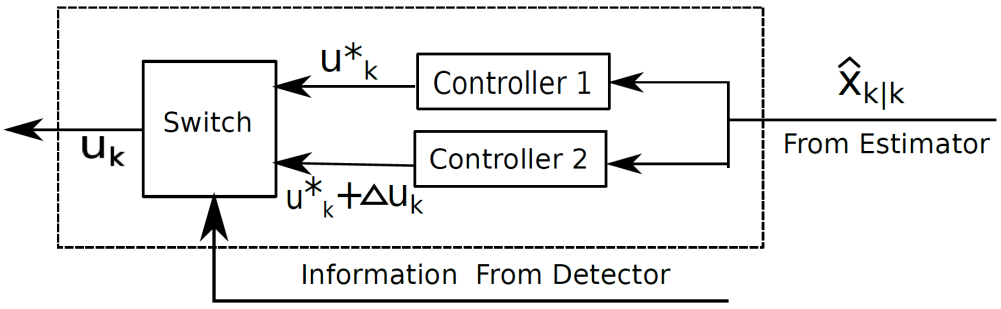

Figure 3.1: Block Diagram of switching controller [1]

To reduce the system performance loss, Tran and Shin [80] also designed a new detection scheme based on the method provided by Mo and Sinopoli [2]. Instead of adding random signal Δu_k to control input signal continuously through the whole control process, they modified the way by adding the random signal Δu_k periodically only for small time duration with the period T. So the system will operate normally in the remaining time. By adjusting the period T of adding Δu_k, the detection rate can be guaranteed and the system performance loss will also be reduced. The simulation results show that the system performance loss can be reduced when there is no attack, while providing an adjustable detection rate for the replay attack [80].

The method of Mo and Sinopoli provides a basic way of detecting the replay attack at the expense of losing some control performance. The second and third methods are able to reduce the system control performance loss by sacrificing the detection rate. Although these methods are all effective, they can not eliminate the injection of the random noise to the control signal. In addition the exiting detection methods are all based on the LQG controller. The applicability to those non-LQG feedback control systems is questionable. In this chapter and the next, we will propose two different new approaches without adding the Gauss white noise to the control input by recognizing that the use of network and digital channels in SCADA systems, the communication noises can be used to detect the replay attacks.

3.2 Main Results

This section presents our whitening filter approach to detection of the replay attack. The first subsection will prepare the background material prior to presentation of our proposed detection method. Computational issues will be studied in the third subsection.

3.2.1 Preliminaries

A SCADA system is also a networked control system (NCS) that has the feedback controller situated in a different physical location from that of the plant, and it communicates

with the plant via a (often wireless) network. The use of networks in feedback control systems thus creates a vulnerability for malicious attacks which seek to destabilize and damage the physical system. Specifically consider the discrete-time feedback control system in the block diagram next where the time index k is integer-valued.

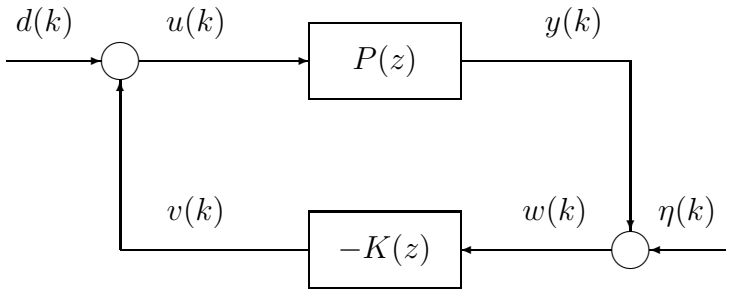

Figure 3.2: Networked feedback control system

In Fig. 3.2, the steady-state responses of the input and output are removed in order to understand better the real issue in secure feedback control. Let q^{-1} be the unit delay operator. In accordance with the block diagram in Fig. 3.2, there holds

$$\begin{aligned}
\begin{bmatrix} v(k) \\ w(k) \end{bmatrix} &= \begin{bmatrix} v(k) \\ y(k) \end{bmatrix} + \begin{bmatrix} 0 \\ \eta(k) \end{bmatrix} \\
&= \begin{bmatrix} 0 & -K(q) \\ P(q) & 0 \end{bmatrix} \begin{bmatrix} u(k) \\ w(k) \end{bmatrix} + \begin{bmatrix} 0 \\ \eta(k) \end{bmatrix} \\
&= \begin{bmatrix} 0 & -K(q) \\ P(q) & 0 \end{bmatrix} \begin{bmatrix} v(k) \\ w(k) \end{bmatrix} + \begin{bmatrix} 0 \\ P(q)d(k) + \eta(k) \end{bmatrix} \\
&= \begin{bmatrix} I & K(q) \\ -P(q) & I \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ P(q)d(k) + \eta(k) \end{bmatrix} \\
&= \begin{bmatrix} -K(q) \\ I \end{bmatrix} [I + P(q)K(q)]^{-1} \begin{bmatrix} P(q) & I \end{bmatrix} \begin{bmatrix} d(k) \\ \eta(k) \end{bmatrix}.
\end{aligned}$$

The above indicates that the closed-loop transfer matrix from the exogenous inputs $\{d(k)\}$ and $\{\eta(k)\}$ to $\{w(k)\}$ and $\{v(k)\}$, the input/output signals of the feedback controller $-K(z)$,

is given by

$$T_K(z) = \begin{bmatrix} -K(z) \\ I \end{bmatrix} [I + P(z)K(z)]^{-1} \begin{bmatrix} P(z) & I \end{bmatrix}. \tag{3.1}$$

A similar derivation yields

$$\begin{bmatrix} u(k) \\ y(k) \end{bmatrix} = \begin{bmatrix} I \\ P(q) \end{bmatrix} [I + K(q)P(q)]^{-1} \begin{bmatrix} I & -K(q) \end{bmatrix} \begin{bmatrix} d(k) \\ \eta(k) \end{bmatrix}.$$

Thus the closed-loop transfer matrix from the exogenous inputs $\{d(k)\}$ and $\{\eta(k)\}$ to $\{u(k)\}$ and $\{y(k)\}$, the input/output signals of the plant model $P(z)$, given by

$$T_P(z) = \begin{bmatrix} I \\ P(z) \end{bmatrix} [I + K(z)P(z)]^{-1} \begin{bmatrix} I & -K(z) \end{bmatrix}. \tag{3.2}$$

Without loss of generality the plant model $P(z)$ with m-input/p-output is assumed to admit a stabilizable and detectable state-space realization. Its transfer matrix is given as

$$P(z) = D + C(zI - A)^{-1}B := \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]. \tag{3.3}$$

As a result a stabilizing state feedback gain F and a stabilizing state estimation gain L exist such that $(A + BF)$ and $(A + LC)$ are both Schur stability matrix. It is well known that $P(z)$ admits left/right coprime factorizations (LCF/RCF)

$$P(z) = \tilde{M}(z)^{-1}\tilde{N}(z) = N(z)M(z)^{-1} \tag{3.4}$$

with $\{\tilde{M}(z), \tilde{N}(z), M(z), N(z)\}$ all stable transfer matrices. Assume that the feedback system in Fig. 1 is internally stable. Then the controller $K(z)$ admits left and right coprime factorizations

$$K(z) = \tilde{V}(z)^{-1}\tilde{U}(z) = U(z)V(z)^{-1} \tag{3.5}$$

with $\{\tilde{V}(z), \tilde{U}(z), V(z), U(z)\}$ all stable transfer matrices satisfying the Bezout identity

$$\begin{bmatrix} \tilde{V}(z) & \tilde{U}(z) \\ -\tilde{N}(z) & \tilde{M}(z) \end{bmatrix} \begin{bmatrix} M(z) & -U(z) \\ N(z) & V(z) \end{bmatrix} = I_{m+p} \ \ \forall \ |z| \geq 1. \tag{3.6}$$

It is emphasized that the LCF and RCF for the plant model and for the stabilizing controller always exist and satisfy the Bezout identity in (3.6). The computation of such LCF and RCF for the plant and controller can be simplified for the observer-based controller

$$K_{\rm o}(z) = F(zI - A - BF - LC - LDF)^{-1}L. \tag{3.7}$$

Recall that $(A+BF)$ and $(A+LC)$ are both stability matrix. In this case $K_{\rm o}(z)$ admits left and right coprime factorizations

$$K_{\rm o}(z) = \tilde{V}_{\rm o}(z)^{-1}\tilde{U}_{\rm o}(z) = U_{\rm o}(z)V_{\rm o}(z)^{-1}$$

with realizations of its coprime factors together with those of coprime factors of $P(z)$ specified as

$$\left[\begin{array}{cc} \tilde{V}_{\rm o}(z) & \tilde{U}_{\rm o}(z) \\ -\tilde{N}(z) & \tilde{M}(z) \end{array}\right] = \left[\begin{array}{c|cc} A+LC & -(B+LD) & L \\ \hline F & I_m & 0 \\ \Omega^{-1}C & -\Omega^{-1}D & \Omega^{-1} \end{array}\right], \tag{3.8}$$

$$\left[\begin{array}{cc} M(z) & -U_{\rm o}(z) \\ N(z) & V_{\rm o}(z) \end{array}\right] = \left[\begin{array}{c|cc} A+BF & B & -L\Omega \\ \hline F & I_m & 0 \\ C+DF & D & \Omega \end{array}\right] \tag{3.9}$$

for any Ω that is nonsingular. The above is a slight modification from the existing literature. Then any stabilizing controller $K(z)$ for the feedback system in Fig. 1 has the form

$$K(z) = \left(\tilde{V}_{\rm o} + J\tilde{N}\right)^{-1}\left(\tilde{U}_{\rm o} - J\tilde{M}\right) = \left(U_{\rm o} - MJ\right)\left(V_{\rm o} + NJ\right)^{-1} \tag{3.10}$$

for some stable $J(z)$. It follows that the coprime factors of $K(z)$ in (3.5) given by

$$V(z) = V_{\rm o}(z) + N(z)J(z), \qquad U(z) = U_{\rm o}(z) - M(z)J(z), \tag{3.11}$$

$$\tilde{V}(z) = \tilde{V}_{\rm o}(z) + J(z)\tilde{N}(z), \qquad \tilde{U}(z) = \tilde{U}_{\rm o}(z) - J(z)\tilde{M}(z), \tag{3.12}$$

satisfy the Bezout identity (3.6). Conversely if the LCF/RCF of $K(z)$ in the above are available, then $J(z)$ can be obtained as

$$J(z) = \tilde{U}_0(z)V(z) - \tilde{V}_0(z)U(z) = \tilde{V}(z)U_0(z) - \tilde{U}(z)V_0(z). \tag{3.13}$$

Normally the exogenous inputs $\{d(k)\}$ and $\{\eta(k)\}$ are unknown disturbances and measurement noises, respectively. They are often white wide-sense stationary (WSS) random processes. However when attacks are present, $\{d(k)\}$ and $\{\eta(k)\}$ are replaced by $\{\alpha_u(k)+d(k)\}$ and $\{\alpha_y(k)+\eta(k)\}$, respectively, as shown next.

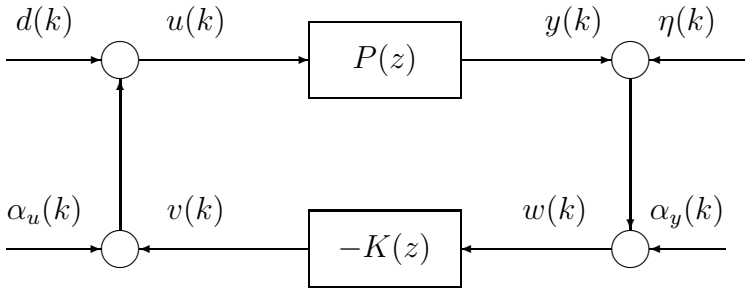

Figure 3.3: Feedback control system under attack

It is assumed that the signals available for monitoring are at the controller site, and hence $\{v(k), w(k)\}$ can be logged by the controller, while the most valuable data $\{u(k), y(k)\}$ from the physical system are unavailable. In fact only $\{w(k)\}$ need be collected for monitoring the underlying feedback control system, in light of the fact that $\{v(k)\}$ can be computed based on $\{w(k)\}$ and the controller model $K(z)$. In order to conceal the attack and induce damages, it is very likely that $\{\alpha_u(k)\}$ are unbounded, but $\{\alpha_y(k)\}$ are bounded. Even if bounded $\alpha_u(k)$ is used for malicious attack, it can be very irregular and disruptive in order to cause hardware damage of the physical system. As a result $\{\alpha_u(k)\}$ and $\{\alpha_y(k)\}$, injected by malicious attacks at the receiver end of the network, have different objectives: the former is aimed at replacing $u(k)$ so to damage the physical system while the latter is aimed at concealing the true output $y(k)$. For this reason early detection of the presence of the malicious attacks based on $\{v(k), w(k)\}$ becomes a very challenging problem facing the NCS control community.

This chapter will be focused on the replay attack studied initially in [2]. The notorious Stuxnet worm is an example of such an attack. It fits to the framework in Fig. 2 by taking

$$\alpha_y(k) = (q^{-\tau_\alpha} - 1)[y(k) + \eta(k)] \implies w(k) = y(k - \tau_\alpha) + \eta(k - \tau_\alpha) \tag{3.14}$$

for $k \geq k_\alpha$, assuming that the replay attack takes place at time k_α aimed at replaying the output τ_α samples before. This way conceals the real-time output of the plant and is probably the easiest way to fake the plant normal output. With $\tau_\alpha >> 1$, the catastrophic result of $y(k)$ under unbounded attack $\alpha_u(k)$ is not observed at the controller site until a very long time later.

Under the replay attack, $\alpha_y(k) = -[y(k) + \eta(k)] + \alpha_0(k)$ with

$$\begin{aligned}\alpha_0(k) &= y(k-\tau_\alpha) + \eta(k-\tau_\alpha) = P(q)u(k-\tau_\alpha) + \eta(k-\tau_\alpha) \\ &= P(q)[v(k-\tau_\alpha) + d(k-\tau_\alpha)] + \eta(k-\tau_\alpha).\end{aligned}$$

There thus holds $w(k) = \alpha_0(k)$ in light of (3.14). The replay attack results in

$$w(k) = P(q)[v(k-\tau_\alpha) + d(k-\tau_\alpha)] + \eta(k-\tau_\alpha) \tag{3.15}$$

for $k \in [k_\alpha,\ k_\alpha + \tau_\alpha)$. The above contrasts to the case in absence of attacks:

$$w(k) = P(q)[v(k) + d(k)] + \eta(k). \tag{3.16}$$

The replay attack is very effective so long as the feedback controller $K(z)$ is stable, by the fact that $w(k)$ in absence of attacks is statistically no different from that in presence of attacks. It is thus difficult to detect the presence of replay attacks. On the other hand when the feedback controller $K(z)$ is unstable, its output $v(k)$ diverges under the replay attack due to lack of real time information from the feedback measurements. Unfortunately many feedback controllers in engineering practice are stable. Mo and Sinopoli in [2] (see also [9])propose a smart strategy by injecting a random noise at the control input, which serves as a time stamp. While being effective, the injected noise degrades the control performance. In addition the results in [2] are restricted to LQG control systems. A true challenge for replay attack lies in its detection without injecting noises, applicable to commonly used control systems other than LQG, which will be studied in next subsections.

3.2.2 Detection algorithm

Detection of replay attacks is a very challenging problem. Injection of a random noise is effective to detect if a replay attack is present. However the noise injected at the plant input has to be large enough in order to achieve good detection performance, which deteriorates the control system performance. So there is a tradeoff between detection performance and control performance. An interesting problem is whether or not the communication noises present in the NCS can be used for detection of replay attacks without injecting noises. In this chapter the underlying NCS is assumed to employ network communications between the plant out and controller input over an additive white Gauss noise (AWGN) channel, and thus $\eta(k)$ present at the controller input has the form

$$\eta(k) = \eta_{\rm o}(k) + \eta_{\rm c}(k) \tag{3.17}$$

with $\eta_{\rm o}(k)$ for the measurement error and $\eta_{\rm c}(k)$ for the communication error induced by the AWGN channel. Both are i.i.d. random noises. It will be shown that the AWGN channel, while introducing information distortion, can help detection of the replay attack without injecting noise at the plant input, provided that the noise power due to the AWGN channel is not too small. This section consider the whitening filter approach. The following result is useful.

Lemma 5. *Assume that the plant model $P(z)$ in* (3.3) *admits a stabilizable and detectable realization, and $\{d(k), \eta(k)\}$ are both temporal white processes with covariance Q_d and Q_η, respectively. Let $Y_{\rm n} \geq 0$ be the stabilizing solution to the discrete-time algebraic Riccati equation* (DARE)*:*

$$Y_{\rm n} = AY_{\rm n}A' - (AY_{\rm n}C' + BQ_dD')Z_{\rm n}^{-1}(AY_{\rm n}C' + BQ_dD')' + BQ_dB' \tag{3.18}$$

where $Z_{\rm n} = Q_\eta + DQ_dD' + CY_{\rm n}C'$. *Then with* $L = L_{\rm n} := -(AY_{\rm n}C' + BQ_dD')Z_{\rm n}^{-1}$, *and the left coprime factors* $\{\tilde{M}(z), \tilde{N}(z)\}$ *of* $P(z)$ *in* (3.8) *replaced respectively by*

$$\left[\begin{array}{cc} \tilde{M}_{\rm n}(z) & \tilde{N}_{\rm n}(z) \end{array}\right] = \left[\begin{array}{c|cc} A + L_{\rm n}C & L_{\rm n} & (B + L_{\rm n}D) \\ \hline Z_{\rm n}^{-1/2}C & Z_{\rm n}^{-1/2} & Z_{\rm n}^{-1/2}D \end{array}\right], \tag{3.19}$$

$P(z) = \tilde{M}_{\rm n}(z)^{-1}\tilde{N}_{\rm n}(z)$, *and* $\{\tilde{M}_{\rm n}(z), \tilde{N}_{\rm n}(z)\}$ *satisfy the normalization condition*

$$\tilde{N}_{\rm n}(z)Q_d\tilde{N}_{\rm n}(z)^* + \tilde{M}_{\rm n}(z)Q_\eta\tilde{M}_{\rm n}(z)^* = I \quad \forall\ |z| = 1. \tag{3.20}$$

Proof: It is noted that the DARE in (3.18) can be written into the form of the discrete-time Lyapunov equation

$$Y_{\rm n} = (A + L_{\rm n}C)Y_{\rm n}(A + L_{\rm n}C)' + L_{\rm n}Q_\eta L_{\rm n}' + (B + L_{\rm n}D)Q_d(B + L_{\rm n}D)'. \tag{3.21}$$

With $Z_{\rm n} = Q_\eta + DQ_dD' + CY_{\rm n}C'$, it can be verified that

$$(A + L_{\rm n}C)Y_{\rm n}C' + L_{\rm n}Q_\eta L_{\rm n}' + (B + L_{\rm n}D)Q_dD' = 0. \tag{3.22}$$

Denote $\tilde{G}(z) = \left[\begin{array}{cc} \tilde{M}_{\rm n}(z)Q_d^{1/2} & \tilde{N}_{\rm n}(z)Q_\eta^{1/2} \end{array}\right]$. There holds

$$\tilde{G}(z)\tilde{G}(z)^* = Z_{\rm n}^{-1/2}(Q_d + DQ_\eta D' + CY_{\rm n}C')Z_{\rm n}^{-1/2} = I$$

for all $|z| = 1$ in light of (3.22) from which the normalized left coprime condition in (3.20) follows. □

Lemma 5 indicates that

$$\mathcal{N}(k) = \tilde{N}_{\rm n}(q)d(k) + \tilde{M}_{\rm n}(q)\eta(k) \tag{3.23}$$

is a white process with mean zero and covariance identity. That is, the power spectral density (PSD) of $\mathcal{N}(k)$ is identity at all frequencies.

Assume that the feedback controller $K(z)$ is stabilizing. Then $K(z) = U_{\rm n}(z)V_{\rm n}(z)^{-1}$ is an RCF for some stable and proper $V_{\rm n}(z)$ and $U_{\rm n}(z)$ by taking $V(z) = V_{\rm n}(z)$ and $U(z) = U_{\rm n}(z)$

in (3.5). Note that $K(z)$ may not be an observer-based controller. In fact it can be PID or lead/lag compensator, provided that it stabilizes the feedback system in Fig. 1 or 2. Now we can choose $L = L_{\rm n}$ so that $\tilde{N}(z) = \tilde{N}_{\rm n}(z)$, and $\tilde{M}(z) = \tilde{M}_{\rm n}(z)$. The fact that $K(z)$ is stabilizing implies that $V_{\rm n}(z)$ and $U_{\rm n}(z)$ can be chosen such that

$$V_{\rm n}(z) = V_{\rm o}(z) + N(z)J(z), \quad U_{\rm n}(z) = U_{\rm o}(z) - M(z)J(z),$$

for some stable $J(z)$ in light of (3.10) by simply setting $V(z) = V_{\rm n}(z)$ and $U(z) = U_{\rm n}(z)$. Recall that $L = L_{\rm n}$ is used at present. In addition there holds

$$\tilde{M}_{\rm n}(z)V_{\rm n}(z) + \tilde{N}_{\rm n}(z)U_{\rm n}(z) = I \ \ \forall \ |z| \geq 1. \tag{3.24}$$

The right coprime factorization of $P(z) = N(z)M(z)^{-1}$ can be obtained by taking some stabilizing state feedback gain F, and thus $V_{\rm o}(z), U_{\rm o}(z)$, and $\tilde{V}_{\rm o}(z), \tilde{U}_{\rm o}(z)$ are also available using $L = L_{\rm n}$ and chosen stabilizing state feedback gain F.

Consider the first case of no attack. There holds

$$T_K(z) = \begin{bmatrix} -U_{\rm n}(z) \\ V_{\rm n}(z) \end{bmatrix} \begin{bmatrix} \tilde{N}_{\rm n}(z) & \tilde{M}_{\rm n}(z) \end{bmatrix}. \tag{3.25}$$

Stability of $K(z)$ implies that $V_{\rm n}(z)^{-1}$ is also a stable and causal transfer matrix. It follows that $w(k) = V_{\rm n}(q)\mathcal{N}(k)$ by (3.1), (3.25), and by the definition of $\mathcal{N}(k)$ in (3.23). Then $V_{\rm n}(z)^{-1}$ represents a whitening filter in the sense that the filtered signal

$$s(k) = V_{\rm n}(q)^{-1}w(k) = \mathcal{N}(k) \tag{3.26}$$

is a white process with covariance identity, in light of the normalized left coprime factorization in (3.20) for the plant model $P(z)$ and the discussion after Lemma 5. Consequently the PSD of $s(k)$ is given by $\Phi_s(\omega) = I$ for all ω.

Suppose that the replay attack takes place at $k = k_\alpha$ for the duration of $\tau_\alpha >> 1$. In this case under the hypothesis of (3.17), the noise $\eta(k)$ corrupted at the controller input consists

of two parts with $\eta_{\rm o}(k)$ from the output measurement error, and $\eta_{\rm c}(k)$ from the error induced by the communication network. In absence of replay attacks, there is no need to separate $\eta_{\rm o}(k)$ and $\eta_{\rm c}(k)$. However in presence of replay attacks, the communication error $\eta_{\rm c}(k)$ plays a crucial role in detection of the replay attack. The PSD $\Phi_w(\omega)$ for $w(k)$ is given in the following result.

Proposition 1. *Suppose that $\eta_{\rm o}(k)$ and $\eta_{\rm c}(k)$ are assumed to be independent white processes for all k. Let $\{\tilde{M}_{\rm n}(z), \tilde{N}_{\rm n}(z)\}$ in* (3.19) *be* LCF *of $P(z)$ satisfying* (3.20)*, and $\{V_{\rm n}(z), U_{\rm n}(z)\}$ be* RCF *of the stabilizing controller satisfying* (3.24)*. Under the replay attack, the* PSD *of $w(k)$ is given by*

$$\begin{aligned}\Phi_w(\omega) = V_{\rm n}(e^{j\omega}) \left[I - \tilde{M}_{\rm n}(e^{j\omega}) Q_{\eta_{\rm c}} \tilde{M}_{\rm n}(e^{j\omega})^* \right] V_{\rm n}(e^{j\omega})^* \\ + N(e^{j\omega}) \tilde{U}(e^{j\omega}) Q_{\eta_{\rm c}} \tilde{U}(e^{j\omega})^* N(e^{j\omega})^* + Q_{\eta_{\rm c}}\end{aligned}$$

where $Q_{\eta_{\rm c}}$ is the covariance of $\eta_{\rm c}(k)$.

Proof: Suppose that $k > k_\alpha$. Denote

$$\mathcal{N}_\alpha(k) = \mathcal{N}(k - \tau_\alpha) = \tilde{N}_{\rm n}(q) d(k - \tau_\alpha) + \tilde{M}_{\rm n}(q) \eta(k - \tau_\alpha). \tag{3.27}$$

Recall $\mathcal{N}(k)$ in (3.23). Over the time interval of $[k_\alpha,\ k_\alpha + \tau_\alpha)$,

$$\begin{aligned} w(k) = w_\alpha(k) &= y(k - \tau_\alpha) + \eta_{\rm o}(k - \tau_\alpha) + \eta_{\rm c}(k) \\ &= w(k - \tau_\alpha) - \eta_{\rm c}(k - \tau_\alpha) + \eta_{\rm c}(k) \\ &= V_{\rm n}(q) \mathcal{N}_\alpha(k) - \eta_{\rm c}(k - \tau_\alpha) + \eta_{\rm c}(k) \\ &= V_{\rm n}(q) \left[\tilde{N}_{\rm n}(q) d(k - \tau_\alpha) + \tilde{M}_{\rm n}(q) \eta_{\rm o}(k - \tau_\alpha) \right] \\ &\qquad - \left[I - V_{\rm n}(q) \tilde{M}_{\rm n}(q) \right] \eta_{\rm c}(k - \tau_\alpha) + \eta_{\rm c}(k). \end{aligned}$$

The Bezout identity in (3.6) can also be written as

$$\begin{bmatrix} M(z) & -U_{\rm n}(z) \\ N(z) & V_{\rm n}(z) \end{bmatrix} \begin{bmatrix} \tilde{V}(z) & \tilde{U}(z) \\ -\tilde{N}_{\rm n}(z) & \tilde{M}_{\rm n}(z) \end{bmatrix} = I_{m+p} \ \ \forall\ |z| \geq 1. \tag{3.28}$$

The above implies $I - V_{\rm n}(z)\tilde{M}_{\rm n}(z) = N(z)\tilde{U}(z)$, leading to

$$w_\alpha(k) = V_{\rm n}(q)\left[\tilde{N}_{\rm n}(q)d(k-\tau_\alpha) + \tilde{M}_{\rm n}(q)\eta_{\rm o}(k-\tau_\alpha)\right] - N(q)\tilde{U}(q)\eta_{\rm c}(k-\tau_\alpha) + \eta_{\rm c}(k). \quad (3.29)$$

The expression of $w_\alpha(k)$ is different from $w(k)$ prior to k_α. As a result $s(k) = V_{\rm n}(q)^{-1}w_\alpha(k)$ is not a white process in general. Since the four terms in (3.29) are all uncorrelated, the PSD of $w(k)$ over $[k_\alpha,\ k_\alpha + \tau_\alpha)$ can be easily obtained as in the proposition in which the normalization property (3.20) is used in obtaining the PSD expression. □

Proposition 1 can be used to derive the PSD for

$$s(k) = s_\alpha(k) = V_{\rm n}(q)^{-1}w_\alpha(k) \quad (3.30)$$

in presence of attacks. It can be easily seen that the PSD of $s(k)$ is given by

$$\begin{aligned}\Phi_s(\omega) = I &- \tilde{M}_{\rm n}(e^{j\omega})Q_{\eta_{\rm c}}\tilde{M}_{\rm n}(e^{j\omega})^* \\ &+ V_{\rm n}(e^{j\omega})^{-1}\left[N(e^{j\omega})\tilde{U}(e^{j\omega})Q_{\eta_{\rm c}}\tilde{U}(e^{j\omega})^*N(e^{j\omega})^* + Q_{\eta_{\rm c}}\right]V_{\rm n}(e^{j\omega})^{*-1}.\end{aligned} \quad (3.31)$$

Hence if $\|\Phi_s(\omega) - I\|$ is significantly greater than zero in most of the frequency range, then successful detection of presence of replay attacks will have high probability.

The above analysis shows that detection of replay attack is equivalent to detecting whether or not $s(k)$ is white, assuming that $s(k)$ deviates from white noise significantly under replay attacks. Although several schemes exist, we adopt the χ^2 detector in this chapter.

3.2.3 Computational issues

This section is focused on the computation of the transfer matrix $V_{\rm n}(z)$ that whitens the observed signal $w(k)$ by the fact that $s(k) = V_{\rm n}(q)w(k)$ is white. The computation can be complicated for a general stabilizing controller. Two different cases will be considered.

Observer-based controller

If we assume an observer-based controller, then

$$K(z) = F(zI - A - BF - LC)^{-1}L \quad (3.32)$$

where $L \neq L_{\rm n}$ in general. In this case, denote

$$K_{\rm o}(z) = F(zI - A - BF - L_{\rm n}C)^{-1}L_{\rm n} \neq K(z). \tag{3.33}$$

The LCF, and RCF of $K_{\rm o}(z)$ have realizations specified as in (3.8), and (3.9), respectively with L replaced by $L_{\rm n}$. Then realization of $V_{\rm n}(z)$ can be obtained in closed form. The general form for any stabilizing controller $K(z) = U_{\rm n}(z)V_{\rm n}(z)^{-1}$ in (3.5) gives

$$\begin{bmatrix} U_{\rm n}(z) \\ V_{\rm n}(z) \end{bmatrix} = \begin{bmatrix} U_{\rm o}(z) \\ V_{\rm o}(z) \end{bmatrix} + \begin{bmatrix} -M(z) \\ N(z) \end{bmatrix} J(z). \tag{3.34}$$

Multiplying $\begin{bmatrix} -\tilde{V}_{\rm o}(z) & \tilde{U}_{\rm o}(z) \end{bmatrix}$ from left and using $K_{\rm o}(z) = U_{\rm o}(z)V_{\rm o}(z)^{-1} = \tilde{V}_{\rm o}(z)^{-1}\tilde{U}_{\rm o}(z)$ yield

$$J(z) = \begin{bmatrix} \tilde{V}_{\rm o}(z) & \tilde{U}_{\rm o}(z) \end{bmatrix} \begin{bmatrix} -U_{\rm n}(z) \\ V_{\rm n}(z) \end{bmatrix} \tag{3.35}$$

by the Bezout identity $\tilde{V}_{\rm o}(z)M(z) + \tilde{U}_{\rm o}(z)N(z) = I$. It follows that

$$J(z) = \left[\begin{array}{c|cc} A + L_{\rm n}C & -(B + L_{\rm n}D) & L_{\rm n} \\ \hline F & I & 0 \end{array}\right] \left[\begin{array}{c|c} A + BF & -L \\ \hline F & 0 \\ C + DF & I \end{array}\right] = \left[\begin{array}{c|c} A + L_{\rm n}C & L_{\rm n} - L \\ \hline F & 0 \end{array}\right]. \tag{3.36}$$

Substituting the above back to (3.34) gives the realization of $V_{\rm n}(z)$ as

$$V_{\rm n}(z) = \left[\begin{array}{c|c} A_{v_{\rm n}} & B_{v_{\rm n}} \\ \hline C_{v_{\rm n}} & D_{v_{\rm n}} \end{array}\right] := \left[\begin{array}{cc|c} A + L_{\rm n}C & 0 & L_{\rm n} - L \\ BF & A + BF & -L_{\rm n} \\ \hline DF & C + DF & I \end{array}\right]. \tag{3.37}$$

Since PSD $\Phi_s(\omega)$ or $\Phi_w(\omega)$ also requires knowledge of $N(z)$ and $\tilde{U}(z)$ that are given next:

$$N(z) = \left[\begin{array}{c|c} A + BF & B \\ \hline C + DF & D \end{array}\right], \quad \tilde{U}(z) = \tilde{U}_{\rm o}(z) - J(z)\tilde{M}_{\rm n}(z),$$

$$\tilde{U}_{\rm o}(z) = \left[\begin{array}{c|c} A + L_{\rm n}C & L_{\rm n} \\ \hline F & 0 \end{array}\right].$$

More general types of controllers

In the case when the feedback controller is neither in observer form, nor in Kalman filter form, then computation of $V_{\rm n}(z)$ becomes more complex. Since $K(z)$ is stable for our replay attack problem, we have that

$$\tilde{R}(z) = \tilde{M}_{\rm n}(z) + \tilde{N}_{\rm n}(z)K(z)$$

is both stable and admits stable inverse. Hence RCF $K(z) = U_{\rm n}(z)V_{\rm n}(z)^{-1}$ can be obtained as

$$V_{\rm n}(z) = \tilde{R}(z)^{-1}, \quad U_{\rm n}(z) = K(z)\tilde{R}(z)^{-1},$$

satisfying the Bezout identity

$$\tilde{M}_{\rm n}(z)V_{\rm n}(z) + \tilde{N}_{\rm n}(z)U_{\rm n}(z) = I \quad \forall\, |z| \geq 1.$$

If $K(z)$ admits realization (A_K, B_K, C_K, D_K), then

$$\begin{aligned}
\tilde{R}(z) = \tilde{M}_{\rm n}(z) + \tilde{N}_{\rm n}(z)K(z) &= \left[\begin{array}{cc} \tilde{M}_{\rm n}(z) & \tilde{N}_{\rm n}(z) \end{array}\right] \left[\begin{array}{c} I \\ \\ K(z) \end{array}\right] \\
&= \left[\begin{array}{c|cc} A + L_{\rm n}C & L_{\rm n} & (B + L_{\rm n}D) \\ \hline Z_{\rm n}^{-1/2}C & Z_{\rm n}^{-1/2} & Z_{\rm n}^{-1/2}D \end{array}\right] \left[\begin{array}{c|c} A_K & B_K \\ \hline 0 & I \\ \\ C_K & D_K \end{array}\right] \\
&= \left[\begin{array}{cc|c} A_K & 0 & B_K \\ (B + L_{\rm n}D)C_K & A + L_{\rm n}C & L_{\rm n} + (B + L_{\rm n}D)D_K \\ \hline \Omega^{-1}DC_K & \Omega^{-1}C & \Omega^{-1}(I + DD_K) \end{array}\right]
\end{aligned}$$

where $\Omega = Z_{\rm n}^{1/2}$. Hence $V_{\rm n}(z) = R(z)^{-1}$ can be obtained as

$$V_{\rm n}(z) = \left[\begin{array}{cc|c} A_K - B_K(I + DD_K)^{-1}DC_K & -B_K(I + DD_K)^{-1}C & B_K(I + DD_K)^{-1}\Omega \\ (B + L_{\rm n}D)C_K - \Pi DC_K & A + L_{\rm n}C - \Pi C & \Pi\Omega \\ \hline -(I + DD_K)^{-1}DC_K & -(I + DD_K)^{-1}C & (I + DD_K)^{-1}\Omega \end{array}\right] \tag{3.38}$$

with $\Pi = [L_{\rm n} + (B + L_{\rm n}D)D_K](I + DD_K)^{-1}$. The above $V_{\rm n}(z)$ has much more complicated expression compared to the previous type of stabilizing controllers.

The computation of $\tilde{U}(z)$ is more involved. Let $P(z) = N(z)M(z)^{-1}$ be RCF with realization in (3.9) where F is a stabilizing state feedback gain. Then the stabilizing assumption for $K(z)$ implies that

$$R(z) = M(z) + K(z)N(z)$$

has a stable and proper inverse. In this case the LCF $K(z) = \tilde{V}(z)^{-1}\tilde{U}(z)$ can be obtained as

$$\tilde{V}(z) = R(z)^{-1}, \qquad \tilde{U}(z) = R(z)^{-1}K(z),$$

satisfying the Bezout identity

$$\tilde{V}(z)M(z) + \tilde{U}N(z) = I \quad \forall\, |z| \geq 1.$$

Again let (A_K, B_K, C_K, D_K) be realization of $K(z)$. Direct calculation gives

$$\begin{aligned} R(z) = M(z) + K(z)N(z) &= \left[\begin{array}{cc} I & K(z) \end{array}\right] \left[\begin{array}{c} M(z) \\ N(z) \end{array}\right] \\ &= \left[\begin{array}{c|cc} A_K & 0 & B_K \\ \hline C_k & I & D_K \end{array}\right] \left[\begin{array}{c|c} A+BF & B \\ \hline F & I \\ C+DF & 0 \end{array}\right] \\ &= \left[\begin{array}{ccc|c} A+BF & 0 & & B \\ B_K(C+DF) & A_K & & B_KD \\ \hline F+D_K(C+DF) & C_K & & I+D_KD \end{array}\right]. \end{aligned}$$

Denote $\tilde{\Pi} = (I + D_KD)^{-1}[F + D_K(C + DK)]$. The inverse of $R(z)$ is obtained as

$$\tilde{V}(z) = R(z)^{-1} = \left[\begin{array}{cc|c} A+BF-B\tilde{\Pi} & -B(I+D_KD)^{-1}C_K & B(I+D_KD)^{-1} \\ B_K(C+DF)-B_KD\tilde{\Pi} & A_K - B_KD(I+D_KD)^{-1}C_K & B_KD(I+D_KD)^{-1} \\ \hline -\tilde{\Pi} & -(I+D_KD)^{-1}C_K & (I+D_KD)^{-1} \end{array}\right].$$

Assume that $p = m$. If $p \neq m$, then additional rows or columns can be appended to $K(z)$ and $N(z)$, respectively, to satisfy $p = m$. After $\tilde{U}(z)$ is obtained, these appended rows and columns can be annihilated by taking limit to zero. The square assumption implies that

$$\tilde{U}(z) = R(z)^{-1}K(z) = [M(z) + N(z)K(z)]^{-1}K(z) = [K(z)^{-1}M(z) + N(z)]^{-1}.$$

For the same reason, D_K can be assumed to be nonsingular. Otherwise small perturbation can be introduced to D_K to satisfy the non-singularity condition. The limit to zero can then be taken for the perturbation. Hence

$$\begin{aligned}
K(z)^{-1}M(z) + N(z) &= \left[\begin{array}{cc} K(z)^{-1} & I \end{array}\right] \left[\begin{array}{c} M(z) \\ N(z) \end{array}\right] \\
&= \left[\begin{array}{cc|cc} A_K - B_K D_K^{-1} C_K & & B_K D_K^{-1} & 0 \\ & & & \\ \hline & & & \\ -D_K^{-1} C_K & & D_K^{-1} & I \end{array}\right] \left[\begin{array}{c|c} A + BF & B \\ \hline F & I \\ C + DF & 0 \end{array}\right] \\
&= \left[\begin{array}{ccc|c} A + BF & & 0 & B \\ B_K D_K^{-1} F & & A_K - B_K D_K^{-1} C_K & B_K D_K^{-1} \\ \hline C + DF + D_K^{-1} F & & -D_K^{-1} C_K & D + D_K^{-1} \end{array}\right].
\end{aligned}$$

Let $\{\tilde{A}_U, \tilde{B}_U, \tilde{C}_U, \tilde{D}_U\}$ be realization of $\tilde{U}(z) = [K(z)^{-1}M(z)+N(z)]^{-1}$. The above realization gives

$$\begin{aligned}
\tilde{D}_U &= (D + D_K^{-1})^{-1} = D_K(I + DD_K)^{-1}, \\
\tilde{C}_U &= \left[\begin{array}{cc} \tilde{C}_{U1} & \tilde{C}_{U2} \end{array}\right] := -\tilde{D}_U \left[\begin{array}{cc} C + DF + D_K^{-1} F & -D_K^{-1} C_K \end{array}\right] \\
&= -(I + D_K D)^{-1} \left[\begin{array}{cc} D_K C + (I + D_K D)F & -C_K \end{array}\right] \\
&= \left[\begin{array}{cc} -F - (I + D_K D)^{-1} D_K C & (I + D_K D)^{-1} C_K \end{array}\right], \\
\tilde{B}_U &= \left[\begin{array}{c} \tilde{B}_{U1} \\ \tilde{B}_{U2} \end{array}\right] := \left[\begin{array}{c} B \\ B_K D_K^{-1} \end{array}\right] \tilde{D}_U = \left[\begin{array}{c} BD_K(I + DD_K)^{-1} \\ B_K(I + DD_K)^{-1} \end{array}\right].
\end{aligned}$$

For $\tilde{A}_U$, partition it into 2×2 sub-blocks compatible with the above, denoted by $\{\tilde{A}_{Uij}\}$. Then

$$\begin{aligned}
\tilde{A}_{U11} &= A + BF - B\tilde{D}_U(C + DF + D_K^{-1}F) \\
&= A + BF - B[F + (I + D_K D)^{-1} D_K C] \\
&= A - B(I + D_K D)^{-1} D_K C, \\
\tilde{A}_{U12} &= B\tilde{D}_U D_K^{-1} C_K = B(I + D_K D)^{-1} C_K, \\
\tilde{A}_{U21} &= B_K D_K^{-1} F - B_K D_K^{-1} \tilde{D}_U (C + DF + D_K^{-1} F) \\
&= B_K D_K^{-1} F - B_K D_K^{-1} [F + (I + D_K D)^{-1} D_K C] \\
&= -B_K (I + D D_K)^{-1} C, \\
\tilde{A}_{U22} &= A_K - B_K D_K^{-1} C_K + B_K D_K^{-1} \tilde{D}_U D_K^{-1} C_K \\
&= A_K - B_K D_K^{-1} C_K + B_K D_K^{-1} (I + D_K D)^{-1} C_K \\
&= A_K - B_K D_K^{-1} [I - (I + D_K D)^{-1}] C_K \\
&= A_K - B_K D (I + D_K D)^{-1} C_K.
\end{aligned}$$

The above derivation yields the realization of $\tilde{U}(z)$ as

$$\tilde{U}(z) = \left[\begin{array}{cc|c}
A - B(I + D_K D)^{-1} D_K C & B(I + D_K D)^{-1} C_K & B D_K (I + D D_K)^{-1} \\
-B_K (I + D D_K)^{-1} C & A_K - B_K D (I + D_K D)^{-1} C_K & B_K (I + D D_K)^{-1} \\
\hline
-F - (I + D_K D)^{-1} D_K C & (I + D_K D)^{-1} C_K & D_K (I + D D_K)^{-1}
\end{array}\right]. \tag{3.39}$$

For completion we will also provide the expression for realization of

$$U_{\mathrm{n}}(z) = K(z)[\tilde{M}_{\mathrm{n}}(z) + \tilde{N}_{\mathrm{n}}(z) K(z)]^{-1} = [\tilde{M}_{\mathrm{n}}(z) K(z)^{-1} + \tilde{N}_{\mathrm{n}}(z)]^{-1}.$$

With $\{A_K, B_K, C_K, D_K\}$ as realization of $K(z)$, there hold

$$\begin{aligned}
\tilde{M}_{\rm n}(z)K(z)^{-1} + \tilde{N}_{\rm n}(z) &= \left[\begin{array}{cc} \tilde{M}_{\rm n}(z) & \tilde{N}_{\rm n}(z) \end{array}\right] \left[\begin{array}{c} K(z)^{-1} \\ \hline I \end{array}\right] \\
&= \left[\begin{array}{c|cc} A+L_{\rm n}C & L_{\rm n} & (B+L_{\rm n}D) \\ \hline Z_{\rm n}^{-1/2}C & Z_{\rm n}^{-1/2} & Z_{\rm n}^{-1/2}D \end{array}\right] \left[\begin{array}{c|c} A_K - B_K D_K^{-1} C_K & B_K D_K^{-1} \\ \hline -D_K^{-1}C_K & D_K^{-1} \\ 0 & I \end{array}\right] \\
&= Z_{\rm n}^{-1/2} \left[\begin{array}{cc|c} A_K - B_K D_K^{-1} C_K & 0 & B_K D_K^{-1} \\ -L_{\rm n} D_K^{-1} C_K & A+L_{\rm n}C & B+L_{\rm n}(D+D_K^{-1}) \\ \hline -D_K^{-1}C_K & C & D+D_K^{-1} \end{array}\right]
\end{aligned}$$

It follows that $U_{\rm n}(z) = [\tilde{M}_{\rm n}(z)K(z)^{-1} + \tilde{N}_{\rm n}(z)]^{-1}$ can be computed based on the above expression. For simplicity purpose, define

$$U_{\rm n}(z) = Z_{\rm n}^{1/2}\left[D_U + C_U(zI - A_U)^{-1}B_U\right].$$

Then $\{A_U, B_U, C_U, D_U\}$ does not involve $Z_{\rm n}^{1/2}$. Direct calculation yields

$$\begin{aligned}
D_U &= (D + D_K^{-1})^{-1} = D_K(I + DD_K)^{-1}, \\
C_U &= -D_U \left[\begin{array}{cc} -D_K^{-1}C_K & C \end{array}\right] \\
&= \left[\begin{array}{cc} (I+DD_K)^{-1}C_K & -D_K(I+DD_K)^{-1}C \end{array}\right], \\
B_U &= \left[\begin{array}{c} B_K D_K^{-1} \\ \hline B + L_{\rm n}(D + D_K^{-1}) \end{array}\right] D_U = \left[\begin{array}{c} B_K(I+DD_K)^{-1} \\ \hline L_{\rm n} + BD_K(I+DD_K)^{-1} \end{array}\right].
\end{aligned}$$

Again partition A_U into 2×2 blocks, denoted as $\{A_{Uij}\}$ with compatible partitions. Then

$$\begin{aligned}
A_{U11} &= A_K - B_K D_K^{-1} C_K + B_K D_K^{-1} D_U D_K^{-1} C \\
&= A_K - B_K D_K^{-1} C_K + B_K (I + D D_K)^{-1} D_K^{-1} C_K \\
&= A_K - B_K [I - (I + D D_K)^{-1}] D_K^{-1} C_K \\
&= A_K - B_K (I + D D_K)^{-1} D C_K, \\
A_{U12} &= -B_K D_K^{-1} D_U C = -B_K (I + D D_K)^{-1} C, \\
A_{U21} &= -L_{\rm n} D_K^{-1} C_K + [B + L_{\rm n}(D + D_K^{-1})] D_U D_K^{-1} C_K \\
&= B D_U D_K^{-1} C_K = B (I + D_K D)^{-1} C_K, \\
A_{U22} &= A + L_{\rm n} C - [B + L_{\rm n}(D + D_K^{-1})] D_U C \\
&= A - B D_U C = A - B D_K (I + D D_K)^{-1} C.
\end{aligned}$$

The above provides realization for $U_{\rm n}(z)$ given by

$$U_{\rm n}(z) = \left[\begin{array}{cc|c}
A_K - B_K (I + D D_K)^{-1} D C_K & -B_K (I + D D_K)^{-1} C & B_K (I + D D_K)^{-1} \\
B (I + D_K D)^{-1} C_K & A - B D_K (I + D D_K)^{-1} C & L_{\rm n} + B D_K (I + D D_K)^{-1} \\
\hline
Z_{\rm n}^{1/2} (I + D_K D)^{-1} C_K & -Z_{\rm n}^{1/2} D_K (I + D D_K)^{-1} C & Z_{\rm n}^{1/2} D_K (I + D D_K)^{-1}
\end{array}\right]$$

3.3 Simulation Results and Concluding Remarks

In the first example, we consider a typical plant model given by

$$P(z) = \frac{0.03(z+1)^2}{(z-1)(z^2 - 0.03z + 0.8)}.$$

with variance $Q_d = 0.01, Q_{\eta_o} = 0.01$ and $Q_{\eta_c} = 0.01$. Often industrial control systems use lead/lag compensators to correct the bad behaviors. In this example, we assume that the feedback controller has a lead/lag form given by

$$K(z) = \frac{2(z - 0.931)(z - 0.89)}{(z - 0.985)(z - 0.81)}.$$

In terms of the step response, the closed-loop system admits 12% for overshot, $13T_s$ for rise time, and $75T_s$ for settling time where T_s is the sampling period. In our simulation study,

we assume that the system is in the steady-state prior to the replay attack, and we assume that the replay attack is launched at time $k = 250T_s$.

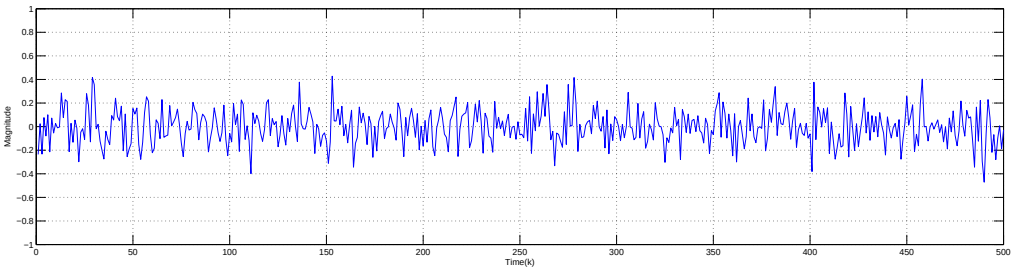

Figure 3.4: feedback signal $w(k)$ under replay attack

The above figure shows the received signal $w(k)$ at the controller site prior and after the replay attack. Although this is one simulation result, but it is very typical. Usually we do not see the difference from $w(k)$ between absence of the replay attack and presence of the replay attack. Indeed the signal $w(k)$ has no significant difference after $250T_s$ than before $250T_s$.

We also applied the whitening filter $V_{\mathrm{n}}(q)^{-1}$ to the feedback signal $w(k)$ to obtain the filtered signal $s(k) = V_{\mathrm{n}}(q)^{-1}w(k)$.

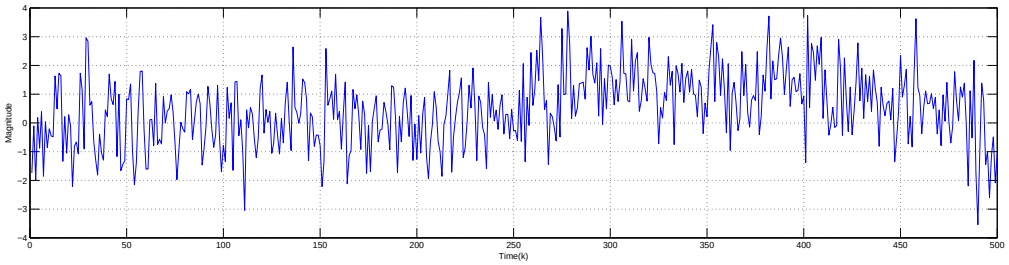

Figure 3.5: Filtered signal $s(k)$ under replay attack

The above figure shows that the filtered signal $s(k)$ shows some changes after $250T_s$, although the changes are not as significant as we want. In fact the simulation results are random. Sometimes the changes are more obvious than others. We used χ^2 detector to analysis $s(k)$ by setting up the window size to be 5 and the false alarm rate at 5%. The detection rate for this example is shown next.

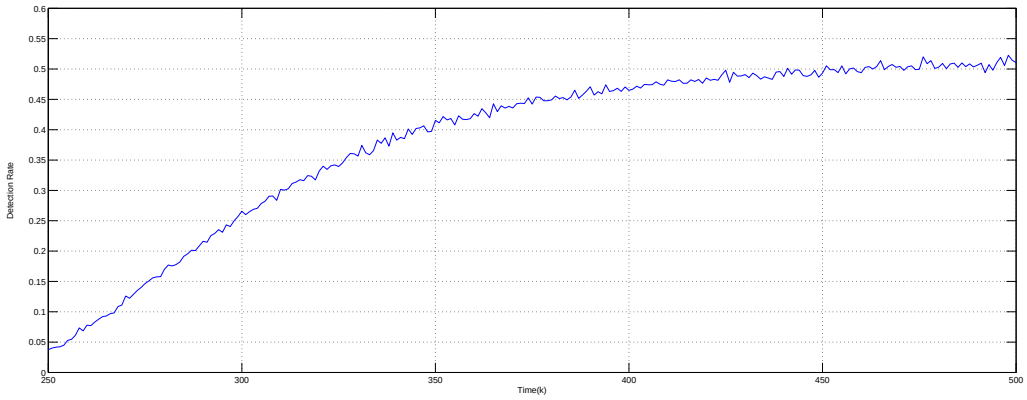

Figure 3.6: detection rate for replay attack

The detection rate is obtained by taking the average of 10000 experimental ensembles. The replay attack starts at time $k = 250T_s$. The above figure shows that the detection rate curve begins at below 5% at time $k = 250T_s$. The detection rate grows as time increases. It can reach more than 50% detection rate after a latency of about $200T_s$. Although 50% detection rate is better than the one in [2], the latency can be a problem.

A second example considers the plant model given by

$$P(z) = \frac{0.004837(z + 0.9673)}{(z^2 - 1.905z + 0.905)}.$$

with the same variance Q_d, Q_{η_o} and Q_{η_c} are used. The feedback controller is a lead compensator given by

$$K(z) = \frac{56.63(z - 0.8533)}{(z + 0.03108)}.$$

In terms of step response, the closed-loop system admits 13% for overshot, $3T_s$ for rise time, and $18T_s$ for settling time. We assume that the system is already in the steady state prior to the replay attack, and the replay attack is lunched at time $k = 50T_s$. The signal $w(k)$ and its filtered version $s(k)$ are shown in the next two figures separately.

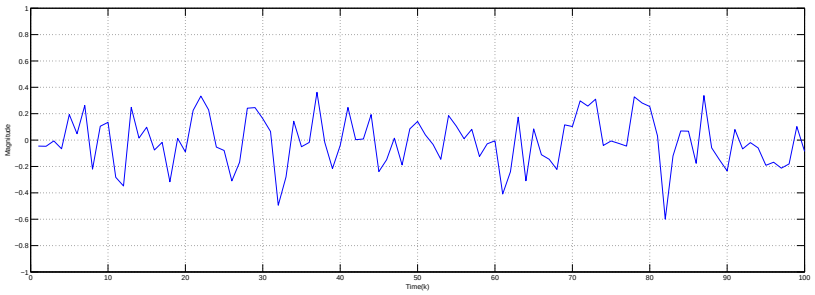

Figure 3.7: Signal $w(k)$ prior and post replay attack

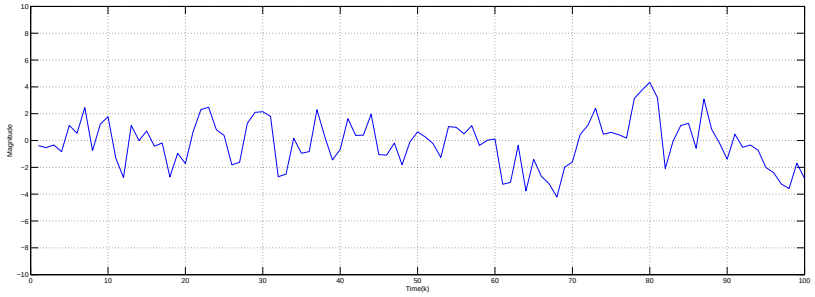

Figure 3.8: Filtered signal $s(k)$ prior and post replay attack

It shows that under the replay attack the filtered signal $s(k)$ contains more information on the replay attack than the signal $w(k)$. However both are less informative than those in the previous example, respectively. The corresponding detection rate is shown below.

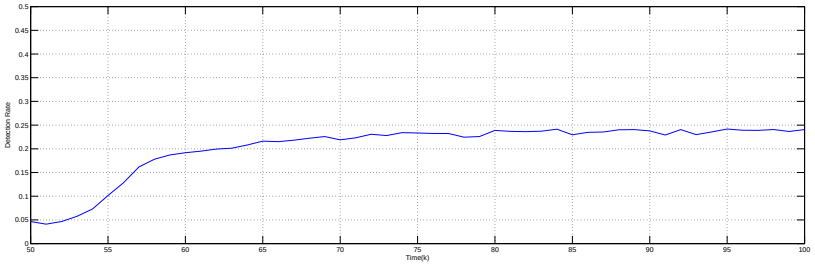

Figure 3.9: detection rate for replay attack

The detection rate is obtained by taking the average of 10000 experiments under the 5% false alarm rate. The replay attack starts at time $k = 50T_s$. Fig. 3.9 shows the line start at 5% at time $k = 50T_s$. The detection rate grows as time increases and it can reach to about

25% detection rate with a shorter latency compared to the previous example. However the detection rate is much smaller.

The two numerical examples demonstrate that the whitening filter approach is effective to detect the replay attack. However the detection performance is not very stable and varies as the underlying plant models and feedback controllers change. The problem lies in the χ^2 detector that computes the variance of the signal. Recall that the variance is the same as integral of the spectral density over all the frequencies. Because the signal $w(k)$ or $s(k)$ exhibits different behaviors at different frequencies, χ^2 detector may not be the right detector. For instance $w(k)$ or $s(k)$ can behave vary differently in some particular frequencies when the replay attack exists compared to the case when the replay attack is absent. Hence a spectral estimator at some special frequencies can serve a better detector. This problem will be studied in the next chapter.

To recap, we introduced the replay attack model and existing detection methods in this chapter. Since the existing detection methods are based on LQG controller and requires to inject noise to the system control signal, the existing χ^2 detector may not effective to detect the replay attack. For this reason, we proposed a whitening filter approach without injecting noises to the feedback control system thereby preserving the control performance. Our results show that by employing a whitening filter to the feedback signal between the plant output and controller input, the filtered signal will not be the white noise process anymore after a replay attack is launched. Our simulation examples demonstrate the effectiveness of this whitening filter approach. However latency exists and the detection rate may not be as good as what we desire especially when the false alarm rate is kept to a small number such as 5%. Hence further work and new approaches are needed, which will be studied in the next chapter.

4 Spectrum Estimation Approach

The previous approach fails if $\|\Phi_s(\omega) - I\|$ is small over majority of the frequency samples when a replay attack takes place. In this case $\Phi_s(\omega)$ is not far away from being white, and thus detection of whiteness may result in high probability of errors. On the other hand, it is possible that the PSD of $s(t)$ is far away from unity at some specific frequency samples. This fact motivates us to examine the spectral content of the signal $s(t)$ or $w(t)$ defined in the previous chapter, and consider a spectrum estimation method to tackle the detection problem for the replay attack. Our work in this chapter shows that the spectrum estimation method provides a viable alternative to detection of the replay attack.

An important assumption for the spectrum estimation method is that $\Phi_s(\omega)$ is far away from the identity at some frequency samples although it is close to being white over majority of the frequency samples. Hence detection of replay attacks can be accomplished by spectrum estimation for $\Phi_s(\omega)$ at those frequency samples it is far away from the identity after the replay attack takes place. A threshold can be set up for the resulting spectrum estimates based on the *a priori* information to determine the existence of the replay attack. However this spectrum estimation method can be very sensitive to the variation of the variances of the input disturbance, and output noises, including both the measurement and communication noises. The reason lies in the fact that the left coprime factorization such as the one in (3.20) makes use of the *a priori* knowledge of these variances. The mismatch between these true variances and those used in computing the coprime factorization in (3.20), respectively may lead to failure of the detection for replay attacks. For this reason we focus on a different

way to compute the left coprime factorization in this chapter, which avoids the use of both disturbance and noise variances in our detection algorithm.

4.1 Spectrum Estimation Approach

The spectrum estimation approach is motivated by the fact that a typical feedback control system has all singular values of $P(e^{j\omega})$ or $K(e^{j\omega})$ tending to large in the vicinity of some frequency $\omega = \omega_h$. An asymptotic case is when the gain approaches infinity at $\omega = \omega_h$. In practice the gain greater than $30 \sim 40$ dB can be regarded as adequately large.

Let us consider the first case when all singular values of $P(e^{j\omega})$ approach infinity at $\omega = \omega_h$. A different left coprime factorization of $P(z)$ from that in (3.20) is used in this section. We again assume white disturbance, white measurement noise, and white communication noises. It is known that left coprime factors of $P(z)$ can be obtained by taking a stabilizing state estimation gain. In this chapter the stabilizing state estimation gain for the chosen left coprime factors of $P(z)$, denoted by $P(z) = \tilde{M}_0(z)^{-1}\tilde{N}_0(z)$, are computed according to

$$L = L_0 := -(AY_0C' + BD')(I + DD' + CY_0C')^{-1} \tag{4.1}$$

where $Y_0 \geq 0$ is the stabilizing solution to DARE

$$Y_0 = AY_0A' - (AY_0C' + BD')(I + DD' + CY_0C')^{-1}(CY_0A' + DB') + BB'. \tag{4.2}$$

That is, realization of the left coprime factors are given by

$$\left[\begin{array}{cc} \tilde{M}_0(z) & \tilde{N}_0(z) \end{array}\right] = \left[\begin{array}{c|cc} A + L_0C & L_0 & (B + L_0D) \\ \hline Z_0^{-1/2}C & Z_0^{-1/2} & Z_0^{-1/2}D \end{array}\right], \tag{4.3}$$

with $Z_0 = I + DD' + CY_0C'$. This is also called normalized left coprime factors satisfying

$$\tilde{N}_0(z)\tilde{N}_0(z)^* + \tilde{M}_0(z)\tilde{M}_0(z)^* = I \quad \forall\ |z| = 1. \tag{4.4}$$

Note the difference between the above equality and (3.20), especially the missing of the covariances Q_d, Q_{η_c} and Q_{η_o}. For this reason, the spectrum estimation approach is more robust

than the white noise approach studied in the previous chapter. Let $K(z) = U(z)V(z)^{-1}$ be a given stabilizing controller. Then in light of the parameterization of stabilizing controllers,

$$V(z) = V_{\rm o}(z) + N(z)J(z), \quad U(z) = U_{\rm o}(z) - M(z)J(z)$$

for some stable and proper $J(z)$. It follows from $P(z) = \tilde{M}_0(z)^{-1}\tilde{N}_0(z)$ that $T_K(z)$ in (3.25) becomes

$$T_K(z) = \begin{bmatrix} -U(z) \\ V(z) \end{bmatrix} \begin{bmatrix} \tilde{N}_0(z) & \tilde{M}_0(z) \end{bmatrix}. \tag{4.5}$$

The next result provides the PSD of $w(k)$ at high-gain frequency ω_h with high-gain approaching infinity asymptotically.

Theorem 2. *Suppose that the input noise $d(t)$ is a white process with covariance $Q_d = \sigma_d^2 I$, and the* AWGN *noise have covariance $Q_{\eta_c} = \sigma_{\eta_c}^2 I$. Assume that $P(e^{j\omega})$ has infinity gain at $\omega = \omega_h$. Let $\Phi_w(\omega)$ be the* PSD *for $w(t)$. In absence of replay attacks,*

$$\Phi_w(\omega_h) = \sigma_d^2 V(e^{j\omega_h})V(e^{j\omega_h})^*. \tag{4.6}$$

In presence of replay attacks,

$$\Phi_w(\omega_h) = \sigma_d^2 V(e^{j\omega_h})V(e^{j\omega_h})^* + 2\sigma_{\eta_c}^2 I. \tag{4.7}$$

Proof: Under the assumption that all singular values of $P(e^{j\omega})$ tend to infinity as $\omega \to \omega_h$,

$$\tilde{M}_0(e^{j\omega})\tilde{M}_0(e^{j\omega})^* \;\to\; 0, \quad \tilde{N}_0(e^{j\omega})\tilde{N}_0(e^{j\omega})^* \;\to\; I, \tag{4.8}$$

as $\omega \to \omega_h$ in light of (4.4). It follows that $w(k)$ now has an expression

$$w(k) = V(q)\left[\tilde{N}_0(q)d(k) + \tilde{M}_0(q)\eta(k)\right] \tag{4.9}$$

in absence of attacks. There thus holds $V(e^{j\omega})\tilde{M}_0(e^{j\omega}) \to 0$ as $\omega \to \omega_h$. We can now conclude that the PSD of $w(k)$ at $\omega = \omega_h$ is given by

$$\Phi_w(\omega_h) = V(e^{j\omega_h})\tilde{N}_0(e^{j\omega_h})Q_d\tilde{N}_0(e^{j\omega_h})^*V(e^{j\omega_h})^* = \sigma_d^2 V(e^{j\omega_h})V(e^{j\omega_h})^* \tag{4.10}$$

if $Q_d = \sigma_d^2 I$ that verifies (4.6). In addition the identity (3.28) holds for the case when $\tilde{M}_{\rm n}(e^{j\omega})$ and $\tilde{N}_{\rm n}(e^{j\omega})$ are replaced by $\tilde{M}_0(e^{j\omega})$ and $\tilde{N}_0(e^{j\omega})$, respectively, and $V_{\rm n}(z), U_{\rm n}(z)$, are replaced by $V(z)$ and $U(z)$, respectively. As a result,

$$I - V(z)\tilde{M}_0(z) = N(z)\tilde{U}(z) \ \ \forall\ |z| \geq 1 \ \ \Longrightarrow \ \ N(e^{j\omega_h})\tilde{U}(e^{j\omega_h}) = I.$$

If the replay attack is present, then $\Phi_w(\omega_h)$ changes its value. Specifically (3.29) is modified into

$$w_\alpha(k) = V(q)\left[\tilde{N}_0(q)d(k-\tau_\alpha) + \tilde{M}_0(q)\eta_{\rm o}(k-\tau_\alpha)\right] - N(q)\tilde{U}(q)\eta_{\rm c}(k-\tau_\alpha) + \eta_{\rm c}(k). \quad (4.11)$$

The results in (4.8) and (4.10) then lead to

$$\Phi_w(\omega_h) = \sigma_d^2 V(e^{j\omega_h})V(e^{j\omega_h})^* + 2Q_{\eta_{\rm c}}$$

that verifies the expression of $\Phi_w(\omega_h)$ in (4.7), if in addition $Q_{\eta_c} = \sigma_{\eta_c}^2 I$. □

Corollary 2. *Under the same hypotheses of* Theorem 2 *and let* $s(k) = V(q)^{-1}w(k)$*, then the* PSD *of* $s(k)$ *at* $\omega = \omega_h$ *is given by*

$$\Phi_s(\omega_h) = \sigma_d^2 I. \quad (4.12)$$

in absence of attacks. In presence of replay attacks, there holds

$$\Phi_{s_\alpha}(\omega_h) = \sigma_d^2 + 2\sigma_{\eta_c}^2 V(e^{j\omega_h})^{-1}V(e^{j\omega_h})^{*-1}. \quad (4.13)$$

Since the proof follows from Theorem 2, it is omitted. It is important to observe that entries of $V(e^{j\omega_h})$ do not have large value. Indeed

$$\tilde{N}_0(z)U(z) + \tilde{M}_0(z)V(z) = I \ \ \Longrightarrow \ \ \tilde{N}_0(e^{j\omega_h})U(e^{j\omega_h}) = I$$

by the argument earlier. Since $\tilde{N}_0(e^{j\omega_h})\tilde{N}_0(e^{j\omega_h})^* = I$ by (4.8), $U(e^{j\omega_h})$ is a unitary matrix as well, if the plant has an equal number of inputs and outputs. The above implies that singular values of the controller $K(e^{j\omega_h})$ are the same as those of $V(e^{j\omega_h})^{-1}$ by

$$\sigma_i[K(e^{j\omega_h})] = \sigma_i[V(e^{j\omega_h})^{-1}] \ \ \forall\ i.$$

Therefore $\Phi_{s_\alpha}(\omega_h)$ differs from $\Phi_s(\omega_h)$ significantly provided that $K(e^{j\omega})$ has high gain or equivalently $V(e^{j\omega})$ has small gain at $\omega = \omega_h$.

Remark 2. If the control systems are designed based on the classic Bode method or based on $\mathcal{H}_\infty$ loop-shaping, then $K(e^{j\omega}) = U(e^{j\omega})V(e^{j\omega})^{-1}$ cannot have small gain over those frequencies at which the plant $P(e^{j\omega})$ has high gains. For instance many practical control systems are required to have infinity gain at $\omega_h = 0$ in order to have zero steady-state error for tracking step inputs, and a lag compensator is often employed as part of the feedback controller. It follows that $K(e^{j\omega_h})$ has large gain. In the case when $\omega_h \neq 0$, a notch filter is often employed as part of the feedback controller to boost the gain at frequency ω_h. Since $U(e^{j\omega_h})$ is unitary, $V(e^{j\omega_h})V(e^{j\omega_h})^*$ has very small gain. This shows that $\Phi_w(\omega_h)$ and $\Phi_s(\omega_h)$ under replay attacks can be significantly different from those in absence of replay attacks. □

A dual case is when $K(e^{j\omega_h})$ has large gain or asymptotically approaches infinity. A similar derivation can be carried out to conclude the same results in Theorem 2 and Corollary 2, provided that $P(e^{j\omega_h})$ has a relatively large gain. The detail is omitted.

4.2 Detection Algorithms

This section is focused on detection algorithms for replay attacks using the spectrum estimation approach. For this reason, several popular spectrum estimation algorithms [81] will be reviewed first prior to developing the detection algorithm for replay attacks.

4.2.1 Spectrum estimation methods

Nonparametric estimation

For a given signal samples $\{s(k)\}_{k=0}^{N-1}$, the simplest method to estimate its PSD at frequency ω_h is to set

$$\hat{\Phi}_s(\omega_h) = \frac{1}{N}\left|\sum_{k=1}^{N} s(k)e^{-j\omega_h k}\right|^2.$$

Let $\hat{R}_s(\tau)$ be the estimated autocorrelation sequence (ACS) defined by

$$\hat{R}_s(\tau) = \frac{1}{N} \sum_{k=\tau+1}^{N} s(k)s(k-\tau)', \quad 0 \le \tau < N. \tag{4.14}$$

The above estimate is biased, but satisfies the relation [81]

$$\hat{\Phi}_s(\omega_h) = \sum_{\tau=-(N-1)}^{N-1} \hat{R}_s(\tau) e^{-j\omega_h \tau}.$$

Let $R_s(\tau) = \mathrm{E}\{s(k)s(k-\tau)'\}$. Then there holds

$$\mathrm{E}\{\hat{\Phi}_s(\omega_h)\} = \sum_{\tau=-(N-1)}^{N-1} \left(1 - \frac{|\tau|}{N}\right) R_s(\tau) e^{-j\omega_h \tau}.$$

So the biased ACS estimate has a windowing effect on average.

The unbiased ACS estimate has a different form given by

$$\hat{R}_s(\tau) = \frac{1}{N-\tau} \sum_{k=\tau+1}^{N} s(k)s(k-\tau)', \quad 0 \le \tau < N. \tag{4.15}$$

The use of unbiased ACS estimate suggests a more sophisticated estimation method that introduces the windowing technique by taking the PSD estimate at frequency ω_h as

$$\hat{\Phi}_s(\omega_h) = \sum_{\tau=-(N-1)}^{N-1} \mathrm{win}(\tau) \hat{R}_s(\tau) e^{-j\omega_h \tau}. \tag{4.16}$$

The above covers the case of the biased ACS estimate in a special case in the average sense. Commonly used windows include Barlett, Hanning, Hamming, Blackman, and Kaiser windows. Each has its advantages and disadvantages, which will not be discussed here.

Capon algorithm

Capon algorithm is more sophisticated than the nonparametric method for spectrum estimation in that a specific optimization is carried out in the estimation process. Assume that the value of $\Phi_w(\omega_h)$ in (4.6) is very different from that in (4.7). Quite a few spectrum estimation methods can be employed to estimate $\Phi_w(\omega_h)$. The one attracting our attention is the Capon algorithm [82]. Our description in this subsection is based on the book [81] (page 232 - 238).

For convenience assume that $s(k)$ has dimension 1 at each time index k. Denote

$$a(\omega) = \begin{bmatrix} 1 \\ e^{-j\omega} \\ \vdots \\ e^{-j\ell\omega} \end{bmatrix}, \quad \underline{s}(k) = \begin{bmatrix} s(k) \\ s(k-1) \\ \vdots \\ s(k-\ell) \end{bmatrix}, \quad h = \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_\ell \end{bmatrix}$$

where $(\ell+1)$ is the window size. Let $R_s = \mathrm{E}\{\underline{s}(k)\underline{w}(k)'\}$. Then the Capon algorithm seeks to synthesize the estimate of PSD $\Phi_s(\omega)$ through solving the following optimization problem:

$$\min_h \; h'R_s h \quad \text{subject to} \quad h'a(\omega) = 1. \tag{4.17}$$

It can be easily seen that

$$s_F(k) = \sum_{i=0}^{\ell} h_i s(k-i) = h'\underline{s}(k).$$

It admits the mean power $\mathrm{E}\{|s_F(k)|^2\} = h'R_s h$. The constrained minimization in (4.17) seeks to minimize the total power subject to the constraint that the filter pass the frequency ω undistorted.

For its application to our problem, $\omega = \omega_h$ and R_s can be estimated via

$$R_s \approx R_{k_0} := \frac{1}{N+1} \sum_{k=k_0-N}^{k_0} \underline{s}(k)\underline{s}(k)'$$

at time k where $N > \ell$. That is, the sample size for estimating covariance R_s needs to be greater than the window size. Two versions of the Capon estimator (CM) are given by

$$\mathrm{CM}_1: \quad \widehat{\Phi}_s(\omega_h) = \frac{\ell+1}{a(\omega_h)^* R_{k_0}^{-1} a(\omega_h)},$$
$$\mathrm{CM}_2: \quad \widehat{\Phi}_s(\omega_h) = \frac{a(\omega_h)^* R_{k_0}^{-1} a(\omega_h)}{a(\omega_h)^* R_{k_0}^{-2} a(\omega_h)}.$$

Note that the covariance matrix R can also be calculated based on the *a priori* information on $d(k)$, $\eta(k)$, and the models of $P(z)$ and $K(z)$. Extension to the vector case or $m > 1$ can take each entry of $a(\omega)$ multiple of identity of dimension m and replacing the inverse in CM_i by matrix inverse.

A disadvantage of the Capon algorithm is that the spectrum estimation is a nonlinear function of the signal $w(k)$, and thus distribution of $w(k)$ is more difficult to analyze. However it often outperforms the nonparametric method due to its optimization nature. More importantly it has very close relation to other widely used methods such as autoregressive (AR) method (CM_1). Its weakness in resolution is not of great concern to us because the frequency ω_h is already known to us.

4.2.2 Detection algorithm

As discussed above, at frequency w_h, the PSD can be significantly different from that in absence of replay attack. So, it is possible to improve the detect rate of the replay attack by analyze the PSD at frequency w_h. As many practical control systems are required to have infinity gain at $w_h = 0$ in order to have zero steady-state error for tracking step inputs, we will focus on analyzing the PSD at zero frequency $w_h = 0$. Thus, for the given signal samples $\{s(k)\}$, the simplest way to estimate the PSD at zero frequency with a specified window size $(\ell + 1)$ at time k becomes

$$\hat{\Phi}_s(0;k) = \frac{1}{\ell+1}\left|\sum_{i=k-\ell}^{k} s(i)\right|^2 = \left|\frac{1}{\sqrt{\ell+1}}\sum_{i=k-\ell}^{k} s(i)\right|^2. \tag{4.18}$$

Because all exogenous signals $\{d(k), \eta_{\mathrm{o}}(k), \eta_{\mathrm{c}}(k)\}$ are AWGN, the random variable

$$\overline{s}_\ell(k) = \frac{1}{\sqrt{\ell+1}}\sum_{i=k-\ell}^{k} s(i)$$

is also Gauss distributed with the same mean and covariance as $s(k)$. Hence the PSD estimate for $\omega_h = 0$ at time index $(k + N)$ is given by

$$\hat{\Phi}_s(0;k) = |\overline{s}_\ell(k)|^2$$

and has a Rayleigh distribution. Let the hypothesis H_0 be the case in absence of the replay attack, and the hypothesis H_1 be the case in presence of the replay attack. Then the detection

algorithm is given by

$$H_0: \quad \hat{\Phi}_s(0) < \tau,$$

$$H_1: \quad \hat{\Phi}_s(0) > \tau,$$

for some threshold $\tau > 0$. If the variance of $s(k)$ is known for both the case of the absence and presence of the replay attack, then τ can be determined to satisfy the given false alarm rate.

To improve the detection rate, the Capon algorithm can be implemented to estimate the PSD of $\{s(k)\}$. Thus at frequency $w_h = 0$, we have

$$a(0) = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \qquad \underline{s}(k) = \begin{bmatrix} s(k) \\ s(k-1) \\ \vdots \\ s(k-\ell) \end{bmatrix}, \qquad R_k := \frac{1}{N+1} \sum_{i=k-N}^{k} \underline{s}(i)\underline{s}(i)'$$

where $N \geq \ell$. In this case the Capon estimator (CM) becomes

$$\text{CM}_1: \qquad \widehat{\Phi}_s(0) = \frac{\ell + 1}{a(0)^* R_{k_0}^{-1} a(0)}, \tag{4.19}$$

$$\text{CM}_2: \qquad \widehat{\Phi}_w(0) = \frac{a(0)^* R_{k_0}^{-1} a(0)}{a(0)^* R_{k_0}^{-2} a(0)}. \tag{4.20}$$

Our simulation results in the next section show that the Capon estimator indeed improves the detection rate for replay attacks.

4.3 Simulation and Conclusion

In this section we present some simulation results for detection of the replay attack using the spectrum estimation method. The control system under consideration is the same as the first system model in Chapter 3. Signal $s(k)$ is used in this simulation study. We use the conventional method (4.18) and the Capon method (4.19) to estimate the PSD of $s(k)$ at $\omega_h = 0$. We again assume that the feedback system is in the steady-state prior to the replay

attack, and the replay attack is launched at $k = 250T_s$. For the conventional method, fix the false alarm rate at 5%, and use different window size by set $(\ell + 1) = 5$, 10, 20 and 50. The resulting detection rate is shown next.

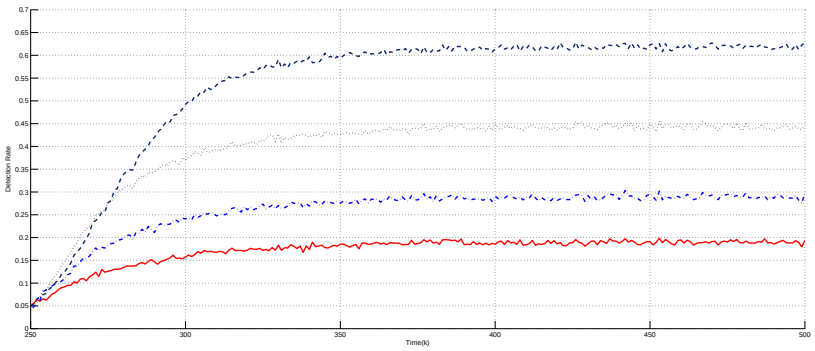

Figure 4.1: Detection rate by using conventional method with different window size

In the above figure, we set the window sizes equal to 5, 10, 20 and 50, respectively. The vertical axis represents the detection rate while the horizontal axis giving the time samples. The red solid line shows the detection rate when the window size is equal to 5; The blue dash dot line shows the detection rate when the window size is 10; The black dot line shows the detection rate when the window size is equal to 20; The purple dash line shows the detection rate when the window size is equal to 50.

The detection rate is obtained by taking the average of 10,000 ensembles under the false alarm rate of 5%. The detection rate varies for different window sizes. It grows as the window size increases. For each window size, the detection rate grows as time increases. The detection rate curve start at $250T_s$ with 5% detection rate. The detection latency is about $200T_s$ that seems to be constant with respect to the window size. When the window size is equal to 5, the curve reaches 19% detection rate; When the window size is equal to 10, the detection rate increases to 29%; When the window size is equal to 20, the detection rate reaches 43%; When the window size is equal to 50, the detection rate reaches 63%.

To improve the detection rate and shorten the latency, we scale the system by set $P(z) = P(z)/5$ and $K(z) = 5K(z)$ separately. This is commonly used in engineering practice and does not change the system performance. The detection rate for the scaled system is shown below.

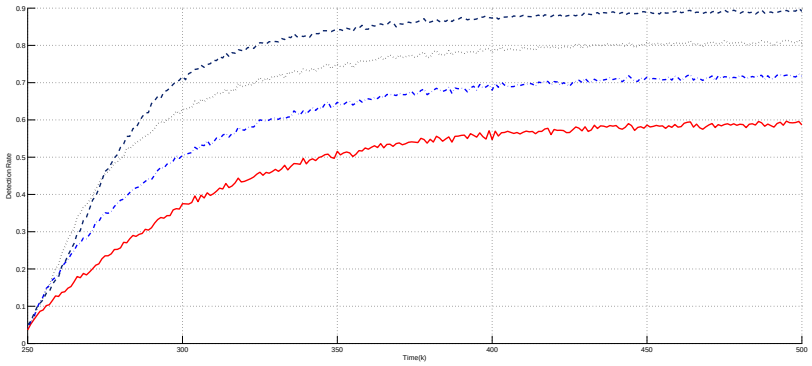

Figure 4.2: Detection rate of scaled system by using conventional method

In the above figure, we again set the window sizes equal to 5, 10, 20 and 50, respectively. The vertical axis represent the detection rate while the horizontal axis giving the time samples. The red solid line shows the detection rate when the window size is equal to 5; The blue dash dot line shows the detection rate when the window size is 10; The black dot line shows the detection rate when the window size is equal to 20; The purple dash line shows the detection rate when the window size is equal to 50.

The detection rate is again obtained by taking the average of 10,000 ensembles under the false alarm rate of 5%. For different window sizes, the detection rate curves vary. The detection rate curves start at $250T_s$ with 5% detection rate. The detection latency seems to have increased more than $100T_s$ for all the cases. When the window size is equal to 5, the detection rate reaches 57%; When the window size is equal to 10, the detection rate reaches 71%; When the window size is equal to 20, the detection rate reaches 80%; When the window size is equal to 50, the detection rate reaches 89%. Compare with the unscaled system model, the detection rate is significantly improved.

In comparison with the whitening filter approach with the window size equal to 5, the detection rate is 14% higher than the result obatined using the whitening filter approach (Fig. 3.6) under the same detection latency. It also shows that the 40% detection rate is reached with latency of $65T_s$, which is 25% less than $100T_s$ using the whitening filter approach (Fig. 3.6).

We also used Capon algorithm for detection of the replay attack, which outperforms the conventional non-parametric method as presented above. Our simulation result shows that by using the scaled system model, it is possible to increase the detection rate and shorten the latency further. Our simulation study is carried out for using the Capon algorithm with the same fixed window size $(\ell+1) = 5$ under the same false alarm rate of 5%. The detection rate using the Capon method is shown next.

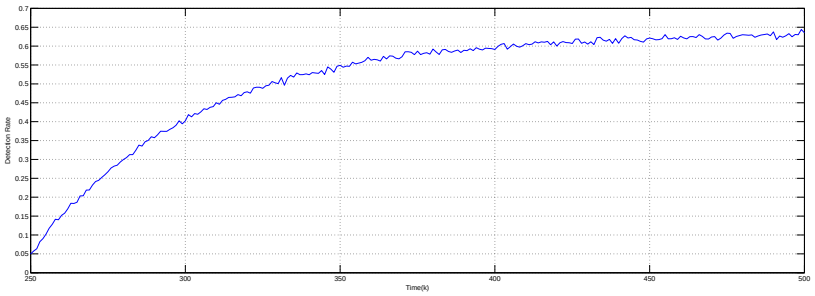

Figure 4.3: Detection rate by using capon method

The detection rate is obtained by taking the average of 10,000 ensembles under the false alarm rate of 5%. The detection rate curve starts with 5% at time = $250T_s$ at which the replay attack is launched. After about latency of $150T_s$, the detection rate reaches more than 60% that is 5% higher than that of the conventional method (Fig. 4.2), and 22% higher than the whitening filter approach (Fig. 3.6). When the same detection rate reaches 40%, the latency is $50T_s$, which is 23% less than $65T_s$ by using conventional method (Fig. 4.2) and 50% less than $100T_s$ by using the whitening filter approach (Fig. 3.6).

To compare our detection performance with the results in the existing literature [2], we use both conventional and Capon methods to test the model in [2]. For the conventional method, we assume that the replay attack is launched at $250T_s$, and set up the window sizes equal to 5, 10, 20 and 50, respectively. The simulation results of the detection rate are shown next.

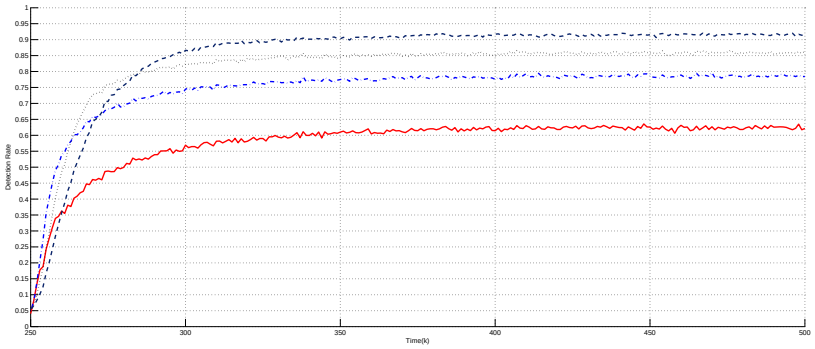

Figure 4.4: Detection rate of model [2] by using conventional method

In the above figure, the vertical axis represents the detection rate while the horizontal axis giving the time samples. The red solid line shows the detection rate when the window size is equal to 5; The blue dash dot line shows the detection rate when the window size is 10; The black dot line shows the detection rate when the window size is equal to 20; The purple dash line shows the detection rate when the window size is equal to 50. The detection rate is obtained by taking the average of 10,000 ensembles under the false alarm rate of 5%. The curve starts at $250T_s$ with 5% detection rate. Let us examine the case of latency = $200T_s$ next. When the window size is equal to 5, the detection rate reaches 63%; When the window size is equal to 10, the detection rate reaches 78%; When the window size is equal to 20, the detection rate reaches 86%; When the window size is equal to 50, the detection rate reaches 92%. For the case when the window size is equal to 5, the detection rate is almost 50% higher than that obtained by using Mo and Sinopoli's method [2]. To improve the detection rate,

The Capon method is used for replay attack detection with window size $(\ell + 1) = 5$. The detection rate curve is shown next.

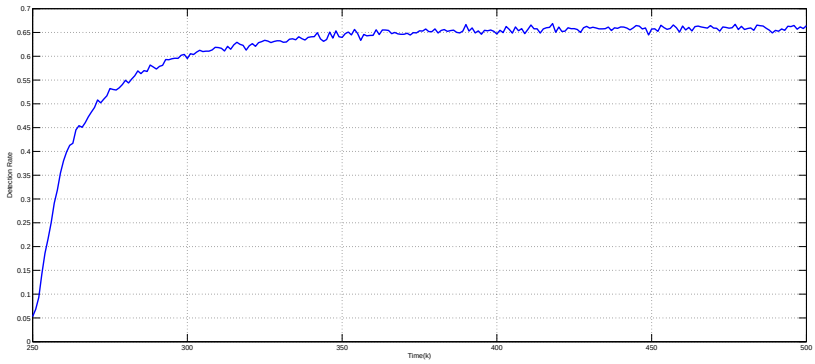

Figure 4.5: Detection rate of model [2] by using Capon method

The detection rate is obtained by taking the average of 10000 examples. The false alarm rate is set to 5%, and replay attack is launched at $250T_s$. The detection rate is 66%, which is 5% more than that using the conventional method (figure 4.4). At 40% detection rate, the latency is $11T_s$, which is 21% smaller than the case of using the conventional method (figure 4.4). In comparison with the result in [2], it shows that the latency by using the spectrum estimation approach is greater. The reason lies in the fact the filter $V(q)^{-1}$ in 4.13 is IIR filter which induces large time delay when filtering the signal. However the saturated detection rate is is much higher than that in [2].

The results for the above example show that the spectrum estimation approach at some special frequency is more effective than the whitening filter approach for replay attack detection. It significantly improves the detection rate and also shortens the latency. By using the Capon estimator, the detection rate improves more than 30% compared to the result of the whitening filter approach. Also the latency is reduced by half. Thus the spectrum estimation approach is an effective way to improve the detection performance for replay attacks.

To recap this chapter, we introduced the spectrum estimation approach and proved that at some specific frequency the PSD of filtered signal $s(k)$ under replay attack is significantly

different from that in absence of replay attacks. Motivated by this fact, both nonparametric estimation method and the Capon method are employed to estimate the PSD for the detection of the replay attacks. Simulation results are presented to show that these two estimation methods can effectively improve the detection rate and shorten the latency for replay attack detection.

5 Conclusion

SCADA system is widely used in many different industries. It is also an essential part and plays a crucial role in monitoring and controlling the power grid. As the expansion of the power grid, efficiency issue and high cost issue arise which require the power grid to evolution. With the advent of the SCADA technology, it now is feasible to build the next generation of power grid – smart grid. With the implement of the smart grid, it will improve the power supply efficiency and reduce power lost, and make the power system more reliable and greener. However due to many sites over large distance and use of information technology, especially wireless communications and networking, SCADA systems give rise to the security problem making the smart grid vulnerable to penetration and malicious attacks that become a dominant research problem in developing the smart grid. Once the adversaries gain access to the power control network, they can perform a wide range of attacks. Possible attack actions can be separated into two groups and thus pose two major challenges. One is to take control of the target region system for future crime actions. It is based on cyber system and can be solved by the cyber security approach. The other one is to modify data or inject malicious data into control systems overseen by the SCADAs. Such attacks can bypass the cyber security, and poses more serious danger to the smart grid. This dissertation is focused on the second type of attacks comprising in particular the unobservable attack, and the replay attack. Although both are studied in the existing literature, the solutions are not satisfactory which motivates this dissertation research, as presented in the previous three chapters. Our results provide viable solution approaches to both the unobservable attack and

the replay attack, validated by our simulation studies. This conclusion chapter summarizes the contribution of this dissertation and in addition discusses the possible future research directions on the security problem of the smart grid.

5.1 Dissertation Contributions

This dissertation is aimed at solving the detection problems for two major threads to the smart grid: the unobservable attack and the replay attack, introduced and dwelled in Chapter 1. Our contributions to these two detection problems are summarized next:

- In Chapter 2, we propose a consensus dynamic estimation approach. We assume that the power system is described by a linear DC power flow model and is operating in a quasi-steady state. Each bus load is considered as an individual agent with different state-space model and state vector dimension. The simplest case is when the dynamic model is described by the random walk Markov process. Then we formulate the detection problem into the dynamic estimation rather than static estimation problem, which is in fact a distributed estimation problem in the sense of consensus. The reason lies in the fact that the existing conventional detection methods are based on static state estimation which is only related to the present measurement information. It can be bypassed by unobservable attacks. In contrast the dynamic state estimation is related to both the present measurement information and the past system states. It can help to validate the data as the states are predicted one more time against the measurement samples. The latest results from the consensus control and the positive real concept are extended to discrete-time multi-agent systems, and a distributed estimation algorithm is derived distributed estimation algorithm in this dissertation. The distributed consensus estimation is developed based on the local dynamic model for each bus. So it is possible to monitor the dynamic status in each bus. Contrasting to conventional estimation method, the consensus method can not only detect the existence of

unobservable attacks but also detect which agent data is modified by adversary. Therefore by using the consensus dynamic approach, unobservable attacks can be effectively detected, which are validated by our extensive simulation studies.

- This dissertation proposes two different new approaches to tackle the detection problem for replay attacks. Contrasting to the existing methods that employ injection of white Gauss noises to the control signal and deteriorate the control system performance, we make use of the communication noise induced by the digital network and wireless channels commonly seen in SCADA systems. Two different solution approaches are proposed without injecting the white Gauss noises to the control signal, yet effective for detecting the replay attacks. Our contributions are described next.

  - In Chapter 3, we propose a whitening filter approach. Specifically the received signal transmitted from the plant model to the controller has a smaller noise variance under the normal operation in absence of the replay attack than in presence of the replay attack. The reason lies in the fact that the replayed signal contains the channel noise in the past uncorrelated to the present channel noise. To highlight this variance difference, a filter is employed to whiten the feedback signal received at the controller input, assuming no replay attack. The filtered signal becomes nonwhite with a greater variance when the replay attack is launched. Hence we are able to detect the replay attack based on the whitening feedback signal without sacrificing the control system performance compared to the existing results in the literature. In addition our approach is applicable to any stable feedback control systems, because there is no assumed Kalman filter in the feedback controller as in the LQG control system although our results apply to the LQG control system as well. The simulations shows that this solution approach

is effective but it comes with large latency due to the use of the whitening filter that is IIR in general.

Although the whitening filter approach works, it may fail if the PSD of filtered signal is close to 1 over majority of the frequency samples when a replay attack takes place. In this case the PSD of filtered signal is not far away from being white, and may result in high probability of detection errors. On the other hand, it is possible that the PSD of filtered signal is far away from unity at some specific frequency samples. This fact motivates us to propose another detection method as described next.

– In Chapter 4, we propose the spectrum estimation approach. First we analyze the PSD of the feedback signal and its filtered version, and find some particular frequency samples at which its PSD is small under the normal operation in absence of the replay attack but become large in presence of the replay attack. In fact the PSD of the filtered signal at these frequency samples is unity in absence of the replay attack and becomes significantly greater than the unity in presence of the replay attack. The set of these frequency samples is not empty due to the stability and performance requirement for the underlying control system. By estimating the PSD of the filtered signal at these frequency samples, we are able to detect the presence of the replay attack by setting the appropriate threshold determined by the false alarm rate. Both nonparametric estimation method and the Capon method are employed to estimate the PSD at these frequency samples. Our simulation results show that the spectrum estimation approach at some specific frequencies is more effective than the whitening filter approach in terms of the detection rate for replay attacks. Both these two methods can effectively improve the detection rate and shorten the latency for detection of the replay attack.

5.2 Future Direction

In this dissertation, we propose new solution approaches to the unobservable attack and replay attack threatening the SCADA systems. These two approaches are shown effective, validated by the simulation studies. However the work on security of the smart grid and SCADA systems is far away from over. In the following, we highlight our point of views for future research in security against the unobservable attack and replay attack.

Unobservable Attacks

Although our proposed solution approach is effective, there are still some inadequacies that need to addressed in the future.

- This dissertation has assumed that the phase angles in the power system are approximately decoupled in order to simply the problem. It is important to note that the phase angles may not be decoupled, and the cross product terms may not be negligible in practice. While we believe that the coupling among different phase angles are weak during the normal operation of the power grid, further qualitative and quantitative studies are necessary. If our intuition is correct, some analysis is needed to show the approximate decoupling among the phase angles. Otherwise we need to find a remedy to our assumption on the approximate decoupling.

- This dissertation has proposed consensus estimation used to detect the unobservable attack and shown to be effective. However there is no optimal consensus estimation is not studied. Since positive real property in the continuous-time case corresponds to optimal estimation, it will be interesting to find out if there is a similar connection. In the least current consensus estimation algorithm ensures only stability of the estimator, which may be inadequate in many other applications. This can be a good future research topic.

Replay Attacks

For the replay attack, two effective approaches are proposed in this dissertation to detect the replay attack. However the research work in this problem area is far from over due to the imminent threat of the replay attack. We point out a few future directions in this problem area next.

- Although additive white Gauss noise (AWGN) channels are widely used in wireless communications, other network distortions also exist in the networked control systems common in SCADAs. This dissertation has not studied other network distortions different than the AWGN channel. For instance, we are not sure if the packet drops and quantization errors at the plant input and output can be effectively used for detection of the replay attack. Both are difficult to analyze statistically, and are lack of the nice features in the AWGN channel.

- Our two solution approaches reply on filtering of the feedback signal. Because the filter has infinite impulse response (IIR) in general, it induces the latency causing the time delay in detecting the replay attack. How to reduce the detection latency poses a significant challenge. At present we are not sure if more advanced PSD estimation methods can help to improve detection rate and shorten the latency.

The security problem in smart grid with SCADA system is complex and challenging. Our proposed approaches are just only one of the initial attempts to tackle the attack detection problems. We hope that our contributions in this dissertation speed up the research activity and solving the security issues in smart grid security issues.

References

[1] F. Miao, M. Pajic, and G. J. Pappas, "Stochastic game approach for replay attack detection," in *Decision and Control,2013 IEEE 52nd Annual Conference on Digital Object Identifier*, pp. 1854–1859, 2013.

[2] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, pp. 911–918, IEEE, 2009.

[3] P. King, "SCADA systems - looking ahead," tech. rep., Control Microsystems Inc., August 2005.

[4] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on the smart grid," *Smart Grid, IEEE Transactions on*, vol. 2, no. 4, pp. 645–658, 2011.

[5] L. Xie, Y. Mo, and B. Sinopoli, "False data injection attacks in electricity markets," in *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pp. 226–231, IEEE, 2010.

[6] Y. Zhang, L. Wang, W. Sun, R. C. Green, and M. Alam, "Distributed intrusion detection system in a multi-layer network architecture of smart grids," *Smart Grid, IEEE Transactions on*, vol. 2, no. 4, pp. 796–808, 2011.

[7] S. Roy, M. Xue, and S. K. Das, "Security and discoverability of spread dynamics in cyber-physical networks," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 23, no. 9, pp. 1694–1707, 2012.

[8] M. Blanc, J. Briffaut, P. Clemente, M. G. El Rab, and C. Toinard, "A collaborative approach for access control, intrusion detection and security testing," in *Collaborative Technologies and Systems, 2006. CTS 2006. International Symposium on*, pp. 270–277, IEEE, 2006.

[9] Y. Mo, T.-H. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli, "Cyber–physical security of a smart grid infrastructure," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 195–209, 2012.

[10] J. Berst, "Electronomics: Why we need smart grid technology and infrastructure today," *Xconomy*, February 2009.

[11] B. Chameides, "The new smart grid: 21st century tech for the 21st century," *Popular Science*, February 2009.

[12] D. Coalition, "Smart meters, demand response, & lowincome customers," *Demand Response and Smart Grid Coalition*, August 2007.

[13] D. Coalition, "Demand response, smart grid, & renewable energy," *Demand Response and Smart Grid Coalition*, August 2007.

[14] E. Miller, "Smart grid overview," *Trilliant, Inc.*, February 2009.

[15] S. Amin, A. A. Cárdenas, and S. S. Sastry, "Safe and secure networked control systems under denial-of-service attacks," in *Hybrid Systems: Computation and Control*, pp. 31–45, Springer, 2009.

[16] M. Benattou and K. Tamine, "Intelligent agents for distributed intrusion detection system," *World Academy of Science, Engineering and Technology*, vol. 6, pp. 190–200, 2005.

[17] E. Byres and J. Lowe, "The myths and facts behind cyber security risks for industrial control systems," in *Proceedings of the VDE Kongress*, vol. 116, 2004.

[18] V. Chatzigiannakis, G. Androulidakis, M. Grammatikou, and B. S. Maglaris, "An architectural framework for distributed intrusion detection using smart agents.," in *Security and Management*, pp. 193–199, 2004.

[19] S. Forrest, S. A. Hofmeyr, and A. Somayaji, "Computer immunology," *Communications of the ACM*, vol. 40, no. 10, pp. 88–96, 1997.

[20] M. G. Gouda and X.-Y. Liu, "Firewall design: Consistency, completeness, and compactness," in *Distributed Computing Systems, 2004. Proceedings. 24th International Conference on*, pp. 320–327, IEEE, 2004.

[21] Z. M. Fadlullah, M. M. Fouda, N. Kato, X. Shen, and Y. Nozaki, "An early warning system against malicious activities for smart grid communications," *Network, IEEE*, vol. 25, no. 5, pp. 50–55, 2011.

[22] M. HADLEY, N. Lu, and A. DEBORAH, "Smart-grid security issues," *IEEE Security and Privacy*, vol. 8, no. 1, pp. 81–85, 2010.

[23] G. G. Helmer, J. S. Wong, V. Honavar, and L. Miller, "Intelligent agents for intrusion detection," in *Information Technology Conference, 1998. IEEE*, pp. 121–124, IEEE, 1998.

[24] R. J. Turk, *Cyber incidents involving control systems*. Idaho National Engineering and Environmental Laboratory, 2005.

[25] M. Long, C.-H. J. Wu, and J. Y. Hung, "Denial of service attacks on network-based control systems: impact and mitigation," *Industrial Informatics, IEEE Transactions on*, vol. 1, no. 2, pp. 85–96, 2005.

[26] B. Mukherjee, L. T. Heberlein, and K. N. Levitt, "Network intrusion detection," *Network, IEEE*, vol. 8, no. 3, pp. 26–41, 1994.

[27] R. Mitchell and I.-R. Chen, "Effect of intrusion detection and response on reliability of cyber physical systems," *Reliability, IEEE Transactions on*, vol. 62, no. 1, pp. 199–210, 2013.

[28] B. Parno, A. Perrig, and V. Gligor, "Distributed detection of node replication attacks in sensor networks," in *Security and Privacy, 2005 IEEE Symposium on*, pp. 49–63, IEEE, 2005.

[29] T. T. Kim and H. V. Poor, "Strategic protection against data injection attacks on power grids," *Smart Grid, IEEE Transactions on*, vol. 2, no. 2, pp. 326–333, 2011.

[30] M. Zeller, *Common Questions and Answers Addressing the Aurora Vulnerability.* Schweitzer Engineering Laboratories, Inc., 2010.

[31] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security (TISSEC)*, vol. 14, no. 1, p. 13, 2011.

[32] J.-M. Lin and H.-Y. Pan, "A static state estimation approach including bad data detection and identification in power systems," in *Power Engineering Society General Meeting, 2007. IEEE*, pp. 1–7, IEEE, 2007.

[33] W. J. Broad, J. Markoff, and D. E. Sanger, "Israeli test on worm called crucial in iran nuclear delay," *The New York Times*, January 2011.

[34] R. Chabukswar, Y. Mo, and B. Sinopoli, "Detecting integrity attacks on scada systems," in *Proceedings of the 18th IFAC World Congress, Milano, Italy*, pp. 11239–11244, 2011.

[35] A. J. Wood and B. F. Wollenberg, *Power generation, operation, and control.* John Wiley & Sons, 2012.

[36] A. Monticelli, *State estimation in electric power systems: a generalized approach*, vol. 507. Springer, 1999.

[37] D. E. Denning, "An intrusion-detection model," *Software Engineering, IEEE Transactions on*, no. 2, pp. 222–232, 1987.

[38] S. Janakiraman, V. Vasudevan, and S. Radhakrishnan, "Agent based intrusion detection system: A computational biology approach," in *India Conference, 2006 Annual IEEE*, pp. 1–4, IEEE, 2006.

[39] J. Salmeron, K. Wood, and R. Baldick, "Analysis of electric grid security under terrorist threat," tech. rep., DTIC Document, 2004.

[40] N. H. Abbasy and W. El-Hassawy, "Power system state estimation: Ann application to bad data detection and identification," in *AFRICON, 1996., IEEE AFRICON 4th*, vol. 2, pp. 611–615, IEEE, 1996.

[41] Y. Gu, T. Liu, D. Wang, X. Guan, and Z. Xu, "Bad data detection method for smart grids based on distributed state estimation," in *Communications (ICC), 2013 IEEE International Conference on*, pp. 4483–4487, IEEE, 2013.

[42] E. N. Asada, A. V. Garcia, and R. Romero, "Identifying multiple interacting bad data in power system state estimation," in *Power Engineering Society General Meeting, 2005. IEEE*, pp. 571–577, IEEE, 2005.

[43] J. Chen and A. Abur, "Improved bad data processing via strategic placement of pmus," in *Power Engineering Society General Meeting, 2005. IEEE*, pp. 509–513, IEEE, 2005.

[44] J. Chen and A. Abur, "Placement of pmus to enable bad data detection in state estimation," *Power Systems, IEEE Transactions on*, vol. 21, no. 4, pp. 1608–1615, 2006.

[45] E. Handschin, F. Schweppe, J. Kohlas, and A. Fiechter, "Bad data analysis for power system state estimation," *Power Apparatus and Systems, IEEE Transactions on*, vol. 94, no. 2, pp. 329–337, 1975.

[46] A. Garcia, A. Monticelli, and P. Abreu, "Fast decoupled state estimation and bad data processing," *Power Apparatus and Systems, IEEE Transactions on*, no. 5, pp. 1645–1652, 1979.

[47] S. Gastoni, G. Granelli, and M. Montagna, "Multiple bad data processing by genetic algorithms," in *Power Tech Conference Proceedings, 2003 IEEE Bologna*, vol. 1, pp. 1–6, IEEE, 2003.

[48] V. Quintana, A. Simoes-Costa, and M. Mier, "Bad data detection and identification techniques using estimation orthogonal methods," *Power Apparatus and Systems, IEEE Transactions on*, no. 9, pp. 3356–3364, 1982.

[49] F. C. Schweppe, "Power system static-state estimation, part i: Exact model," *Power Apparatus and Systems, IEEE Transactions on*, no. 1, pp. 120–125, 1970.

[50] F. C. Schweppe and D. B. Rom, "Power system static-state estimation, part ii: Approximate model," *power apparatus and systems, ieee transactions on*, no. 1, pp. 125–130, 1970.

[51] F. C. Schweppe, "Power system static-state estimation, part iii: Implementation," *Power Apparatus and Systems, IEEE Transactions on*, no. 1, pp. 130–135, 1970.

[52] X. Nian-De, W. Shi-Ying, and Y. Er-Keng, "A new approach for detection and identification of multiple bad data in power system state estimation," *Power Apparatus and Systems, IEEE Transactions on*, no. 2, pp. 454–462, 1982.

[53] E. Handschin, F. C. Schweppe, J. Kohlas, and A. Fiechter, "Bad data analysis for power system state estimation," *Power Apparatus and Systems, IEEE Transactions on*, vol. 94, no. 2, pp. 329–337, 1975.

[54] B. Singh, N. Sharma, A. Tiwari, K. Verma, and S. Singh, "Applications of phasor measurement units (pmus) in electric power system networks incorporated with facts controllers," *International Journal of Engineering, Science and Technology*, vol. 3, no. 3, 2011.

[55] N. Shivakumar and A. Jain, "A review of power system dynamic state estimation techniques," in *Power System Technology and IEEE Power India Conference, 2008. POWERCON 2008. Joint International Conference on*, pp. 1–6, IEEE, 2008.

[56] M. Do Coutto Filho, J. D. Glover, and A. L. da Silva, "State estimators with forecasting capability," *11th PSCC Proc*, vol. 2, pp. 689–695, 1993.

[57] S. Chohan, *static and tracking state estimation in power systems with bad data analysis.* Phd dissertation, Centre for Energy Studies, IIT-Delhi, July 1993.

[58] Y. Cheng, W.-J. Lee, S.-H. Huang, and J. Adams, "Dynamic parameter identification of generators for smart grid development," in *Power and Energy Society General Meeting, 2011 IEEE*, pp. 1–7, IEEE, 2011.

[59] E. Farantatos, R. Huang, G. J. Cokkinides, A. Meliopoulos, B. Fardanesh, and G. Stefopoulos, "Advanced disturbance recording and playback enabled by a distributed dynamic state estimation including bad data detection and topology change identification," in *Power and Energy Society General Meeting, 2012 IEEE*, pp. 1–8, IEEE, 2012.

[60] D. Pauli, "Copycat weapons not limited to nation-states, boffin says," *SC Magazine, Australia*, November 2013.

[61] F. C. Schweppe and J. Wildes, "Power system static-state estimation, part I: Exact model," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-89, no. 1, 1970.

[62] A.-M. Borbely and J. F. Kreider, *Distributed Generation: The Power Paradigm for the New Millennium.* Boca Raton, FL, USA: CRC Press, 2001.

[63] H. Zhu and G. Giannakis, "Sparse overcomplete representations for efficient identification of power line outages," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 2215–2224, 2012.

[64] A. Pinar, J. Meza, V. Donde, and B. Lesieutre, "Optimization strategies for the vulnerability analysis of the power grid," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1786–1810, 2010.

[65] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, 2007.

[66] W. Ren and R. W. Beard, "Consensus seeking in multiagent systems under dynamically changing interaction topologies," *IEEE Transactions on Automatic Control*, vol. 50, no. 5, pp. 655–661, 2005.

[67] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*. New Jersey: Prentice Hall, 1996.

[68] B. D. Anderson, "Algebraic properties of minimal degree spectral factors," *Automatica*, vol. 9, pp. 491–500, 1973.

[69] G. Gu, X. Cao, and H. Badr, "Generalized LQR control and Kalman filtering with relations to computations of inner-outer and spectral factorizations," *IEEE Transactions on Automatic Control*, vol. 51, no. 4, pp. 595–605, 2006.

[70] C. Xiao and D. J. Hill, "Generalizations and new proof of the discrete-time positive real lemma and bounded real lemma," *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, vol. 46, no. 6, pp. 740–743, 1999.

[71] A. Abur and Antonio Gómez Expósito, *Power System State Estimation: Theory and Implementation*. New York, USA: Marcel Dekker Inc., 2004.

[72] H. Zhang, F. Lewis, and A. Das, "Optimal design for synchronization of cooperative systems: State feedback, observer and output feedback," *IEEE Transactions on Automatic Control*, vol. 56, no. 9, pp. 1948–1952, 2011.

[73] L. Alvergue, A. Pandey, G. Gu, and X. Chen, "Output consensus control for heterogeneous multi-agent systems," in *52nd IEEE Conference on Decision and Control*, (Firenze, Italy), 2013.

[74] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. New Jersey: Prentice Hall, 2005.

[75] L. Ljung, *System Identification - Theory for the User*. New Jersey: PTR Prentice Hall, 1999.

[76] F. Pasqualetti, F. Dorfler, and F. Bullo, "Cyber-physical attacks in power networks: Models, fundamental limitations and monitor design," in *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, pp. 2195–2201, IEEE, 2011.

[77] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems–part i: Models and fundamental limitations," *arXiv preprint arXiv:1202.6144*, 2012.

[78] IEEE, *DATA SHEETS FOR IEEE 14 BUS SYSTEM*.

[79] D. Kushner, "The real story of stuxnet," *Spectrum, IEEE*, vol. 50, no. 3, pp. 48–53, 2013.

[80] T.-T. Tran, O.-S. Shin, and J.-H. Lee, "Detection of replay attacks in smart grid systems," in *Computing, Management and Telecommunications (ComManTel), 2013 International Conference on*, pp. 298–302, IEEE, 2013.

[81] P. Stoica and R. L. Moses, *Spectral analysis of signals*. Pearson/Prentice Hall Upper Saddle River, NJ, 2005.

[82] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

Vita

Bixiang Tang was born in the city of Nanjing of Jiangsu Province in China 1983. He is the son of Huixin Tang and Xueqin Bi. He received the bachelor degree of electrical engineering from Jiangsu University in 2002. And later, he became a graduate student in Jiangsu University and got the master degree of electrical engineering in 2006. Then, he came to American and joined the Division of Electrical and Computer Engineering, School of Electrical Engineering and Computer Science (EECS) at Louisiana State University in 2009 in doctoral program. He obtained a M.S. degree in Electrical Engineering in 2011, LSU. He is currently a candidate for the degree of Doctor of Philosophy in Division of Electrical and Computer Engineering, School of Electrical Engineering and Computer Science.