DATA-DRIVEN APPROACHES FOR CYBER-PHYSICAL ATTACK AND DEFENSE IN MODERN POWER GRIDS

by

Youqi Guo

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Electrical Engineering

at

The University of Wisconsin-Milwaukee

August 2022

ABSTRACT

DATA-DRIVEN APPROACHES FOR CYBER-PHYSICAL ATTACK AND DEFENSE IN MODERN POWER GRIDS

by

Youqi Guo

The University of Wisconsin-Milwaukee, 2022
Under the Supervision of Professor Lingfeng Wang

With the tighter integration of advanced communication and computing technologies, electrical power system is being transformed to more complex, efficient and sustainable smart grid. Today, almost every sector of physical power system including generation, transmission, distribution and consumption has to be monitored, protected and controlled to interact and communicate with each other through cyber infrastructure. The use of information and communication technologies (ICTs) has not only enhanced the efficiency and reliability of smart grid, but also created new vulnerabilities if they are not accompanied by advisable security reinforcements. Various vulnerabilities that ICTs bring about may leave some sectors of the power system to a wide range of cyber-physical attacks, which are implemented in cyberspace and may adversely affect the physical infrastructure.

To improve the resiliency of power grids against the threats of cyber-physical attacks, it is imperative that we identify the risk of such attacks and thereby implement effective security strategies to protect power systems from the attacks. Therefore, the cyber-physical security has become a key concern for both industry and academia communities. As a result, a large amount of efforts have been devoted to the research of cyber-physical attack and defense in smart grid. In recent years, with the rapid development of artificial intelligence, the data-driven machine learning approaches have received wide attentions because of their salient advantages in areas of attack identification and detection, and system control and

risk mitigation. This dissertation investigates the cyber-physical attack detection, defense and mitigation by utilizing state-of-the-art data-driven machine learning techniques. Three research studies are presented in this dissertation, providing useful insight to power system administrators to identify cyber-physical attacks, mitigate the risk of attacks and protect critical power system equipments, and thereby enhance power system resiliency.

The first study proposes a novel reinforcement-learning-based dynamic defense strategy against dynamic load altering attack (D-LAA). A two-player zero-sum Markov game is formulated to analyze the complex interactions between the attacker and the defender, in which all players are rational and tend to maximize their own benefits. The proposed minimax-q algorithm is applied to derive the attacker/defender's Nash equilibrium strategies. The performance of the proposed model is evaluated on the IEEE 39-bus system. Comparisons between the dynamic defense strategy and the passive defense strategy are conducted, and the results verify the advantage of the proposed dynamic defense strategy. To improve the power system resiliency, this defense strategy can be deployed in advance when such cyber-physical attacks are anticipated.

The second research presents a new TripleGAN-based defense framework against the stealthy FDI attacks, which aims to accurately detect the attack and effectively mitigate the impact at the same time with a few labeled historical measurements. In this model, the detection is performed by the classifier and the mitigation is carried out by replacing the tampered measurements with the produced measurement data from the generator. To improve the detection accuracy and recovery efficiency, an extended loss function integrated with feature matching is proposed. The simulation results demonstrate that the proposed defense model is able to accurately detect the stealthy FDI attacks and the recovered state estimation is sufficiently close to the normal operation status, which thus improves the power system resilience. Furthermore, under various circumstances (with different numbers of targeted measurements, different intensities of environmental noise, and fewer historical data), the obtained results confirm that the proposed techniques exhibit advantages over other machine learning based detection and recovery methods.

In the last research work, a data-driven FDI attack model against load frequency control (LFC) system is proposed based on multi-agent deep reinforcement learning (MA-DRL). Instead of using legacy linearized LFC model, AC state estimation (ACSE) is integrated with LFC to reduce measurement noises and perform bad data detection. Thus, the system environment becomes more practical and complex, and more requirements need to be satisfied for the attacker to perform successful FDI attacks. In order to achieve two attack objectives simultaneously, i.e., stealthily maximizing the frequency deviation and minimizing the number of compromised measurements, a modified Multi-Agent Deep Deterministic Policy Gradient (MA-DDPG) algorithm is devised in this study, which treats the two objectives separately by global and local individual critic networks other than a simple linear combination. The impact of FDI attack on the LFC with ACSE is also analytically derived. The simulation results on the New-England 39-bus system demonstrate a good performance of the proposed FDI attack model compared with other methods. In addition, corresponding countermeasures based on the critical measurements are discussed and verified.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to take this opportunity to send my utmost and sincere gratitude to my advisor Dr. Lingfeng Wang for your guidance, patience and constant encouragement throughout my journey. Dr. Wang has introduced me to the field of power systems. He has always guided my research work with valuable insight and provided me with the best environment to study and conduct research. His extensive knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life. There are no words enough to describe my appreciation to him.

I would also like to appreciate the valuable feedback from my committee members. They are also the ones who help lay the foundation of my research and future career. Dr. Zeyun Yu teaches me scientific computing and Dr. Jun Zhang introduces me to the field of machine learning. Dr. Wei Wei and Dr. Weizhong Wang's research papers always help me solve problems on my own research work. I'm grateful to get a chance to learn from your classes and work.

In addition, acknowledgment should also go to all my group members, Dr. Zhaoxi Liu, Yitong Shen, Yunfan Zhang and Pikkin Lau, for a cherished time working together and lots of inspiring discussions. The work would not have been finished without your kind help.

Special thanks to my beloved family: my parents and my wife, for their loving support. Even if not being by my side, the constant love and tremendous encouragement from my parents keep me motivated and confident. Deepest thanks to my wife, who is always supportive of my adventures. Thank you for being with me whenever I need you and willing to do everything to support me.

I am also grateful to the financial support for this work. This work was supported in part by the U.S. National Science Foundation under Award ECCS1711617.

Chapter 1

INTRODUCTION

1.1 Overview

As one of the most fundamental and complicated artificial system in modern society, the electrical power system has been going through more than a hundred years of development since the first electrical power station was constructed in New York City in 1882. Over the past few decades, a mass of technical innovations have been adopted by power industry to cope with unprecedented challenges on power system planning, operation and maintenance. As of now, with the recent invention and advancement in control, sensing, monitoring and communication, the legacy power system has evolved into smart grid, which is envisaged to achieve higher requirements of efficiency, resiliency and sustainability. One of the main features of smart grid is the integration of cyber infrastructure, which enables the collection and analysis of data from millions of distributed devices, such as phasor measurement units (PMU), smart meters, circuit reclosers and breakers. For example, a number of utilities in the United States deploy Advanced Distribution Management Systems (ADMS) to support distribution management and optimization. By the collection of a great number of data from distributed end-points, the control center can achieve volt-VAr optimization (VVO), conservation voltage reduction (CVR), isolation and restoration, and peak demand management through directly communicating with regulating devices, such as load tap changers, line voltage regulators and line capacitor banks.

However, the beauty of smart grid innovation cuts both ways: deep integration of the

power system operation with advance computation and communication technologies not only enhances the overall system efficiency, but at the same time the vulnerability to potential cyber-attacks would exceedingly soar because of massive network operations. Critical computation and control process in smart grid, such as state estimation, optimal power flow and CVR, highly rely on a secure and robust cyber infrastructure. The cyber vulnerabilities may enable malicious adversaries to manipulate system measurements, meter measurements or even directly sabotage vulnerable critical control process. Such attacks that is implemented in cyberspace and adversely impact the physical infrastructure are referred as cyber-physical attacks.

In the past decades, several cyber-physical attacks have been reported in power industry. In 2003, the control system of the David-Besse nuclear plant in Ohio was invaded through the Slammer worm, leading to 5 hours offline of the supervisory system. In 2010, a cyber worm "Stuxnet" attacked the Iranian nuclear fuel enrichment plant. The adversaries targeted the programmable logic controllers of the supervisory control and data acquisition (SCADA) system using four zero-day vulnerabilities to bypass the attack detection. As a result, the centrifuges was forced to work in a unexpected range of speed because of the malicious alternation on current frequency. This facility had to cease several times because of a series of major technical problems [1, 2]. In 2015, attackers remotely switched off the breakers in a series of substations by pre-installed malware, resulting in a widespread outage affecting approximately 225,000 customers in the Ukrainian power grid. This blackout is the first publicly acknowledged incident caused by cyber-physical attacks, which is even more destructive than natural disasters [3]. As mentioned in [4], cyber security issues are considered as one of the top priorities for the smart grid construction. Hence, considering the increasing cyber-physical threats, it is imperative that we understand the risks resulting from cyber-physical attacks and thus implement effective security strategies against them.

1.2 Research Objectives and Contributions

These real cyber-threats examples facilitate synergistic efforts from industry practitioners and research communities towards to the security issues in smart grid. This dissertation investigates cyber-physical attacks detection, defense and mitigation through data-driven approaches. The three research studies in this dissertation provide useful insight into power system operators to identify cyber-physical attacks and protect critical infrastructures, and thereby power system resiliency gets enhanced. The major contributions are listed as following.

- A two-player zero-sum multistage game considering both dynamic of the attacker and the defender is proposed. As a newer type of cyber-physical attacks, the dynamic load altering attacks (D-LAA) is investigated and extended as sequence attacks in this game, which allow the attacker takes offensive actions one by one based on the states of system and protection policy. A minimax-q learning scheme is adopted in this research to effectively find out the optimal defense sequence against chronological D-LAA. Comparisons between the dynamic defense strategy and the passive defense strategy are conducted, and the results verify the advantage of the proposed dynamic defense strategy.

- A novel Triple Generative Adversarial Network (TripleGAN) based defense framework is developed against the stealthy False Data Injection (FDI) attacks. The proposed model aims to effectively detect and mitigate the attacks by limited historical measurement data. In this model, the detection is performed by the classifier and the mitigation is carried out by replacing the tampered measurements with the produced measurement data from the generator. Further, A new regularization scheme to the TripleGAN model is developed by integrating feature matching in the loss function to enhance the detection accuracy and recovery efficiency. The simulation results demonstrate that the proposed technique can effectively defend the grid against the FDI

attacks and outperform other machine learning methods

- A data-driven FDI attack model against Load Frequency Control (LFC) system is proposed. Instead of legacy linearized LFC model, AC state estimation (ACSE) is coupled with LFC to reduce environment noise and filter bad data before sending measurements to LFC controller, which is more precise and practical for the analysis of interactions among control areas. Multi-Agent Deep Reinforcement Learning (MA-DRL) method is used to solve the problem of increasing frequency deviation and minimizing the number of tampered measurements. In order to achieve these two objectives simultaneously, a modified Multi-Agent Deep Deterministic Policy Gradient (MA-DDPG) algorithm is proposed. The simulation results validate the good performance of proposed FDI attack model compared with other methods. Furthermore, corresponding countermeasures based on checking out the critical measurements are discussed and verified.

1.3 Outline of the Dissertation

The rest of this dissertation is organized as follows. Chapter 2 presents related background knowledge including the description of modern power systems, cyber threats in smart grid, and data-driven methods and applications. Chapter 3, 4, and 5 present three research work in detail respectively. Conclusion and future work are given in Chapter 6.

Chapter 2

BACKGROUND

2.1 Smart Grid

In general, a modern power system consists of generation, transmission, distribution, customers, service providers and grid operations. As shown in Fig. 2.1, the electrical energy and data continually flow among these sectors and maintain all power grids in operation, which highly rely on wire or wireless communication. In smart grid, the electrical grid is often referred as physical layer and the communication network is referred as cyber layer. In the context, all the equipments that directly involved in delivering electricity from points of generation to customers are categorized into physical layer, while all the network devices, intelligent electronic devices and servers which are used to monitor, manage and protect this electricity delivering process belong to cyber layer.

2.1.1 Physical Layer

The physical layer of smart grid is composed of generation, transmission, distribution and consumption. The physical layer never stops the pace of expansion and development since it was constructed hundred year ago due to higher and higher electricity demand. Especially for the last couple of decades, with the integration of Distributed Energy Resources (DERs), High Voltage Direct Current Transmission (HVDC), electric vehicles (EV), micro-grids, etc, the electrical grid becomes more assorted, while the new elements also bring in new challenges

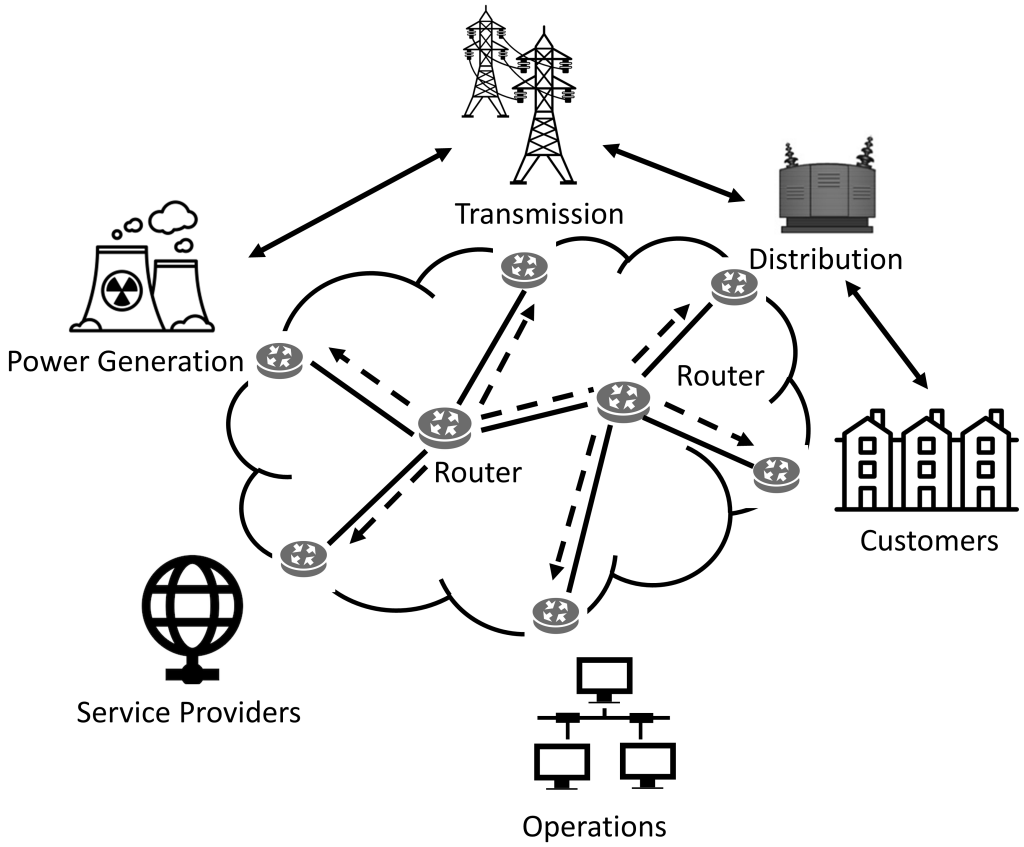

Figure 2.1: Communication and control system of the modern power grids

for power system researchers and operators.

Generation

Electrical generation is the process of converting sources of primary energy to electric power. As of now, fossil fuels such as coal, peat and natural gas are still the most common resources utilized in electricity generation. However, methane leaks and carbon dioxide emission from fossil fuel-based electricity generation are the part of reasons for the greenhouse effect and climate change. Due to the environmental concerns, more renewable resources generation are developed, such as solar, wind, geothermal energy, etc. As per the International Energy Agency (IEA), 85% of global electrical output need to be generated from low-carbon elec-

tricity generation by 2040 to prevent the worst effects of climate change [5]. Thus, there is long way to go to build up environmental-friendly power generation and meet the electricity demand at the same time.

Transmission

Electrical transmission system is to deliver the electricity from power station to the places near to customers over long distance. To reduce the energy loss, voltage level needs to be increased by transformers before sending out from power plants. In the US, common transmission systems are overhead High Voltage Alternate Current (HVAC) systems, including voltage level 138kV, 230kV, 345kV, 500kV and 765kV. Recent years, High Voltage Direct Current (HVDC) systems attract widespread attentions due to the benefits of economic, operation and control.

Distribution

The electric distribution system is the final stage in the delivery of electricity, which carries electricity from transmission system to customers. The voltage is stepped down to medium voltage level ranging between 4 kV and 35kV at distribution substation with the use of transformers. Some industrial customers can directly connect to the medium voltage, while for residential and commercial users, the voltage has to be further stepped down to 120/240V by service transformer located near the customer's premises.

Conventionally, the distribution systems only operate as simple distribution lines to carry electricity to customers. Today, as a result of heavy integration with renewable energy generations at the distribution level, such as wind farms and solar farms, distribution systems become more independent from transmission networks. It requires the use of advanced technology and management to cope with the challenges, such as ADMS, battery storage

power station, data analytics, etc.

Consumption

Traditionally, the consumption means nothing but passive load at industrial or residential customers in power industry. Nowadays, the electricity consumers not only use electrical energy but have options to feed back to the power grid. For example, the maturer DER techniques enable the customers to install their own DER generation, such as solar rooftop. Another example is Vehicle to Grid (V2G) technology that enables energy to be pushed back to the power grid from the battery of electric cars. Although such new types of consumption loads bring in significant advantages on energy efficiency, high requirements of power grid capacity, operation reliability and security have to be met. It is also one of the pressing issues to tackle for power system researchers.

2.1.2 Cyber Layer

The operation of physical power system equipments are undergoing a rise in complex cyber-infrastructures, which is a critical component of modern power grid. Since today power system has been evolved into a unprecedented large-scale infrastructure and has to accomplish more complicated tasks, almost every part of the physical layer has to be monitored, protected and controlled in order to interact and communicate with each other well. This is where the cyber layer come into play.

The cyber layer of smart grid is with the incorporation of Information-Communication Technologies (ICTs), which comprise four essential categories for the implementation on the operation of power system, including acquisition, implementation, processing and communication of subsystems [6–8].

- **Acquisition**: Distributed sensors collect system data such as power flow measurement, bus voltage, frequency, the state of circuit breakers and reclosers and then transmit all the data via communication medium to the processing stage.

- **Processing**: Processes data collected from distributed sensors including data cleaning, data integrity checking, and state estimation. The system operator will know the status of the real-time system and make decisions for implementation.

- **Implementation**: Performs the required actions based on the data processing results. The actions may include sending control signal to corresponding controllable devices, activating the protecting circuit breakers and relays, and enable or disable some capacitor banks for volt-var control.

- **Communication**: The coordination of all subsystems in power network relies on Information Technology (IT) network, which hosts function such as long-distance transmitting, corporate web server, cloud service, etc. IT network enhances the situational awareness of the network and helps make correct decisions for the operators and efficiently improves the power system reliability. Today, IT network is typically isolated with firewalls and operated in Virtual Private Network (VPN).

2.2 Cyber Threats in Smart Grids

2.2.1 Cyber-physical Security Concerns of Smart Grids

The use of information and communication technologies (ICTs) has not only enhanced the efficiency and reliability of the power grids, but also created new vulnerabilities. Cyber-physical vulnerability is one of existing major security concerns, which are identified as the weakness resulting from the integration of cyber layer with the physical layer of smart grids.

Cyber-physical vulnerabilities can be usually targeted through network communication and intelligent devices vulnerabilities.

- **Network communication vulnerabilities**: Power system infrastructures rely on Ethernet-based local area network to communicate and interact with each other. However, the Ethernet-based networks are vulnerable to interception and Man-in-the-Middle attacks. Attackers are able to exploit the known vulnerabilities of the existing internet protocol standards to launch attacks on the network, such as injecting false data, impersonating components and releasing classified information [9–11]. Most of communication protocols used in power industry such as Modbus and DNP3 do not provide enough security measures. The lack of encryption and authentication mechanisms enables the adversaries to alter, spoof and tamper the data in transmission.

- **Intelligent devices vulnerabilities**: The components that establish the cyber infrastructure networks in smart grid may pose serious security concerns, including smart meters, routing devices, intelligent electronics devices and installed software. Well-known vulnerabilities of these equipments can be directly utilized by the attackers. The attackers may also tamper the same devices and then find out the zero-day vulnerabilities by comprehensive testing. For example, smart meters may have back doors that could give full access and control to the users who exploit the factory login account and credential. Once the attackers take over the control of smart meters, false data could be injected and the operator might make wrong decisions leading to power disruption. In addition, the attackers may also use the meters as bot to launch attack against other components within the network [12].

2.2.2 Cyber-Physical Attacks

In smart grid, attacks that are implemented in cyberspace and adversely impact the physical infrastructure are referred as cyber-physical attack. An assessment in [13] has presented that the major cybersecurity concerns range from exploiting well-known protocols vulnerabilities to the leakage of confidential information. As one of the most threaten attacks against the security of smart grid, tremendous research have been addressed on the risk of cyber-physical attacks. According to the tech report [14], cyber-physical attacks can be classified into three categories based on the attack target:

1. **Control signal attacks**: Control signal attacks aim to acquire the physical devices authority by bypassing the data authentication and integration examinations, and then operate the devices to achieve malicious objects. Such attacks are usually constructed to target critical devices in smart grid such smart inverters, Flexible AC Transmission System (FACTS) devices, Automatic Generation Control (AGC) system, circuit breakers, etc. The adversaries may utilize the knowledge about the targeted devices, such as load flow information and generator ramping limits, to efficiently achieve attack goals. Although N-1 contingency has been developed and deployed in most power grid network, a lot of research show that most vulnerable devices might cause cascading failure by exploiting the clustering-based vulnerability. The well-known control signal attacks include Aurora attacks [15] and pricing attacks [16].

2. **Measurement attacks**: Measurement attacks target on altering the sensor measurement data transmitted by the communication channels or directly falsify the sensor units in the field. Depending on the attackers' capabilities, they may eavesdrop the measurement data for reconnaissance, manipulate the firmware of the devices, or even control sensors to transfer tampered measurement data. For example, an attacker may implement Domain Name Systems (DNS) hijacking attacks, which changes The DNS server of the device gateway to a controller server, to transmit all measurement data

to a malicious server instead of the legitimate server. This type of attack may disable the system operator's situational awareness to the status of the power grid and induce wrong operations based on the manipulated measurements. Typical measurement attacks include False Data Injection (FDI) attacks [17], AGC attacks [18], load redistribution attacks [19], topology attacks [20] and GPS spoofing attacks [21].

3. **Control-signal-measurement attacks**: This type of attacks are coordinated attacks that target on both the control signals and measurements. The control signal attacks can cause significant consequences in short period, while the measurement attacks can cover the ongoing control signal attacks. In this way, the manipulated measurements could pass anomaly detection mechanisms. Some researches reveal that the attackers may use the coordinated attacks to design stealthy control signal attacks through disguise the attack consequences and without being detected by attack detection and mitigation systems. Typical control-signal-measurement attacks include line outage masking attacks [22] and Stuxnet-like attacks [23].

2.2.3 Cyber-Physical Defense

Cyber-physcial defense is obviously the focus of ongoing research efforts, where various reactive (acting as mitigation) and proactive (acting in anticipation) defense methodologies are proposed to increase the ability to detect and identify attacks, reduce the risk of threats and initiate mitigation countermeasures to restore the system operations. According to different methodologies of defense approaches, cyber-physical defense can be categorized as follows.

1. **Protecting measurement sensors**: As mentioned above, the majority of attacks are launched on the system control and measurement signal. Thus, a common approach is to strategically select and secure critical control or measurement signal. In [24], the authors present that attackers aim to design attack vector avoiding some well-protected

measurements and states. From the defender's perspective, the system administrator should find out the sets of the critical measurements and the verified state that ensure the adversaries are not able to craft stealthy attack vectors. The concern here is the economic cost of the protection and verification of a large number of measurement sensors. Zhao et al. [25] develop an enhanced detection method against FDI attacks by checking the statistical consistency of measurements from a limited number of sensors. In [26] a short-term measurement forecasting is proposed to improve the MNU data redundancy. The authors propose a vector autoregression method as the prediction process that captures the interdependencies between the events in different time slots to detect FDI attacks in [27]. An optimal PMU placement study is conducted in [28], which maximizes the determinant of the empirical observability of Gram matrix. In [29], a multi-layer detection method is proposed against simultaneous GPS spoofing attacks for limited number of PMUs.

2. **Modeling and algorithmic improvement**: Another major category the defense approaches is the enhancement of the detection models and algorithms. In [30], Matrix Separation is introduced to defend against the sparse FDI attack sequence based on the separation of nominal states and anomalies matrices, which is solve by two methods: the nuclear norm minimization and low-rank matrix factorization. A new detection method is derived in [31] by calculating the metrics of Kullback-Leibler distance (KLD) between two probability distributions, i.e., historical and tampered measurements, to track the dynamics of the measurement variations. Ashok et al. [32] state that the existing cyber-physical defense focuses on either redundant measurements or the cybersecurity of sensors and communication channels. These offline approaches make specific assumptions about the attacks and systems, which are restrictive. One solution of PMUs placement or security mechanism may not be generalized to another system configuration. Thus, the authors propose an online anomaly detection that covers more attack scenarios. The proposed method leverages online information ob-

tained from load forecasts, generation schedules, and real-time data from PMUs to detect anomaly measurements.

3. **Moving target defense**: Computational burden and real-time implementation are the major concerns of aforementioned operational defense approaches. As an emerging technique, moving target defense (MTD), is originally proposed to improve network security [33]. MTD can dynamically change the system configuration thus increasing the difficulty of successful exploits and required effort for the malicious attackers. With the properly crafted MTD perturbation, the attacker's knowledge about the system is always outdated. Recently, MTD has been introduced in the cyber-physical security of smart grid to provide proactive defense, which is an alternative way to traditional remedial defense. Compared with the MTD in the network system, MTD in smart grid which is a cyber-physical system is more complicated since the physical dispatch of control and measurement is required. The concept of MTD was first introduced into the physical layer of the power system by Davis et al. [34]. In general, MTD dynamically modify impedance perturbations using distributed flexible AC transmission system (D-FACTS) to invalidate the knowledge of power network configuration for the attackers. Two essential steps are required to construct MTD: planning and operation. In the MTD planning stage, the utilities install D-FACTS strategically selected subset of transmission lines, which is referred as D-FACTS placement problem. After the allocation of D-FACTS devices, the system operator continuously determine the setpoint of D-FACTS devices under various demand conditions to ensure all devices work as expected.

4. **Watermarking**: Watermarking is originally used to identify the ownership of the copyright of noise-tolerant signals such as audio, video, or image data. It also can be used to check the integrity and authenticity of a signal. The first use of watermarking to defend the replay attack employed in Stuxnet was introduced by [35], where the physical watermarking is utilized as control theoretic method to authenticate the correct control

operation. The methodology is that by introducing a probe signal into the system, an expected response is supposed to be found in the true measurement feedback because of system dynamics. As a result, if the attacker is not aware of the watermarking in advance, the launched attack would be detected. In [36], the authors extend the physical watermarking to dynamic watermarking against more intelligent adversaries. The proposed approach ensures that the amount of distortion that the attacker can add are constrained to an average of power of zero by implementing the correlation detector, no matter in which way the adversarial sensors collude.

Besides the aforementioned model-based defense methods, another noteworthy category of defense approaches is on data-driven machine learning methods. The data-driven defense approaches are presented in Section 3 in this chapter.

2.2.4 Challenges and Opportunities

Although the tremendous research efforts have been put into the cyber-physical security in smart grid, some challenges and problems remain to be addressed. Various potential cyber-physical attacks and corresponding countermeasures still need to be further investigated. Meanwhile, the firmware of some critical infrastructures such as advanced meters, controlled regulation devices and data servers are outdated while no further upgrading plan due to financial issues. In addition, the emerging integration with new technologies including electric vehicles, solar panels, big data and 5G wireless communication have strong impacts on the smart grid but also will create new vulnerabilities in future. Although these are the challenges to modern power systems, they also provide a lot of opportunities to upgrade the current infrastructures, just as how power system has been evolving the last hundreds years.

2.3 Data-Driven Approaches and Applications in Power Systems

With the booming development of artificial intelligence (AI) recent years, the data-driven machine learning approaches have been gaining traction because of two salient advantages: 1) the construction of the data-driven approaches may depend on the network topology; 2) this approach is usually sensitive to time-variance measurement, which can be very effective in detecting one time interval cyber-physical attacks created based on the spatial-relationship of smart grids. Many researches have been conducted to the design and application of data-driven machine learning methods in power systems. Machine learning techniques have demonstrated great competence in extracting information from large amount of raw data available in smart grid. Some general techniques are summarized as below.

- **Classification**: When data samples are well labelled by experts indicating the class they should belong to, supervised learning based classification can be utilized. Supervised learning algorithm can analyzed the training data samples and generate an inferred function. After fully training, a model can be derived to determine the class labels for unseen instances. Many different classification algorithms have been developed and widely used in various areas, such as support vector machines (SVM), random forest, logistic regression, neural networks, etc. Recent years, the supervised learning techniques have been widely investigated to tackle power system security issues. For example, a trained binary-classifier can be used to check that the transmitted measurement data are secured or attacked. As for specific algorithm, for instance, a SVM attains is a hyperplane specified by a group of optimized parameters. During the detection stage, a new given data sample will be classified according to its location in terms of the hyperplane.

- **Clustering**: When the amount of data samples is too large to label, semi-supervised

or unsupervised machine learning techniques can be applied, which is referred as clustering. The unlabelled data will be grouped in light of likeness or contrast through this technique. An example of application in power system is to cluster a large amount of Advanced Measurement Infrastructure (AMI) data collected from various end-users to predict the electricity consuming pattern of different areas. As for the algorithms, K-means clustering and hierarchical clustering are distance based methods that are appropriate for cases with evenly distributed data samples, and density based methods such as deep brief network perform better when the clusters have different densities.

- **Reinforcement learning**: Reinforcement learning (RL) is another machine learning technique that concerned with how intelligent agents take actions in an environment to maximize the cumulative reward. In the area of cyber-physical security in power systems, RL is always used to investigate the interaction between the attackers and the system operators. As a result, the optimal defense actions for system administrators can be obtained and the defense strategy may be deployed in advance when corresponding cyber-physical attacks are anticipated.

All machine learning techniques try to discover certain unseen rules or patterns from available data, while it may be hard or even impossible to model-based methods. As we know, the model obtained via such machine learning techniques could only be as good as the data, and this reveals one of the major drawbacks of the data-driven methods - the model after training is only able to make inductions but not deductions considering that it might not have fully incorporated the inherent nature of the system that it oversees. However, data-driven methods often turn out to be the best options in practice rather than model-based methods when developing an accurate model-based solution is too difficult, even impossible at all. Besides, the data-driven methods can automate and speed up the induction process, and provide extra benefits in terms of efficiency and efficacy compared to human experts. Several specific applications of data driven machine learning techniques in power systems are introduced in the following research work.

2.4 Summary

This chapter provides background knowledge about smart grids, cyber vulnerabilities and threats of smart grid, and data-driven approaches and applications in power system. Detailed research work are presented in the following chapters.

Chapter 3

GAME-THEORETIC BASED DYNAMIC DEFENSE STRATEGY

The work reported in this chapter has been partially published in the following article:

Youqi Guo, Lingfeng Wang, Zhaoxi Liu, Yitong Shen, "Reinforcement-learning-based dynamic defense strategy of multistage game against dynamic load altering attack", *International Journal of Electrical Power & Energy Systems*, Volume 131, 2021, https://doi.org/10.1016/j.ijepes.2021.107113.

3.1 Overview

As the current power grid is highly interconnected and more information and communication technologies (ICTs) are being deployed recently, it could be the target of malicious cyber-physical attacks. Dynamic load altering attacks (D-LAAs), as a special case of load altering attacks, could be performed to interfere the demand response and ultimately force certain generators off-line. Cascading failures due to transmission line overloads may also be triggered. In this chapter, we propose a new dynamic defense strategy against D-LAAs through a multistage game between the attacker and the defender which is solved by minimax-q learning. Different from the static game, the multistage game considers the attacker and defender's action sequences and the optimal strategies at each state are learned. After each time step, the cascading failure is measured, and the load shedding is used as the feedback for

the attacker to generate the next action strategy. The performance of the proposed model is evaluated on the IEEE 39-bus system. Comparisons between the dynamic defense strategy and the passive defense strategy are conducted, and the results verify the advantage of the proposed dynamic defense strategy. To improve the power system resilience, this defense strategy can be deployed in advance when such cyber-physical attacks are anticipated.

3.2 Related Work

Load Altering Attack (LAA) is a representative cyber-physical attack with the aim to maliciously control and alter a group of remotely accessible yet unsecured controllable loads. A successful LAA can disturb the balance between the power demand and supply, causing frequency and angle instability and consequently system blackout through circuit overflow or generator tripping. The potential vulnerable loads to LAAs can be frequency-responsive loads [37, 38], data center's computational load [39], loads with direct load control (DLC) which is one of the most common demand side management programs [40, 41], etc.

LAAs can be categorized into static load altering attack (S-LAA) (which is mainly focused on the amount of vulnerable loads) and dynamic load altering attack (D-LAA) (which is additionally concerned with the trajectory of the changes that are made in the vulnerable loads). Reference [42] introduces and models S-LAA in smart grids, and the studies in [43–45] address the prevention or detection of LAAs. Unlike these investigations, reference [46] introduces, characterizes and classifies D-LAAs as a new class of cyber-physical attacks against the power grids. In [47], the authors present a protection scheme using energy storage systems to improve the power grid's reaction to D-LAAs.

Game theory is oftentimes used to help people understand the situations in which decision-makers interact, e.g., between attackers and defenders. There is a wide range of situations to which game theory can be applied: political candidates competing, companies competing

in business, bidders bidding in an auction, and so on [48]. Various games are formulated to illuminate different economic, political, engineering phenomena, such as general sum, zero-sum and potential games. Recently, researchers recognized the critical role of game-theoretic approaches in power grid security. The security games introduce an analytical framework with a rich mathematical basis for modelling the interactions between intentional attackers whose aim is to disrupt the power grid and operators defending it [49, 50]. The games in power grid security are classified into two categories: *static* and *dynamic* games. The static game can be considered as a one-shot process, which means players only take one action. A wealth of research [51–56] has emerged on the static defense schemes against malicious attacks in the smart grid. In [51], the authors present a comprehensive and quantitative static game framework for the power system security problem. Under this framework, a new criterion is derived to seek reliable defense strategies. In [52], a zero-sum static game model is proposed to provide security policies in the cyber layer with corresponding resilient control in the physical layer. In [53], Farraj *et al.* analyze the cyber switching attacks and corresponding mitigation method by the zero-determinant strategy in an iterative game. The strategy allows the electric power utility (EPU) to stabilize the power grid in the face of cyber switching attacks. A game equilibrium is obtained by a zero-sum static game between intentional attackers and defenders to provide a reliable fusion-based defense scheme for the communication network of power systems in [54]. In [55], the effect of the compromised active power measurements on the electricity price is quantified. This situation is modeled as a zero-sum game between the defender and the attacker who performs the bad data injection attack on the measurements. For defending against denial-of-service (DoS) attacks, Li *et al.* [56] investigate the interaction between the sensor nodes and adversaries.

On the other hand, dynamic games have been a largely underexplored domain in the power grid security area. Most existing work mentioned previously are focused on static games or static defenses without considering dynamic processes. In dynamic or multistage games, attackers can compromise multiple components in a time sequence [57]. For some

practical cases, to obtain the maximum profit or achieve the attack objective, attackers have to take offensive actions one by one based on the defender's protection policy and the next steady state of system. Note that full knowledge and observation of the target system are required for the players. In [58], the authors propose a stochastic game to protect the power system against coordinated cyber-physical attacks. Although two states are considered, the game proposed in [58] is more like a one-shot game because the attacker can only target one element at a time. There is no more dynamic evolution in this game. Ma *et al.* consider a multi-act dynamic game in the electricity market for defending against jamming attacks in [59]. Dynamic programming is adopted to solve the game. To carry out the recursions, knowing the model of environment is necessary. In [60], a q-learning method is devised to solve a multistage game. The attacker's actions are considered while the defender's actions are pre-defined rather than evolving by interacting with the attacker's action and system state.

Furthermore, machine learning methods are being applied to address cyber-physical security issues in power systems for attack detection, analysis of defense strategy, and fault diagnosis. In [61], a deep-learning-based algorithm to detect power theft and false data injection (FDI) attack on real-time measurements is proposed. The authors in [62] use Q-learning to analyze vulnerabilities of the power system in sequential topological attacks. Wang *et al.* [63] develop a deep learning method for fault diagnosis of power plants. A hierarchical deep domain adaptation (HDDA) approach is proposed to apply a classifier with labeled data under one loading condition to detect faults with unlabeled data under another loading condition.

3.3 Research Contribution

Thus far, the focus in power grid security against malicious attacks has been mainly on static defense schemes. In contrast, in this chapter, we address a new dynamic defense policy against multistage D-LAA, which is concerned with dynamic interactions between the attacker and defender. The attack-defense interaction is modeled by a two-player zero-sum multistage game and the solution is obtained based on minimax-q learning. Unlike dynamic programming solution given in [59] that requires exact knowledge of the model of environment, the proposed minimax-q learning based solution in this chapter goes from experience to policies by learning a model rather than needing a model. The main contributions of this research are summarized as follows:

- The one-shot dynamic load altering attacks (D-LAA) in [46] is extended to a sequence attack. The corresponding cascading failures caused by D-LAA are studied holistically. It allows the attacker takes offensive actions one by one based on the states of system and adversary's protection policy to achieve much higher attack objective.

- A two-player zero-sum multistage game considering both dynamic of the attacker and the defender is proposed.Different from the one-shot games that lack dynamic evolvement of the attack-defense sequence and passive defense strategies where the evolvement of defender's actions is neglected in the existing literature, a minimax-q learning scheme is adopted in this chapter to effectively find out the optimal defense sequence against chronological D-LAA considering dynamic interactions between the attacker and the defender. This is also the main difference between the proposed dynamic defense and the existing research.

- This dynamic defense strategy is compared with the static (passive) defense policy. The simulation results show that the power grid with the proposed defense strategy does have lower load loss due to D-LAAs.

3.4 Related Preliminaries

In this section, some related preliminaries are presented including the mathematical model of D-LAAs, optimal load shedding problem and cascading failures.

3.4.1 Dynamic Load Altering Attack

D-LAA Implementation Principle

The basic threat model is adopted from reference [46]. As mentioned, D-LAA is concerned with the volume as well as the trajectory of the changes in the vulnerable load. In a closed-loop D-LAA, referring to Fig. 3.1, the attacker tries to manipulate the vulnerable load (P1) with constant monitoring at the sensor bus for the grid conditions. Although there are various approaches to measure the grid conditions and alter the load, in this research we limit our scope to the power system frequency obtained from the installed frequency sensors and frequency-responsive loads. A successful D-LAA can be conducted only if there are sufficient potential vulnerable loads to be compromised. The attack objective is to deviate the frequency from the system's nominal value and eventually push one generator off-line. To implement a D-LAA, there are three main steps that the attacker must undertake:

- Install the frequency monitor at the sensor bus and constantly send frequency acquisitions to the D-LAA controller. In general, it is not difficult to monitor the frequency of power system using an inexpensive commercial sensor.
- Based on the mechanism of the attack controller and the feedback signal, calculate the amount of vulnerable load which needs to be compromised at the victim bus.
- Remotely control and alter the victim load at the amount that is calculated in the last step. The feasibility of remotely altering the load is discussed in [64].

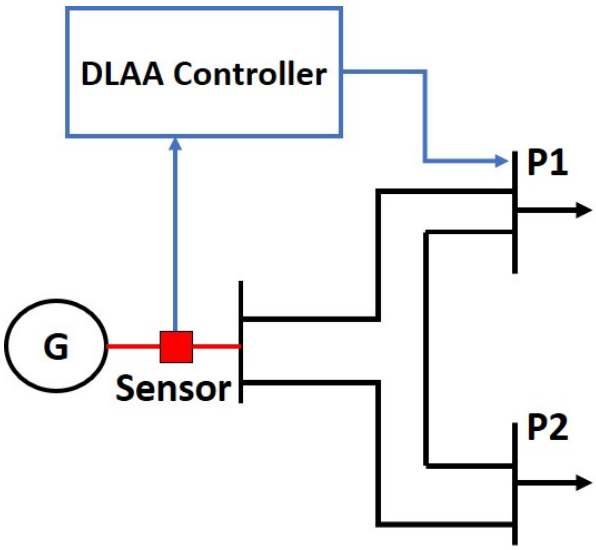

Figure 3.1: Single-point closed-loop D-LAA

Attack Model

In power systems, theoretically, the power flow between buses i and j is a nonlinear function of bus voltages and the impedance of transmission lines. The active power flow can be given as follows:

$$P_{ij} = V_i V_j \left[G_{ij} cos(\phi_i - \phi_j) + B_{ij} sin(\phi_i - \phi_j)\right] \tag{3.1}$$

where V is the voltage magnitude, ϕ is the phase angle in the corresponding bus, and G and B are the real and imaginary parts of the impedance, respectively. Note that G_{ij} can be considered as zero because the resistance of the transmission lines is significantly less than the reactance in practice. Furthermore, the difference of voltage phase angle between two buses is small and the voltage magnitude in each bus is very close to unity in the per-unit system. Thus, further approximation for the power flow equation can be written as:

$$P_{ij} = B_{ij}(\phi_i - \phi_j). \tag{3.2}$$

Specifically, consider a power system with $\mathcal{G}$ generator buses and $\mathcal{L}$ load buses. Let then

$\mathcal{N} = \mathcal{G} \cup \mathcal{L}$ represents the set of all buses in this grid. For a bus $i \in \mathcal{N}$, the total amount of power flow can be separated into the power injection of the generator P_i^G at bus $i \in \mathcal{G}$ and power absorbed by load P_i^L at bus $i \in \mathcal{L}$. Defining δ_i as the voltage phase angle of the i-th generator bus, θ_i as the voltage phase angle at i-th load bus and B_{ij} as the admittance value between buses i and j, the linearized power flow equations based on equation (3.2) can be written as:

$$P_i^G = \sum_{j \in \mathcal{G}} B_{ij}(\delta_i - \delta_j) + \sum_{j \in \mathcal{L}} B_{ij}(\delta_i - \theta_j), \tag{3.3}$$

$$-P_i^L = \sum_{j \in \mathcal{G}} B_{ij}(\theta_i - \delta_j) + \sum_{j \in \mathcal{L}} B_{ij}(\theta_i - \theta_j). \tag{3.4}$$

To model the dynamic behavior of each generator, the swing equations are used for the generator bus:

$$\dot{\delta}_i = \omega_i, \tag{3.5}$$

$$M_i \dot{\omega}_i = P_i^M - P_i^G - D_i^G \omega_i, \tag{3.6}$$

where ω_i is the rotor angular frequency deviation of generator bus i, M_i is the rotor inertia of each generator, P_i^M is the mechanical power input and D_i^G represents the damping coefficient. Note that P_i^M and D_i^G must be positive.

Specifically, the turbine-governor controller and the load-frequency controller can be integrated together as a proportional-integral (PI) controller, aimed at maintaining the rotor angular frequency at its nominal level to affect the mechanical power input [65]. The PI controller is represented as:

$$P_i^M = -(K_i^P \omega_i + K_i^I \int_0^t \omega_i), \ K_i^P, K_i^I > 0, \tag{3.7}$$

where K_i^P and K_i^I are the proportional and integral controller coefficients, respectively. As

a result, the rotor frequency dynamics in equation (3.6) can be rewritten by expressing the mechanical power for each generator in terms of frequency deviation ω_i, as defined in equation (3.7). It becomes:

$$M_i\dot{\omega}_i = -(K_i^P\omega_i + K_i^I\int_0^t \omega_i) - P_i^G - D_i^G\omega_i. \tag{3.8}$$

According to equation (3.3), we obtain:

$$M_i\dot{\omega}_i = -(K_i^P + D_i^G)\omega_i - K_i^I\delta_i - \sum_{j\in\mathcal{G}} B_{ij}(\delta_i - \delta_j) - \sum_{j\in\mathcal{L}} B_{ij}(\delta_i - \theta_j). \tag{3.9}$$

In this way, expressions (3.5), (3.4), (3.9) formulate the complete dynamical model and can be written as the following linear state-space descriptor system:

$$\begin{bmatrix} I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & M \end{bmatrix} \begin{bmatrix} \dot{\delta} \\ \dot{\theta} \\ \dot{\omega} \end{bmatrix} = \begin{bmatrix} 0 & I & 0 \\ B^{LG} & B^{LL} & 0 \\ -(K^I + B^{GG}) & -B^{GL} & -(K^P + D^G) \end{bmatrix} \begin{bmatrix} \delta \\ \theta \\ \omega \end{bmatrix} + \begin{bmatrix} 0 \\ P^L \\ 0 \end{bmatrix}, \tag{3.10}$$

where B is the imaginary part of the admittance matrix:

$$B_{bus} = \begin{bmatrix} B^{GG} & B^{GL} \\ B^{LG} & B^{LL} \end{bmatrix}. \tag{3.11}$$

Now, we consider a single-point closed-loop D-LAA that is performed at victim load bus v and the frequency sensor is installed at a generator bus s aiming to push this particular generator off-line. Suppose a proportional-integral controller is used by the attacker, creating a large deviation while less load is needed. Let K_p^L and K_I^L denote the attack controller's proportional and integral gains at the generator bus (sensor bus) s, respectively. We can

write the compromised power consumption level $\bar{P}_v^L$ at victim bus v:

$$\bar{P}_v^L = P_v^L - K_p^L \omega_s - K_I^L \int_0^t \omega_s. \tag{3.12}$$

As a result, the system dynamics subjects to the above D-LAA becomes

$$\begin{bmatrix} I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & M \end{bmatrix} \begin{bmatrix} \dot{\delta} \\ \dot{\theta} \\ \dot{\omega} \end{bmatrix} = \begin{bmatrix} 0 & I & 0 \\ B^{LG} - K_I^L & B^{LL} & -K_P^L \\ -(K^I + B^{GG}) & -B^{GL} & -(K^P + D^G) \end{bmatrix} \begin{bmatrix} \delta \\ \theta \\ \omega \end{bmatrix} + \begin{bmatrix} 0 \\ P^L \\ 0 \end{bmatrix}. \tag{3.13}$$

From equation (3.13), when the system is under attack, the attacker can affect the system dynamics and compromise the system stability by adjusting the attack controller matrices K_p^L and K_I^L. In particular, the system becomes unstable if the attacker is capable of moving the system poles to the right-half complex plane assuming the vulnerable load is large enough. Considering generators are generally equipped with various relays in the modern power system [66], a single-point closed-loop D-LAA may push the generator off-line and disconnect it from the grid.

Control Scheme of Closed-loop D-LAA

It is worth mentioning that PI control in (3.7) is not the only option to successfully implement D-LAA. In general, the closed-loop D-LAA can be viewed as a frequency controller which makes the compromised loads react to frequency deviants in the opposite direction of the normal demand response for frequency regulation. It has been discussed in [46] that "*the attacker may use a bang-bang, P, PI, or PID controller, or any other more complex feedback control system mechanism*" for the closed-loop D-LAA, and a P control model is used in [46] to formulate the D-LAA. In this work, the P control based D-LAA attack model in [46] is extended to a more general PI control based model. The models provide effective methods

for the modeling of malicious D-LAA actions against the frequency control of the grid. The proposed model keeps the direction of D-LAA actions opposite to the normal frequency control of the grid. Furthermore, the focus of this work is not to design attack controller but rather to develop a dynamic defense framework plan by minimax-q learning against such multistage attacks. This method and idea are not affected by the controller selection and can potentially be extended to other multistage attacks.

D-LAA Implementation

The D-LLA is launched by altering the remotely controllable demands instead of the outputs of the generation units in the grid. Referring to Fig. 3.1 the adversary only needs to hack into the remote load control systems to adjust the power consumption trajectory by constantly monitoring the frequency signals to implement D-LAA. Such remote load control systems extensively exist in demand response programs. Specifically, an attacker may aim to compromise command signals in Direct Load Control (DLC) programs that often involve two-way communications between the power system operator and loads or aggregators [67]. The adversary may utilize the vulnerability in any of these communications infrastructures to gain direct and remote access and control over the load through the load control mechanism. These loads that are potentially vulnerable to D-LAA attack include air conditioners [68], building lighting system [69], water heaters [70] and electric vehicles [71]. For example, considering the heating, ventilation and air conditioning (HVAC) demand in buildings, after intruding into the communication between the building and grid operator, the attacker could generate desired aggregated load profile by orchestrated periodic on/off signals to each component, e.g., air conditioner and fan.

To set the parameters in the controller-based model of D-LAA, the attacker needs to know or estimate the system dynamic model including the system frequency control settings and grid topology, which do not change frequently and are considered constant during the attack.

The only signal that needs to be updated in real time by the attacker when implementing the closed-loop D-LAA is the frequency signal. Thus, following the work in [46], it is assumed in this research that the attacker can constantly monitor the frequency signals via the attacker's installed sensors or by hacking into an existing monitoring infrastructure of the grid.

3.4.2 Optimal Load Shedding

As mentioned, when a system is attacked and the topological structure of the system is changed, such as a generator being off-line, a transmission line being tripped by relay/man or a system partition being caused, load shedding must be performed to regain stability. Considering a power system with n buses the optimal load shedding problem can be formulated as a constrained optimization problem with the physical constraints of the power flows [58, 72, 73]:

$$\min_{z_g, z_l} = \sum_{i=1}^{n} w_{li} z_{li}, \tag{3.14}$$

subject to,

$$\Lambda' B \Lambda \phi - (p + z) = 0, \tag{3.15}$$

$$p_{gmin} \leq p_g + z_g \leq p_{gmax}, \tag{3.16}$$

$$z_{gmin} \leq z_g \leq z_{gmax}, \tag{3.17}$$

$$p_{lmin} \leq p_l - z_l \leq p_l, \tag{3.18}$$

$$\phi_{min} \leq \Lambda \phi \leq \phi_{max}, \tag{3.19}$$

where $w_l = [w_{l1}, w_{l2}, ..., w_{ln}]^T$ is the weight vector representing the relative importance of different load buses; vector $z = [z_g; z_l]$, in which vector z_g refers to the re-dispatched power at each generation bus; vector z_l is the load to be shed at each load bus; vector $p = [p_g; p_l]$, in which vector p_g represents the original active power output at each generation bus; vector p_l is the demand at each load bus; vector ϕ represents the phase angle at each bus; Λ

is the incidence matrix for the topology of the grid; and B is the diagonal matrix of the transmission-line admittances. Constraint (3.15) represents the power balance at each bus; constraint (3.16) is the power output limit of the generation; constraints (3.17) and (3.18) are the constraints of the generation redispatch and load shedding respectively; Constraint (3.19) limits the phase angle difference of the connected buses of each transmission line in the grid.

3.4.3 Cascading Failures and D-LAA in Sequence

A successful single D-LAA with the aim to disconnect a generator may cause cascading failures during the post-attack stage. Due to the excessive load demand after attack, load shedding is an inevitable option for the system operator [42]. The optimal load shedding technology has been discussed and presented in Section 3.4.2. After the load shedding is carried out, a DC power flow analysis is performed to check for overloads on the transmission lines. If a transmission line is overloaded by over 50%, it will be tripped by the operator. Then the balance between the generation and demand is checked again and these steps are repeated until entering into the next steady state.

On the other hand, for causing a more severe damage to the power system such as more load shedding or generation losses, the attacker may perform the D-LAAs in sequence (one-by-one). For example, the attacker may perform a D-LAA to force a generator to be disconnected from the grid and trigger cascading failures. Then, based on the current state and system topology of the post-attack stage, the new proper victim and sensor buses are selected and another D-LAA can be performed. The attacker may repeat this process until the attack goal is achieved.

There is a main concern for the D-LAA sequence: *how to choose the best attack controller gain for each step?* For the ease of analysis, it is assumed the frequency sensor is placed

at the generator bus, that is, the sensor bus is always one of the generator buses, and all portions of loads are controllable at each vulnerable load bus. As mentioned previously, the attacker may destabilize the system by changing the controller gain matrix K_P^L. From the control perspective, the locations of system poles change with the increase of K_P^L and once the pole(s) are moved to right-half plane the system becomes unstable. Fig. 3.2 shows how the root locus analysis helps the attacker find the minimum attack gain.

The minimum vulnerable load that must be compromised can be calculated by equation (3.12). If the minimum amount of load is not larger than the total load at this load bus, the selections of victim load bus and sensor bus are feasible. The attacker tries to make the least effort to achieve the attack goal, so for the same sensor bus when there are two feasible victim buses the attacker tends to choose the one with less minimum compromised load. Once a D-LAA is successfully performed and the cascading failures are triggered, the system enters into the next steady state and the attacker may choose the new feasible victim and sensor buses to conduct the next attack. In simulations, we can change the entries of matrix B in equation (3.13) based on the current system topology because $B_{ij} = 0$ if the transmission line between buses i and j is tripped.

3.5 Game-theoretic Analysis of Attack-defense Interactions

In this section, the behaviors of the attacker and defender are modeled using a two-player zero-sum multistage game. As introduced in Section 1, game theory helps people understand the interactions between the decision-makers. For the analysis of power system security, the attacker and defender are considered as two decision-makers or players. The attacker can be hackers, organized terrorists or other criminals. The defender is the system administrator who monitors the power system network and implements security measures. The attacker

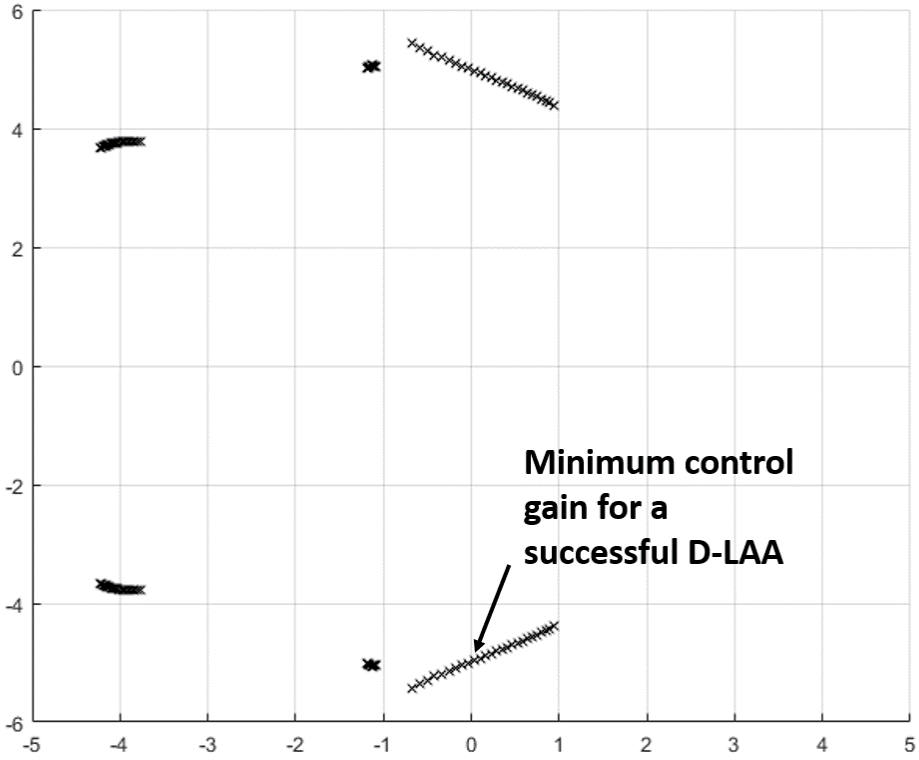

Figure 3.2: Root locus plot of power system under a D-LAA attack

intends to cause the maximum damage to power grid while the defender strives to minimize the impact. Thus, the defender's gain is regarded as the opposite of the attacker's gain. In the attack, the adversary may compromise components in sequence instead of at the same time, in order to cause more damage and decrease the risk of being detected. Similarly, the defender has to change defense actions with a dynamic attacker. Thus, both the attacker and defender have to adjust their actions based on the observation of their past actions and current states. In this way, the attack-defense game falls exactly into the category of two-player zero-sum games.

This game can be considered as a 5-tuple ($\mathcal{S}^{\mathcal{S}}$, $\mathcal{A}^{A}$, $\mathcal{A}^{D}$, $\mathbf{R}^{A}$, $\mathbf{R}^{D}$) Markov game, where

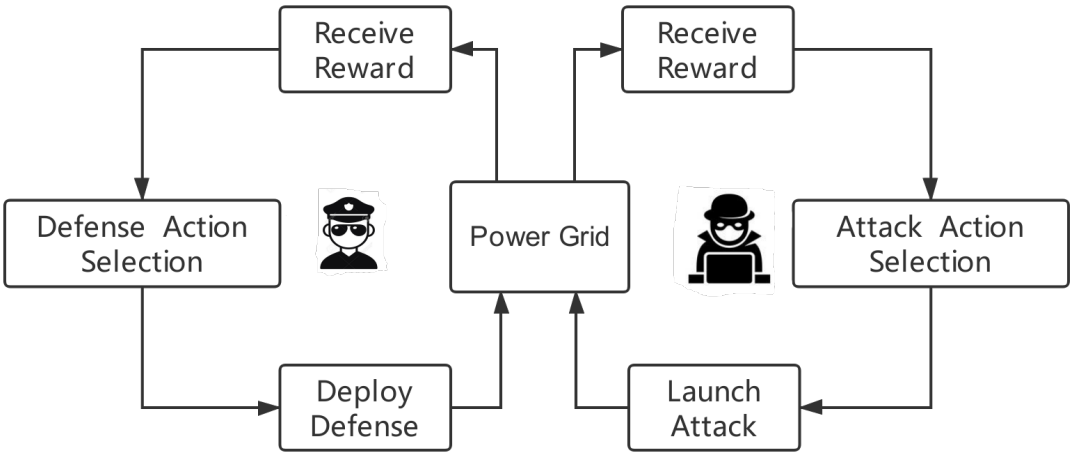

Figure 3.3: The interaction between players and the system

- $\mathcal{S}^S \stackrel{def}{=\!=} \{s_1, ..., s_{Ns}\}$ denotes the system's state space;
- $\mathcal{A}^A \stackrel{def}{=\!=} \{a_1, ..., a_{N_A}\}$ denotes the attacker's action space;
- $\mathcal{A}^D \stackrel{def}{=\!=} \{d_1, ..., d_{N_D}\}$ denotes the defender's action space;
- $\mathbf{R}^A = [R^A_{a,d}(s)]_{N_A \times N_D}$ denotes the attacker's expected reward associated with attack action $a \in \mathcal{A}^A$ against defense action $d \in \mathcal{A}^D$ in state $s \in \mathcal{S}^S$; and
- $\mathbf{R}^D = [R^D_{a,d}(s)]_{N_A \times N_D}$ denotes the defender's expected reward associated with defense action $d \in \mathcal{A}^D$ against attack action $a \in \mathcal{A}^A$ in state $s \in \mathcal{S}^S$.

Fig. 3.3 illustrates a typical player-system interaction for the two-player game. The attacker obtains system state s and takes the attack action a, and will receive reward R_A. Meanwhile, the defender will conduct the same process and receive reward R_D.

3.5.1 Action Spaces

Attacker's target is to implement D-LAA aiming to disconnect one generator and cause cascading failures. Attacker's action $a \in \mathcal{A}^A$ means trying to force one generator disconnecting from the main grid at a time step. The defender's action is related to protect a generator bus. However, the defender can restrain this attack on generators by protecting the load on corresponding victim buses. As discussed in Section II-B, a successful D-LAA needs a victim bus and a sensor bus. The defender may follow the same method in Section II-B to obtain all vulnerable loads that can be potentially controlled by the attacker. Thus, protecting the victim load is an effective method against D-LAA. That is, the physical meaning of protection action is to protect the potential victim load rather than protecting these generators. Currently, the load can be protected by implementing reinforced security measures, e.g., adding hardware and software based security components, at both the communication level [74] and device level [75, 76]. For example, reference [74] proposes a method in which the administrator can temporarily revoke the certificate of some nodes. In this way, these nodes are excluded from the grid's communication network, that is, the attacker is not able to remotely alter these loads.

In this research, it is assumed that when a defense action is performed on a load bus, the load at the protected bus P_v^L cannot be manipulated by the attacker. Note that not all loads can be targeted by attacker to implement D-LAA because some loads are traditional types which cannot be remotely manipulated.

3.5.2 System States

This game is played over a finite state space denoted by $\mathcal{S}^S$. States are formulated as a combination of the statuses of all transmission lines of the power grid. For each state s, the

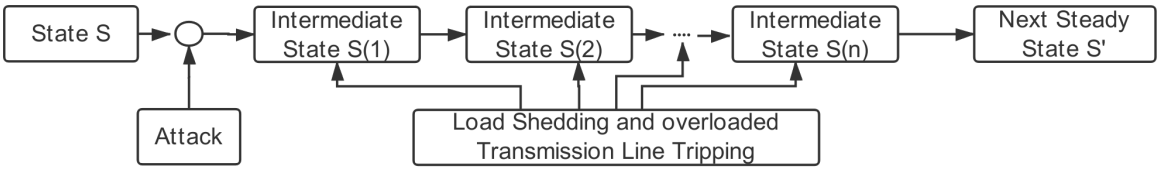

Figure 3.4: Transition from one state to the next steady state

status of the transmission lines is represented by a binary number '1' or '0'.

$$S_t(l) = \begin{cases} 1 & \text{if line } l \text{ works properly} \\ 0 & \text{if line } l \text{ is out-of-service,} \end{cases} \tag{3.20}$$

Fig. 3.4 shows the transition process from one state to the next steady state through some intermediate states. At state s, attack (D-LAA in this research) is launched and one generator is pushed to be disconnected from the grid due to instability. The load is shed based on the model in Section 3.4.2 and the demand is balanced. Then, a DC power flow analysis is applied to decide if any overloads occur on the transmission lines. Generally, the transmission line is tripped if the overload exceeds 50%. The system repeats this process until the generation and demand is balanced and there is no overloaded transmission line. Please see Section 3.4.3 for more details.

3.5.3 Attacker and Defender's Policies and Rewards

There are two players in the game: the attacker and the defender. At state s, the players choose their respective actions $a \in \mathcal{A}^A$ and $d \in \mathcal{A}^D$ independently, and immediately receive rewards $R^A_{a,d}(s)$ and $R^D_{a,d}(s)$, respectively. In this zero-sum game, the defender's expected reward is opposite to the attacker's expected reward, denoted by $R^A_{a,d}(s) = -R^D_{a,d}(s)$. The

rewards for players are assigned following the conditions given as:

$$R^A_{a,d}(s) = \begin{cases} 0 & \text{if load shedding } z(t) = z(t-1) \\ 1 & \text{if } z(t-1) < z(t) < N \\ 10 & \text{if } z(t) \geqslant N, \end{cases} \tag{3.21}$$

and

$$R^D_{a,d}(s) = \begin{cases} 0 & \text{if load shedding } z(t) = z(t-1) \\ -1 & \text{if } z(t-1) < z(t) < N \\ -10 & \text{if } z(t) \geqslant N, \end{cases} \tag{3.22}$$

where z represents the total load shedding caused by the D-LAA attack and N is the attack objective.

Now we have specified the immediate rewards of the attacker and the defender at each state, but have not indicated how these rewards are aggregated into an overall payoff. The most commonly used aggregation method is the discounted-sum reward. For an attack action a and a defense action d, the discounted-sum rewards of the attacker and defender considering deterministic state transition are represented as:

$$Q_A(s, a, d) = \sum_{t=0}^{\infty} \gamma^t R^A_{a,d}(s(t)), \tag{3.23}$$

$$Q_D(s, a, d) = \sum_{t=0}^{\infty} \gamma^t R^D_{a,d}(s(t)), \tag{3.24}$$

where Q_A and Q_D represent game values for the attacker and the defender, respectively; and $\gamma \in (0, 1)$ is the discount factor. A smaller value of γ implies the agent emphasizes the immediate reward while a larger value indicates more concerns about future rewards. For a given state s, the attacker's strategy is defined as probability distributions over action space $\mathcal{A}^A$, i.e.,

$$\pi^A(s) = [Pr(a(s) = a_1), Pr(a(s) = a_2), ..., Pr(a(s) = a_{N_A})]^T, \tag{3.25}$$

which satisfies $\sum_{i=1}^{N_A} Pr(a(s) = a_i) = 1 \mid a_i \in \mathcal{A}^A$. Similarly, the defender's strategy is given as:

$$\pi^D(s) = [Pr(d(s) = d_1), Pr(d(s) = d_2), ..., Pr(d(s) = d_{N_D})]^T \tag{3.26}$$

and $\sum_{i=1}^{N_D} Pr(d(s) = d_i) = 1 \mid d_i \in \mathcal{A}^D$.

When only one entry of the strategies described above is nonzero (and equal to 1), π^A and π^D are called pure strategy and players always adopt this action at state $s(t)$. Otherwise, they are mixed strategies which are adopted in this research. Note that in this multi-stage game, the attacker and defender choose their different targets in time sequence until the attack objective is achieved.

3.5.4 Nash Equilibrium

In this game, the defender tries to minimize the discounted sum of expected reward Q_D while the attacker aims to maximize it. *Nash equilibrium* is a common solution to solve the players' optimal strategies for such a Markov game [48, 77]. Nash equilibrium is a state that no player has a unilateral incentive to change actions as that would reduce their rewards, that is, each agent plays best response to their opponents. For the proposed game model, a Nash equilibrium can be mathematically defined as follows:

Definition 1 *In the proposed zero-sum two-player Markov game, a Nash equilibrium is a pair of mixed optimal strategies* (π_A^*, π_D^*) *for all mixed strategies* π_A *and* π_D *for all states* $s \in S$

$$Q_A(s, \pi_A^*, \pi_D^*) \geqslant Q_A(s, \pi_A, \pi_D^*), \tag{3.27}$$

$$Q_D(s, \pi_A^*, \pi_D^*) \geqslant Q_D(s, \pi_A^*, \pi_D). \tag{3.28}$$

For such a two-player game, it is proved that unique Nash Equilibrium exists in stationary strategies (for all t) by Shapley [78]. That is, the mixed optimal attack/defense strategies can be solved for each state instead of each time t. In general, the stationary optimal strategy can be solved recursively by dynamic programming if environment of the model is known such as in [58, 59].

3.6 Proposed Solution Approach

3.6.1 Minimax-q Learning

In this section, we propose a new dynamic defense solution for the two-player zero-sum Markov game based on the minimax-q learning approach. Our objective is to characterize the attacker's and the defender's Nash equilibrium strategies for each state $s \in \mathcal{S}^S$ and their attack/defense actions in time sequence, where all players are rational and tend to maximize their own benefits. The attacker and the defender are completely competitive and do not cooperate with each other. Minimax-q learning [79] is used in conjunction with Markov games. As a modification of q-learning which just considers the opponent as part of the environment, this algorithm treats the Q function not just from the state/action pairs to values, but from the state/action/action to values, i.e., $Q(s, a, d)$. Thus, both players' actions and their interactions are modeled more explicitly. The minimax-q learning algorithm is adopted to approach real unknown *state value function* $V(s)$ by interacting with the environment and then players obtain the optimal Nash strategies by the learned state value function. Furthermore, a *state-action value function* (Q function) is to quantify the performance for a player to apply a particular action following a policy π in a state. From the defender's perspective, the Q function can be defined as:

$$Q_D(s, a, d) = (1 - \alpha_t)Q_D(s, a, d) + \alpha_t(R^D + \gamma \sum_{s' \in S} V_D(s')), \tag{3.29}$$

and the state value function $V(s)$ is defined as:

$$V_D(s') = \min_{\pi_D} \max_{a} \sum_{a} Q(s', a, d)\pi_D(s'), \tag{3.30}$$

where α_t denotes the learning rate for adjusting the step size. To improve the convergence rate of this algorithm, a polynomial learning rate is adopted as $1/t^{\beta}$ where $\beta \in (1/2, 1)$.

Note that in equation (3.30), minmax is adopted to find the best response instead of playing actions with the highest Q in [60]. Equation (3.30) can be converted to a linear constraint optimization problem to obtain the optimal strategy at state s:

$$\begin{aligned} &\min_{\pi_D} \ V_D(s), \\ s.t.\ &V_D(s) \geq \sum_{d} Q(s, a, d)\pi_D(s), \ \forall a \in \mathcal{A}^A. \end{aligned} \tag{3.31}$$

Similarly, the attacker's state value function and Q function can be dually derived:

$$Q_A(s, a, d) = (1 - \alpha_t)Q_A(s, a, d) + \alpha_t(R^A + \gamma \sum_{s' \in S} V_A(s')), \tag{3.32}$$

$$V_A(s') = \min_{\pi_A} \max_{d} \sum_{d} Q(s', a, d)\pi_A(s'). \tag{3.33}$$

The optimal strategy of the attacker can also be obtained by linear programming in (3.33):

$$\begin{aligned} &\max_{\pi_A} \ V_A(s), \\ s.t.\ &V_A(s) \geq \sum_{a} Q(s, a, d)\pi_A(s), \ \forall d \in \mathcal{A}^D. \end{aligned} \tag{3.34}$$

The procedure to compute the Nash equilibrium at each state and the attack/defense sequence are detailed in Algorithm 3.

Algorithm 1 Minimax-q Learning Algorithm

Initialize $Q_0(s, a, d)$, $V(s)$, π_A, and π_D
Obtain feasible D-LAA target for initial state discussed in Section 3.4.3 as action space for attack/defense
Define exploration probability ϵ and learning rate α
for number of episodes **do**
  while Attack objective is not reached **do**
    Select current state s
    if Generated random number $< \epsilon$ **then**
      Take random attack and defense action
    else
      Take attack and defense action based on Q-table
    end if
    Execute actions
    Calculate load shedding by (3.14), overloads, and cascades
    Determine next s'
    Assign reward by equations(3.21) and (3.22)
    Update state-action value function Q by equations (3.29) and (3.32)
    Solve state value functions (3.30) and (3.33) by linear programming and update $V(s)$ and $\pi_A(s)$ $\pi_D(s)$
    Update feasible D-LAA target for attack/defense's action space
    Update $s = s'$
  end while
end for
Find optimal strategies and sequences of actions for attacker and defender

In the proposed algorithm, the game starts with the initialization of Q function, state value function V and attacker/defender's policy. Then the system is evaluated to obtain feasible D-LAA attack discussed in Section 3.4.3. Note that for simulations, instead of observing the root-locus plot, we can analytically obtain the minimum compromised load by gradually increasing controller K_P^L until the system is unstable. In the beginning, the initial state is assumed at the normal operation condition, that is, all transmission lines and generators are active and work properly. A ϵ-*greedy* strategy is also adopted to balance the exploration and exploitation [80]. With ϵ-*greedy*, the agent plays a random action with a probability $0 < \epsilon < 1$, instead of making the best decision given in the Q-function. With the execution of the actions, a certain generator is disconnected from the power grid due to the D-LAA if the corresponding load is not protected by the defender. Load shedding of the current state is calculated and cascading overloads on the transmission lines may be triggered until the system enters into the next steady state s'. Instant rewards are assigned to the attacker and the defender by equations (3.21) and (3.22) and the value of Q-function is updated. Strategies $\pi_A(s)$ and $\pi_D(s)$ and state value V are solved by linear programming. Then, based on the new topology of system and state, feasible targets for the attack/defense's action spaces are decided. The game is repeated until the attack objective is reached. Ideally, if the process above (from step 4 to step 20) repeats for enough times, i.e., Q matrix is updated at each state by enough times, the players will learn the real complex relationships between the actions and outcomes. Thus, such relationships are reinforced in this process and eventually the players find their optimal Nash equilibrium strategies. Note that the optimal attack/defense sequence is not unique and in this study we evaluate the performance by computing the average impact to the system.

Furthermore, this defense strategy is not a real-time one but is more like a pre-stipulated plan against low-probability high-impact attacks, such us D-LAA in this research, to minimize the damage. According to the features of D-LAA attacks, the defender can find an optimal policy for each state against potential vicious attacker if both play rationally. The

defense strategy can be deployed in advance when such attacks are anticipated to improve the resilience of the power system.

In this chapter, the proposed work focuses only on the D-LAA scenario. Nevertheless, because the proposed defense method is based on minimax-q learning which works with Markov game, this work can be extended to other multistage attacks, e.g., Load Redistribution (LR) attacks and line switching attacks, as long as the action spaces and cascading failures are redesigned following specific attack mechanisms. The power grid operator may make multiple such stored plans based on different types of attacks and defense actions.

3.6.2 Discussion on Computational Complexity

The computational complexity is $\mathcal{O}(S^2 M_A M_D)$ per iteration in Algorithm 1, where M_A and M_D are the numbers of strategies for the attacker and the defender, respectively. Because single-point D-LAA is considered in this research, i.e., the attacker and the defender select one bus to compromise and protect at one time, $M_A = \begin{pmatrix} 1 \\ A \end{pmatrix}$ and $M_D = \begin{pmatrix} 1 \\ D \end{pmatrix}$, where A and D represent the numbers of total possible attack and defense actions, respectively. It can be seen that the computational complexity increases linearly with more attack/defense options. As for S, more possible states will cause relatively quicker increase of the computational complexity.

3.7 Simulation Results and Analysis

Now we evaluate the performance of the proposed minimax-q learning for this two-player zero-sum Markov game on the IEEE 39-bus system that consists of 46 transmission lines and 10 generators. The results of dynamic defense strategy may provide useful insight for grid operators to improve the resiliency of power systems. Comparisons with the existing passive

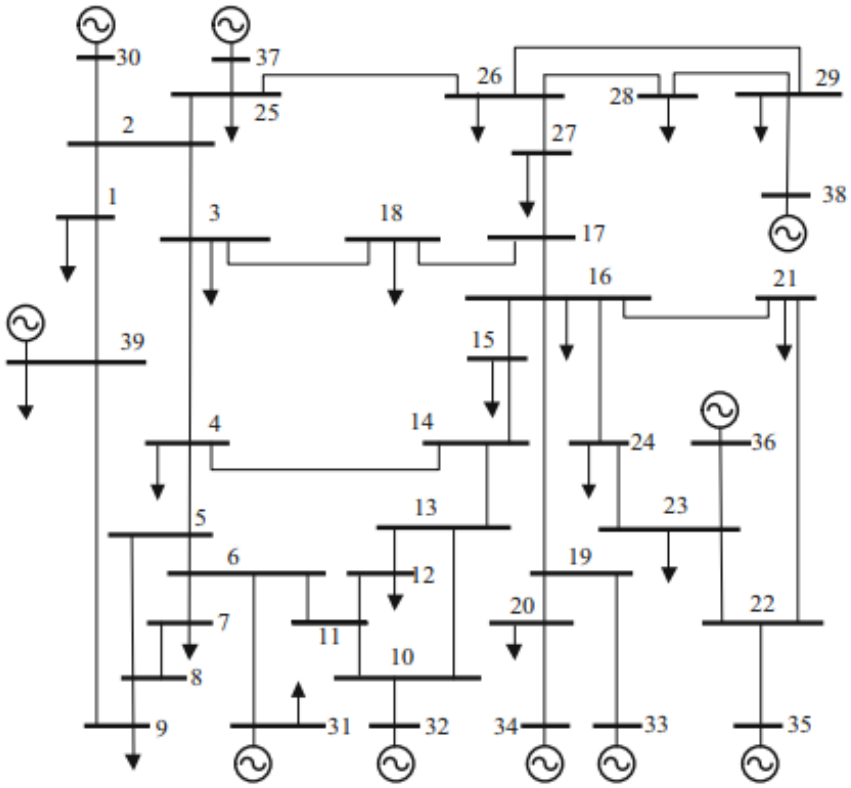

Figure 3.5: IEEE 39-bus system

and dynamic defense strategies are conducted to illustrate the importance of deploying the proposed dynamic strategy against D-LAA.

3.7.1 System Parameters

Fig. 5.4 shows the IEEE 39-bus system based on a 10-machine New-England power network. There are 10 generators, 46 transmission lines and 19 loads. There are two loops in the simulation: episodes and runs. The episodes loop is the main loop in which the attacker and the defender interact to learn the optimal policy. The attacker and the defender complete a bunch of actions in sequence. As the number of episodes increases, the attacker tends to approach the optimal policy. At the end of the episodes, the attacker and the defender

Table 3.1: Simulation parameters for IEEE 39-bus system

No.	Parameter	Value
1	Number of Generators, $\mathcal{G}$	10
2	Total Transmission lines	46
3	Discount Factor, γ	0.8
4	Learning Rate Coefficient, β	0.7
5	Initial Exploration Probability, ϵ	0.9
6	Number of Episodes	1000
7	Number of Runs	50
8	Maximum Iteration per Episode	100
9	Total Capacity	6245MW
10	Attack Objective	at least 50% load shedding

reach the Nash equilibrium point. A number of runs are conducted to deduce different Nash equilibria. Therefore, the whole game simulation is conducted for many runs. Each run includes a number of episodes. The number of episodes is the required number of trials for the agent in the learning process. The initial exploration rate ϵ is 0.9 and decreases 10% every 20 episodes to ensure the convergence. Other simulation parameters are given in Table 3.1.

3.7.2 Selection of Vulnerable Bus and Attacker/defender's Action Space

As discussed in Section II, not all loads can be considered vulnerable to D-LAAs. Some loads are traditional ones and may not even have smart meters or any demand response equipment, which the attacker cannot remotely manipulate. In this case, we assume that only eight load buses have vulnerable loads. They can potentially become victim buses, i.e., $\mathcal{V} = \{4, 6, 7, 12, 18, 19, 23, 29\}$. On the other hand, according to [81], generators $\{31, 35, 37, 38\}$ represent nuclear stations which are fully protected. Thus, frequency sensors are assumed to be placed only at $\mathcal{S} = \{30, 32, 33, 34, 36, 39\}$ that are considered as fossil and hydro

Table 3.2: Minimum portion of vulnerable load that must be compromised at initial state

Victim Bus \ Sensor Bus	30	32	33	34	36	39
4	62	92.5	79.1	69	125	46.2
6	4.9	0.91	1.2	3.7	3.6	128
7	0.72	12.4	0.6	64.5	5.1	5.1
12	73.9	23.5	48.6	77.2	89.5	89
18	146	8.5	117	222	189	46.5
19	48.1	0.77	7.4	1.5	0.62	66.8
23	280	1.9	15.6	2.8	1.9	72
29	4.6	58.5	12.7	4.7	4	0.54

stations. Thus, the attacker's action space is $\mathcal{A}^A = \{30, 32, 33, 34, 36, 39\}$. Table 3.2 shows the minimum portion of the vulnerable load that must be compromised to guarantee a successful D-LAA at the initial state. We assume K_I^L a pre-tuned parameter and there is no need to change it for simplicity of calculation. The highlighted cells indicate the attacker could launch D-LAA on the corresponding sensor bus and victim bus. For the initial state, the attacker is not able to compromise generator 34 because there are not enough loads to be manipulated for the given vulnerable buses. Therefore, for the initial state, the attacker can perform D-LAA to disconnect generators $\{30, 32, 33, 36, 39\}$. Furthermore, at each visited unique state, Table. 3.2 is updated for the next selection of the attack target. Based on the same table, the defender also decides the protection action that should be taken. The defender's action space is denoted as $\mathcal{A}^D = \{30, 32, 33, 34, 36, 39\}$. As mentioned previously, the physical meaning of the protection action is not to protect these generators but to protect the corresponding potential victim load. For example, at the initial state, when the defender selects action "30", it means the load on the corresponding victim bus 7 is protected.

3.7.3 Game-theoretic Attack/Defense

Figs. 3.6 and 3.7 show the convergence curves of the optimal number of attacker/defender's actions. After adequate learning and exploration, we can see that both players reach the

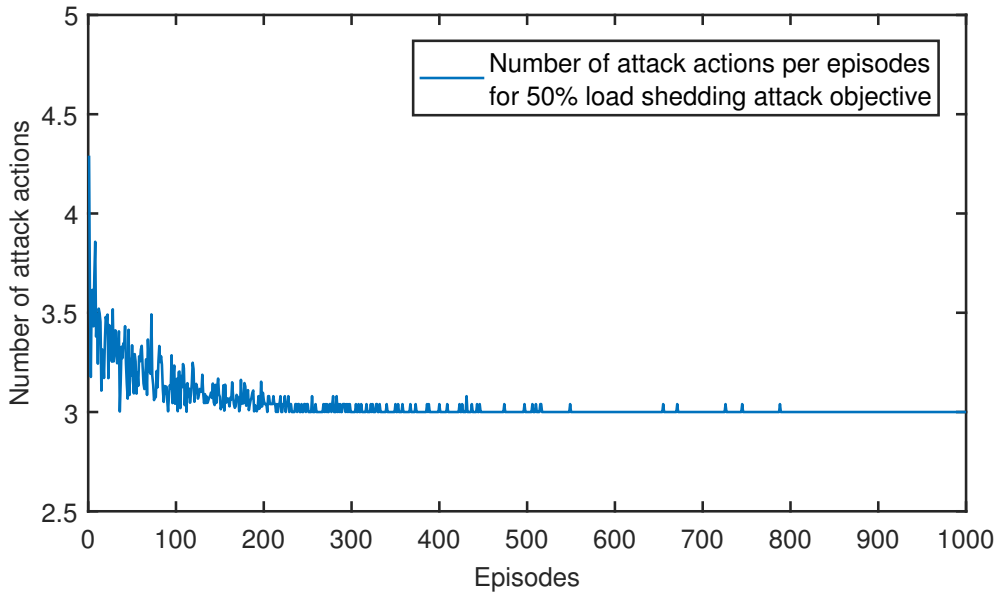

Figure 3.6: Convergence of the defender's number of actions

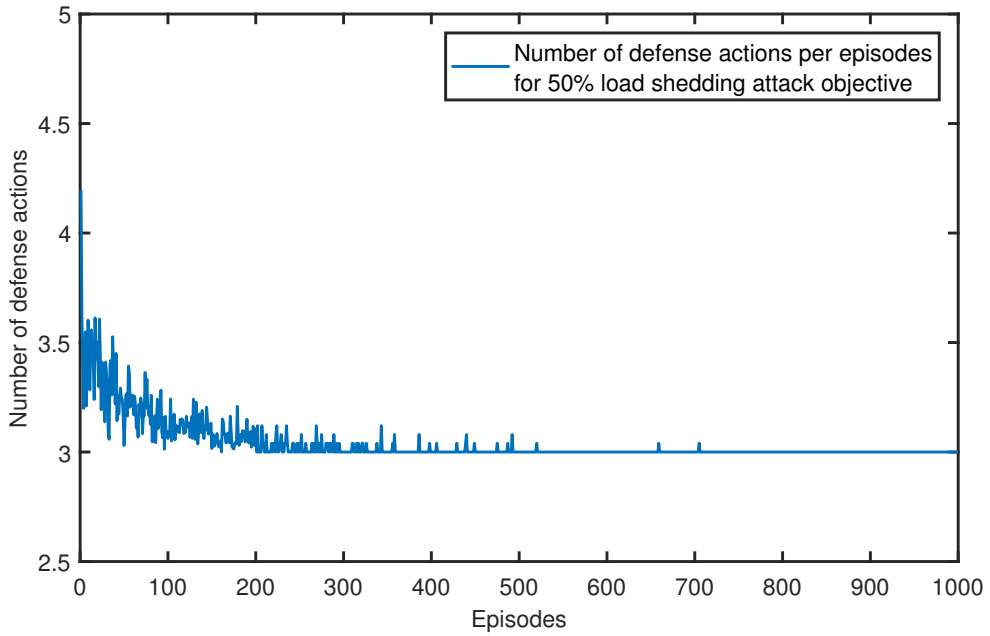

Figure 3.7: Convergence of the attacker's number of actions

optimal number of actions. Among 50 independent runs, the attacker needs three actions in sequence to achieve the objective and the defender also needs the same number of actions

to minimize the load shedding caused by D-LAAs. The average computing time per run is about 564s. We need to emphasize again that the defense strategy is not real-time but off-line trained pre-stipulated plan against the low-probability high-impact D-LAA. The defense plan can be deployed in advance when such attacks are anticipated. Therefore, the proposed strategy can adequately meet the time requirement of practical applications.

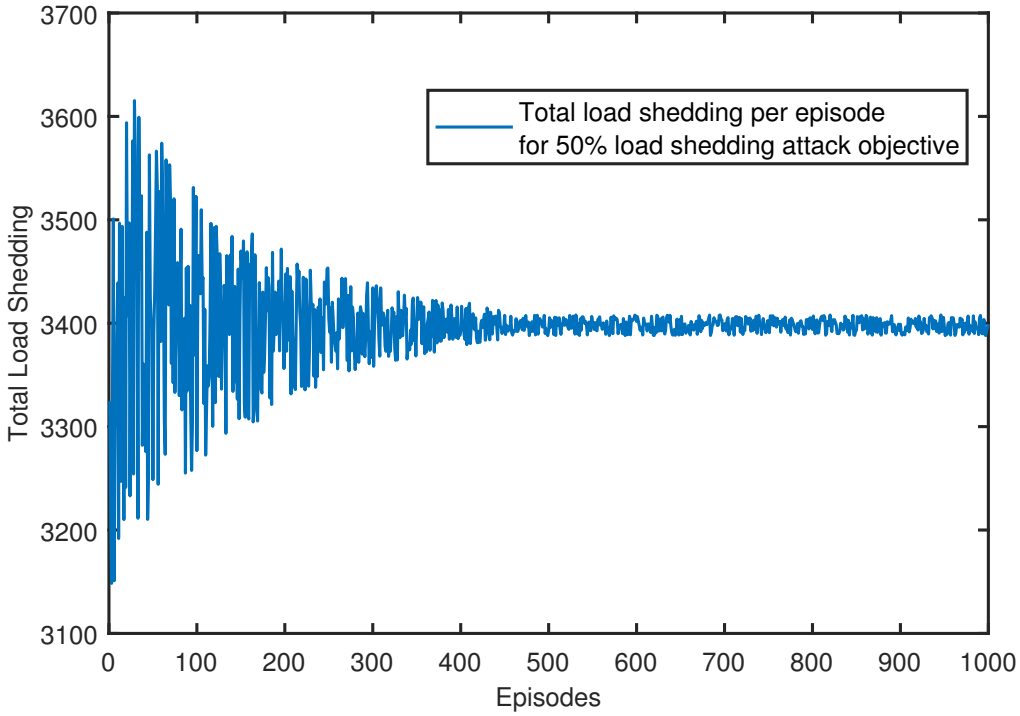

Figure 3.8: Convergence of the total load shedding

Fig. 3.8 depicts the convergence of the total load shedding. The average load shedding converges to around 3400 MW after 50 runs. From the three figures, we can portray how the algorithm works especially in the intermediate process. In the early stage of learning process, because of the large exploration rate ϵ and inaccurate Q matrix, both the attacker and defender take actions randomly or wrongly. Thus, they may take more additional steps and the total amounts of load shedding is not stable. With the decrease of ϵ and the update of Q matrix, the curves gradually converge. At the end of the process, the Q matrix is updated for enough times and the players learn the real relationships between the

actions and outcomes. The policies at each state converge to the optimal Nash Equilibrium strategies.

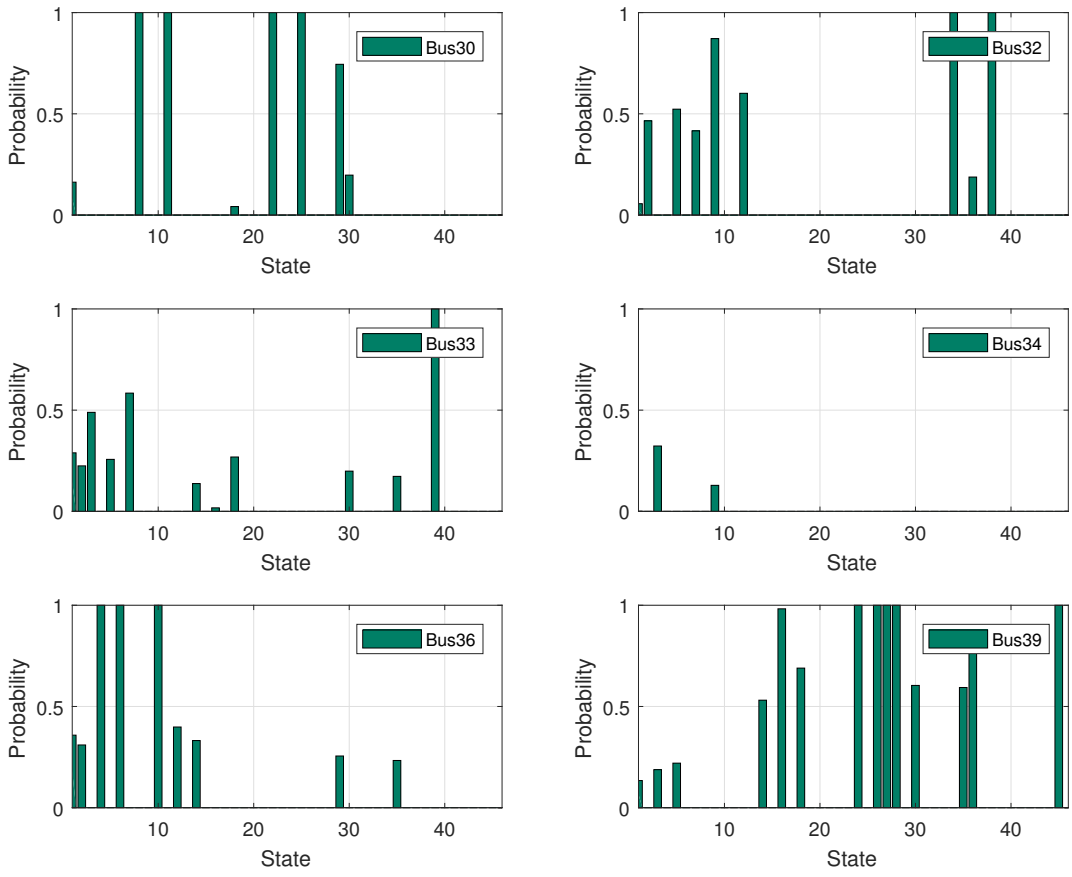

Figure 3.9: Probability of attacker's action at each state

Figs. 3.9 and 3.10 show the attacker's and defender's optimal policies when Nash Equilibrium is reached at each unique state. For this situation, both the attacker and defender have no unilateral incentive to alter their actions, because they have maximized their profits. The physical meaning of the Nash Equilibrium status for this case is that the defender can minimize the damage (load shedding) if they both play rationally their optimal strategies. The system operators are advised to adopt these strategies for each possible state against the D-LAA. Specifically, there is no need to place any defensive strategies for some states

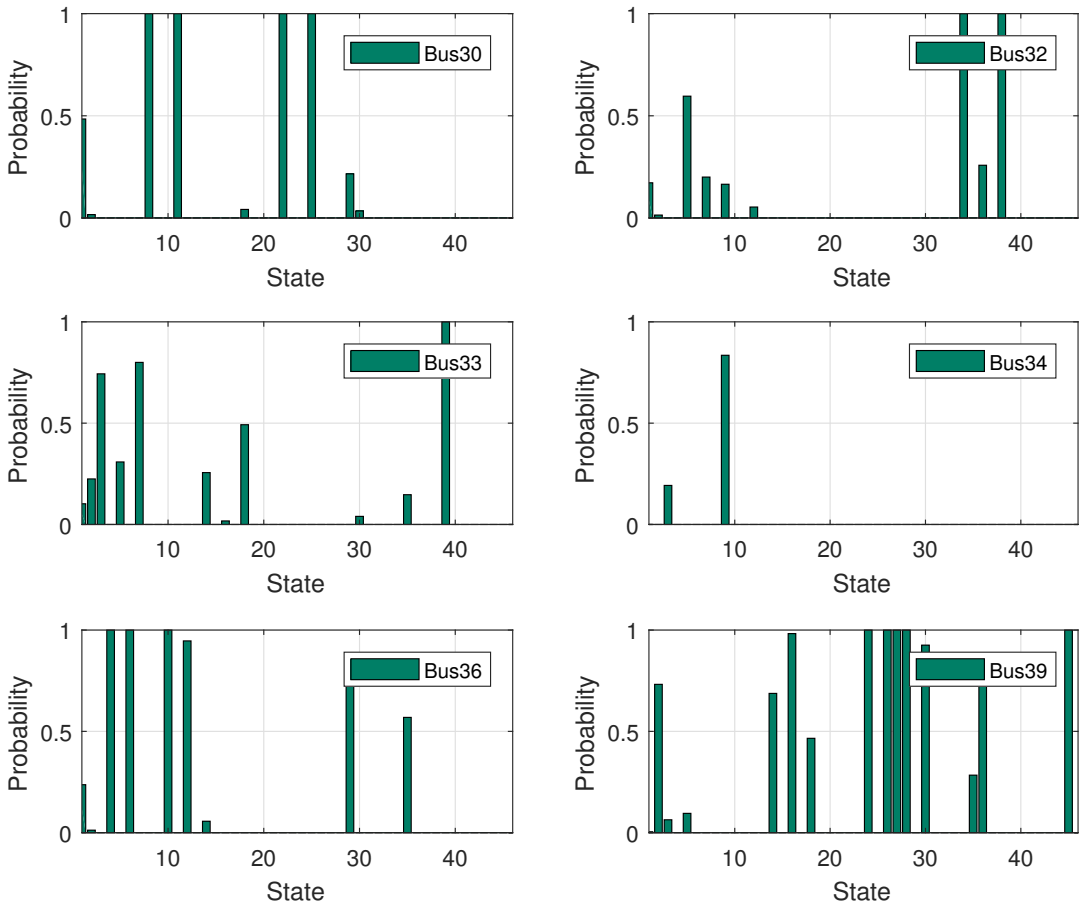

Figure 3.10: Probability of defender's action at each state

because there are not enough vulnerable loads to alter for disconnecting generators from the power grid. Note that mixed strategy at some states. Regarding the actual implementation in practice, the operator may change the defense plan according to the probabilities of the optimal policy. For instance, at state 9, the probabilities of defender's action on generators (30, 32, 33, 34, 36, 39) are (0, 0.165, 0, 0.835, 0, 0), respectively. Thus, the system administrator may plan to take protective actions for bus 34 with a probability of 0.835 and protect bus 32 with a probability of 0.165 at each interval of the actions. The results provide useful information for power system operators to thwart dynamic load altering attacks.

Table 3.3: Defender's action sequences for dynamic defense strategy

Runs	**Defense sequence**	**Physical meaning (protected buses)**	**Total load shedding (MW)**
1	(33, 39, 30)	(7, 19, 23)	3421.1
2	(30, 39, 32)	(7, 29, 4)	3709.6
3	(39, 32, 34)	(29, 12, 18)	3315.2
.....	.....	.....	.....
25	(39, 30, 32)	(29, 19, 12)	4005
26	(33, 32, 30)	(7, 29, 6)	3321.9
.....	.....	.....	.....
49	(36, 34, 33)	(19, 18, 6)	3200.2
50	(39, 32, 34)	(29, 12, 18)	3315.6

The defensive action sequences are shown in Table 3.3. For this game, total 11 unique defense sequences are found and the average load shedding can be calculated as 3398.2MW. As mentioned in Section 3.7.2, the defense action sequence indicates the protected generators identified by the defender while the actual actions of the defense strategy are to protect the corresponding victim load buses. For example, for the first run, this action sequence indicates that the defender tries to protect the loads on victim buses (7, 19, 23).

3.7.4 Comparison with Passive Defense Strategy

To illustrate the importance of dynamic defense strategy, we compare our results with the passive defense strategy in this section. For a passive strategy, the defensive actions are predefined and the attacker is trained to find the optimal attack strategy in the presence of the passive defender. Considering the limited resources the operator has, we assume only two loads can be protected at a time. In this case study, three different predefined protected load sets are considered: (7, 29), (4, 29) and (6, 7). They are denoted as passive defense case I, II and III respectively. We adopt a similar algorithm by calculating the largest value

Table 3.4: Attacker's action sequences for passive defense strategy I

Runs	**Attack action sequence**	**Total load shedding (MW)**
1	(32, 39, 36)	3856.4
2	(32, 39, 36)	3856.4
3	(32, 36, 30)	3725
.....	.....	.....
25	(32, 39, 36)	3856.4
26	(32, 39, 36)	3856.4
.....	.....	.....
49	(32, 36, 30)	3725
50	(32, 39, 36)	3856.4

Table 3.5: Attacker's action sequences for passive defense strategy II

Runs	**Attack action sequence**	**Total load shedding (MW)**
1	(32, 36, 34)	3564.7
2	(32, 34, 36)	3649.2
3	(32, 36, 34)	3564.7
.....	.....	.....
25	(32, 30, 34)	3425.7
26	(32, 34, 36)	3649.2
.....	.....	.....
49	(32, 36, 34)	3564.7
50	(32, 34, 36)	3649.2

instead of solving minimax in equation (3.33). The attack objective is to cause at least 50% load shedding.

Because the defense strategy is passive and unchangeable, we analyze the performance from the attacker's perspective. Table 3.4, 3.5 and 3.6 show the attack sequences of different runs and the total load shedding. It is found in table 3.7 that the attacker's action converges

Table 3.6: Attacker's action sequences for passive defense strategy III

Runs	**Attack action sequence**	**Total load shedding (MW)**
1	(36, 39, 30)	3992
2	(39, 32, 36)	3710.3
3	(36, 39, 30)	3992
.....	.....	.....
25	(36, 39, 30)	3992
26	(36, 39, 30)	3992
.....	.....	.....
49	(39, 32, 36)	3710.3
50	(36, 39, 30)	3992

Table 3.7: Total load shedding of different defense strategies

Proposed dynamic defense	3398.2MW
Passive defense I	3827.5MW
Passive defense II	3601.9MW
Passive defense III	3894.5MW
Dynamic defense in [82]	3709.6MW

to a sequence of three actions, and the total amounts of load shedding for the passive defense I, II and III are 12.6%, 6.0% and 14.6% more than that obtained by the dynamic defense strategy, respectively. The comparison shows the proposed dynamic defense method is more effective against the single point D-LAA.

3.7.5 Comparison with Dynamic Defense Strategy

In this section, the proposed dynamic defense strategy is compared with the dynamic strategy in [82]. In [82], the dynamic defense strategy is obtained by the pre-calculated worst-case dynamic attack, which ignores the adversarial game between the rational attacker and de-

fender, and their future expected gains. This is the main difference between the proposed models in this chapter and [82]. To compare by same standards, the attack objective is still at least 50% load shedding. The last row of table 3.7 shows that the total amount of load shedding by applying the defense plan in [82] is 9.2% higher than that obtained by the proposed strategy in this research. One reason of the result is that the outcome of two players' game, i.e., the attacker and defender, is not always the best for one of them but inclines a Nash equilibrium mentioned in Section 3.4. Thus, the defense strategy derived by the worst-case dynamic attack, i.e., unilateral optimal attack, results in the worse outcome because the interaction between two rational players in each state of the Markov game for D-LAA is not considered. In general, the proposed model formulates a more complex and realistic game considering two rational players' game, which leads to better performance for the defense against D-LAA.

3.7.6 Summary

This chapter proposes a novel reinforcement-learning-based dynamic defense solution against the single point D-LAA in power grid, where considering the attacker/defender's action sequence. We have derived the D-LAA in time sequence considering cascading failures at each state. A two-player zero-sum Markov game is formulated to analyze the complex interactions between the attacker and the defender, in which all players are rational and tend to maximize their own benefits. The proposed minimax-q algorithm is applied to derive the attacker/defender's Nash equilibrium strategies. The IEEE 39-bus system is used to test the proposed algorithm and evaluate the dynamic defense strategy against D-LAA. Simulation results are compared with the existing passive and dynamic defense strategies, which indicates the proposed dynamic strategy exhibits a better performance. The system operator is informed to enforce the optimal dynamic defense strategy at each state in advance to improve the power system resiliency.

Chapter 4

GAN BASED STEALTHY FDIA DETECTION AND MITIGATION

4.1 Overview

False data injection (FDI) attacks are emerging as a severe cyber threat to modern power systems, which manipulate the state estimation results by tampering with the measurements. In this chapter, a novel Triple Generative Adversarial Network (TripleGAN) based defense framework is developed against the stealthy FDI attacks. The proposed model aims to effectively detect and mitigate the attacks with limited historical measurement data. In this model, the detection is performed by the classifier and the mitigation is carried out by replacing the tampered measurements with the produced measurement data from the generator. The advantages are that the detection module is integrated and trained with the mitigation module together, and they reinforce the performance of each other. To improve the detection accuracy and recovery efficiency, an extended loss function integrating feature matching is proposed. Simulation results on the IEEE 118-bus system demonstrate that the proposed defense model can accurately detect the stealthy FDI attacks and reconstruct the state estimation modified by the manipulated measurements, which thus improves the power system resilience. Further, the results confirm that the proposed techniques outperform other existing machine learning detection and recovery methods.

4.2 Related Work

The stealthy FDI attack on electric power systems was first investigated in [83], where the well-constructed attack sequence could mislead the state estimation without being detected by the residual-based bad data detector (BDD). According to the survey [84], the main FDI attack detection algorithms are classified into two categories based on different approaches: model-based [27, 30, 31, 85, 86] and data-based [61, 63, 87–94] algorithms. In [85], the authors utilize recursive Weighted Least Squares (WLS) estimator to enhance the convergence speed in their detection scheme. Zhao *et al.* propose a vector autoregression (VAR) method as the prediction process that captures the interdependencies between the events in different time slots to detect the FDI attack [27]. In [30, 86], Matrix Separation (MS) is introduced to defend against the sparse FDI attack sequence based on the separation of nominal states and anomalies matrices. A new detection method is derived by calculating the metrics of Kullback-Leibler distance (KLD) to track the dynamics of the measurement variations [31]. The associated underlying idea is that if the measurements are corrupted by FDI attacks, the distribution of the measurements variation will deviate and the KLD will become larger.

Aside from the model-based methods, the data-based especially machine learning methods have been widely investigated to tackle the FDI attack detection problem. Support Vector Machine (SVM) is the most well-studied algorithm in this research area because of its simplicity [87–89]. The main drawback of SVM is the choice of the kernel function and the detection accuracy. Artificial neural network (ANN) is another prevalent method in classification and estimation. A number of different types of neural networks have been applied in the FDI detection problems, such as feedforward neural network (FNN) [90], recurrent neural network (RNN) [91, 92] and convolutional neural network (CNN) [63, 93]. In [61], a deep belief network (DBN) is proposed to enable real-time detection, which reduces the time for training the networks. An online detection algorithm is derived by the model-free reinforcement learning based on a partially observable Markov decision process [94]. To defend

against FDI attacks, a greedy approach is deployed to protect certain meter measurements based on the linearized measurement model [95]. Such a greedy approach is also used to promote the PMU deployment for defense against the FDI attacks. Yang *et al.* [96] propose a defense methodology by finding the optimal attack strategy and then protecting the critical sensor nodes. In [97,98], a graphic method is developed for defense against the FDI on state estimation. A twofold mitigation strategy is proposed in [99] using a multi-agent system. In particular, the authors present a voting protocol to arrive at collective decisions among the agents on mitigating the effects of stealthy FDI attacks. In [100], the researchers demonstrate a data analytical method employing the margin setting algorithm. A joint admittance perturbation and meter protection strategy is developed by [101], where the resiliency of the state estimation is enhanced under stealthy FDI attacks.

Recently, a new deep neural network model called Generative Adversarial Network (GAN) has attracted much attention [102]. GAN can be considered as a two-player game between the generator, which learns to produce samples following the distribution of real data, and the discriminator, which attempts to distinguish the real and produced data. The two components contest with each other in the game until the Nash equilibrium is achieved. Although there are a few GAN-based studies addressing the cyber-attacks problem on power systems [103, 104], they solely solve the attack detection or data recovering issue without deeply leveraging the advantage of GAN, where the two components of GAN reinforce each other. It implies that the attack detection and mitigation (data recovering) can be carried out in one GAN-based structure and have better performance than being implemented separately.

4.3 Research Contributions

Even with the use of machine learning techniques discussed above, detecting and defending against stealthy FDI attacks are still highly challenging. Firstly, most of the supervised

methods rely on a large amount of labeled normal and corrupted data to learn from, but in practice we can hardly acquire enough labeled corrupted data as attacks are generally rare [105]. Secondly, though unsupervised learning techniques are good solutions to handling the lack of labeled data, they are inadequate for the attack detection problem due to the concerns of result accuracy and usefulness. Furthermore, their control bounds are not flexible or effective enough for indirect attacks. At last, several GAN-based structures are proposed for this defense task [106, 107]. However, in these methods, the discriminator plays two incompatible roles: the classifier for predicting the labels and discriminator for distinguishing the generated samples. On the one hand, the discriminator should identify the fake sample produced from the generator as the fake one with non-zero probability. On the other hand, the discriminator should always treat it as a real sample and predict the correct class confidently as a classifier converges to the real data distribution. Based on empirical results, the two conflicted convergence points indicate that the generator and the classifier (i.e., discriminator) may not simultaneously converge to the real data distribution. Thus, these challenges motivate us to develop a new model for defending against the stealthy FDI attacks on power systems.

This chapter presents a novel defense methodology to perform attack detection and mitigation by utilizing Triple Generative Adversarial Networks (TripleGAN) [108], which extends the original GAN to semi-supervised learning by employing a classifier. The main focus of this framework is to accurately detect the stealthy attacks and reconstruct the corrupted data at the same time with limited historical labelled measurement data. The classifier and generator are the key components in the proposed model (refer to Fig. 4.2). Specifically, the classifier can determine whether the measurement data are compromised and the generator collectively provides a reconstructed data set in case the attack occurs. Moreover, compared with the TripleGAN [108], the proposed model redesigns the loss function of the generator by integrating feature matching for generating the state estimation data efficiently to replace the tampered measurements. The main contributions of this research are summarized as

follows:

- A new defense framework is proposed to defend the grid against stealthy FDI attacks with limited historical labeled measurement data. By taking advantage of the features of TripleGAN, the proposed method can promptly and accurately detect unobservable FDI attacks and effectively recover the state measurement data manipulated by the attacker. In this way, the impacts of such attacks are mitigated. To the best knowledge of the authors, this is the first paper that uses TripleGAN to perform the detection and mitigation task in power systems at the same time. This research is meaningful and critical to improving the power system security and resiliency.

- A new regularization scheme to the TripleGAN model is developed by integrating feature matching in the loss function to enhance the detection accuracy and recovery efficiency.

- A comparison between the proposed method and a series of semi-supervised and supervised methods is performed. The simulation results from the IEEE 118-bus system demonstrate that the proposed technique can effectively defend the grid against the FDI attacks and outperform other machine learning methods. The results also verify that the classifier and generator reinforce each other to enhance the performance of detection and recovery.

4.4 System Model

This section presents the mathematical formulation of stealthy FDI attacks against the state estimation in power grids.

4.4.1 State Estimation

In power grids, electric power is transmitted from power plants to consumers connected by transmission lines. Theoretically, the power flow between bus i and j is a function of the bus voltages and the impedance of the transmission line. The active power flow can be written as:

$$P_{ij} = |V_i||V_j| \left[G_{ij}cos(\theta_i - \theta_j) + B_{ij}sin(\theta_i - \theta_j)\right], \tag{4.1}$$

where V is the voltage magnitude, θ is the phase angle in the corresponding bus, G and B are the real and imaginary parts of the admittance, respectively. To further narrow the power flow statement, G_{ij} can be considered zero because the resistance of the transmission lines is significantly less than the reactance regarding the transmission systems in practice. For the most typical operating conditions, the voltage phase angle difference between two buses is small and the voltage magnitude on each bus is very close to unity in the per-unit system. Thus, further simplification for the power flow equation can be given as:

$$P_{ij} = B_{ij}(\theta_i - \theta_j). \tag{4.2}$$

Expression (4.2) is the DC power flow model which gives the formulation of line power flows in power systems.

To ensure stable and secure operation of the power system, the states (e.g., voltage phase angles) need to be estimated with the measured data collected by the Supervisory Control and Data Acquisition (SCADA) systems. Accurate state estimation is important to the Energy Management System (EMS) for regulating power flow and contingency analysis [109]. In the state estimation problem, the relationship between measured value z and state vector x is constructed as:

$$\mathbf{z} = \mathbf{h}(\boldsymbol{x}) + \mathbf{e}, \tag{4.3}$$

where $x \in \mathbb{R}^N$ contains the voltage phase angles θ_i, $i \in [1, N]$ (N does not include reference

angle on slack bus), $z \in \mathbb{R}^M$ is the vector of measured active power in transmission lines, $e \in \mathbb{R}^M$ is a random Gaussian measurement noise vector with diagonal variance matrix $\Sigma \in \mathbb{R}^{M \times M}$ and $\mathbf{h}(\cdot)$ dictates the relation between z and x. If the DC power flow equation (4.2) is applied, i.e., $\mathbf{h}(\mathbf{x}) = \mathbf{P} = \mathbf{B}(x_i - x_j)$, the Jacobian matrix $\mathbf{H} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}}$, $\mathbf{H} \in \mathbb{R}^{M \times N}$can be considered as constant matrix.

Hence, equation (4.3) can be rewritten as follows:

$$\mathbf{z} = \mathbf{H}\boldsymbol{x} + \mathbf{e}. \tag{4.4}$$

Specifically, expression (4.4) implies that the N phase angles are estimated by observing M real-time active power measurements.

Given the measurements $\mathbf{z}$, the estimated state vector can be calculated by the weighted least-squares (WLS) method:

$$\hat{\boldsymbol{x}} = (\mathbf{H}^T \mathbf{\Sigma} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{\Sigma} \mathbf{z}. \tag{4.5}$$

In this research, the aforementioned DC state estimation based on DC power flow is used for analysis and simulation with the following assumptions: 1) the resistances of the transmission lines are considered as zero, 2) the voltage phase angle difference between two buses is small, and 3) the bus voltage magnitudes are assumed to be 1.0 per unit. For more details about state estimation please refer to [110, 111].

4.4.2 Stealthy FDI Attack

The goal of the attacker in performing FDI attacks is to inject a false data sequence $\mathbf{a} \in \mathbb{R}^M$ into the measurements without being detected. The tampered observation model is represented as follows:

$$\hat{\mathbf{z}}_a = \mathbf{H}\mathbf{x} + \mathbf{a} + \mathbf{e}. \tag{4.6}$$

For conventional bad data detection (BDD), the measurement residual is examined in l_2-norm $\eta = \|\hat{\mathbf{z}} - \mathbf{H}\hat{\mathbf{x}}\|_2^2$ to detect the attacks. If $\eta > \tau$, where $\tau \in \mathbb{R}$ is a predetermined threshold to balance the probabilities of detection and false alarms, then the power system operator decides that the measurements are attacked [22].

Based on the detection methodology explained above, the corrupted observation $\hat{\mathbf{z}}_a$ that cannot be detected by BDD can be characterized as follows [61]:

$$\begin{aligned} \|\hat{\mathbf{z}}_a - \mathbf{H}\hat{\mathbf{x}}_a\|_2 &= \|\mathbf{z} + \mathbf{a} - \mathbf{H}(\hat{\mathbf{x}} + \mathbf{c})\|_2 \\ &= \|\mathbf{z} - \mathbf{H}\hat{\mathbf{x}} + (\mathbf{a} - \mathbf{H}\mathbf{c})\|_2 \\ &\leq \|\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}\|_2 + \|(\mathbf{a} - \mathbf{H}\mathbf{c})\|_2 \leq \tau, \end{aligned} \tag{4.7}$$

By observing (4.7), if $\|\mathbf{a} - \mathbf{H}\mathbf{c}\| \leq \tau_a$, where $\tau_a = \tau - \|\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}\|_2$, the FDI attacks can bypass the BDD mechanism. Therefore, we can conclude the sufficient condition for stealthy FDI attack as follows:

$$\mathbf{a} = \mathbf{H}\mathbf{c} + \mathbf{t}, \tag{4.8}$$

where $\mathbf{c}$ and $\mathbf{t}$ are vectors designed by the attacker, and $\|\mathbf{t}\|_2 \leq \tau_a$.

4.4.3 Constructing Valid stealthy FDI Attack with Limited Access to Measurements

Note that in the aforementioned discussion, the FDI attack vector $\mathbf{a}$ can be sparse which is reasonable in practice. The modern power system consists of a large number of measurements while the attacker may not be able to attack all of them because of limited resources. Therefore, it is assumed that the attacker has access to k measurements and only can modify these k measurements. As a result, the attacker cannot randomly choose vector $\mathbf{c}$ and use $\mathbf{a} = H\mathbf{c}$ as the attack sequence. For the measurements that are not targeted by the attacker, the injected errors must remain 0. [17] develops a method which allows us to easily generate

a sparse attack vector $\mathbf{a}$ that satisfies the above condition.

Theorem 1 *a=Hc if and only if Ba=0, where* $\mathbf{B} = \mathbf{H}\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}} - \mathbf{I}$.

Let $m \times k$ matrix $\mathbf{B}' = (\mathbf{b}_{i1}, ...\mathbf{b}_{ik})$, where $\mathbf{b_i}$ is the i-th column vector of $\mathbf{B}$ and vector $\mathbf{a}' = (a_{i1}, ...a_{ik})$, where a_i is the desired nonzero element of $\mathbf{a}$.

The solution is $\mathbf{a}' = \left(\mathbf{I} - \mathbf{B}'^{\dagger}\mathbf{B}'\right)\mathbf{d}$, where $\mathbf{B}'^{\dagger}$ is the Moore-Penrose pseudo inverse of B', and d is an arbitrary nonzero vector of length k. With a nonzero solution a', the attacker can generate the attack vector $\mathbf{a}$ by filling 0s as the remaining elements in $\mathbf{a}$.

4.5 Proposed Methodology

The proposed defense framework is built based on TripleGAN with the improved loss functions and the Adam optimization algorithm. This section firstly introduces the overview and mechanism of the whole framework, and then describes the redesigned loss function using feature matching and the Adam optimization algorithm.

4.5.1 Defense Framework Overview

The fundamental idea of the framework is to construct a model against FDI attacks which conducts the detection and mitigation with a small set of labeled historical measurements. Fig. 4.1 shows the proposed detection and mitigation mechanism against stealthy FDI attacks.

Offline training stage: In the offline stage, the TripleGAN model is trained by the collected historical data including a small amount of labeled and a large amount of unlabeled

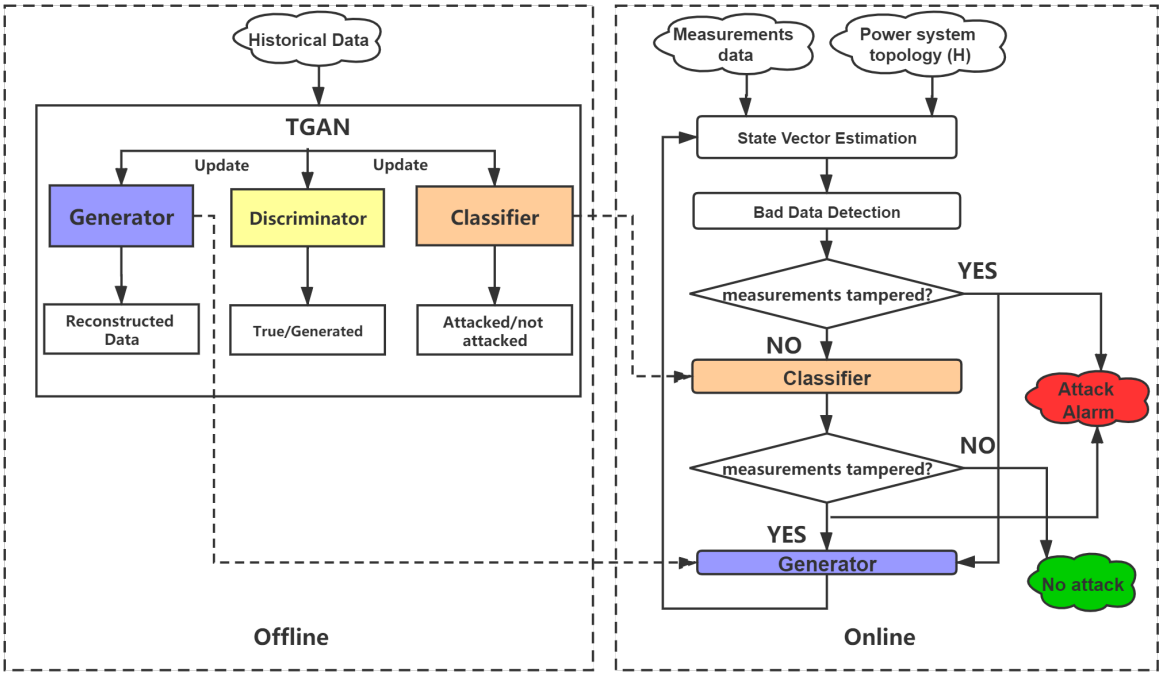

Figure 4.1: The proposed defense framework against FDI attacks

normal/tampered measurement data. The TripleGAN consists of three deep neural networks, namely the generator, the classifier and the discriminator. By iteratively updating each other, the classifier can predict a label (tampered or normal) for the input measurements and the generator can produce fake samples following the normal data distribution to replace the bad measurements.

Online processing stage: In the online stage, the real-time collected measurements are sent to the conventional BDD, checking the deviation of the measurement residual. Once the attack alarm is triggered, these bad data are passed to the generator module. The generated data from the generator is sent back to control center for recovering the state estimation. On the other hand, if the real-time measurements pass the BDD mechanism, the classifier will determine if the measurements are manipulated by the stealthy FDI attack discussed in Section II.

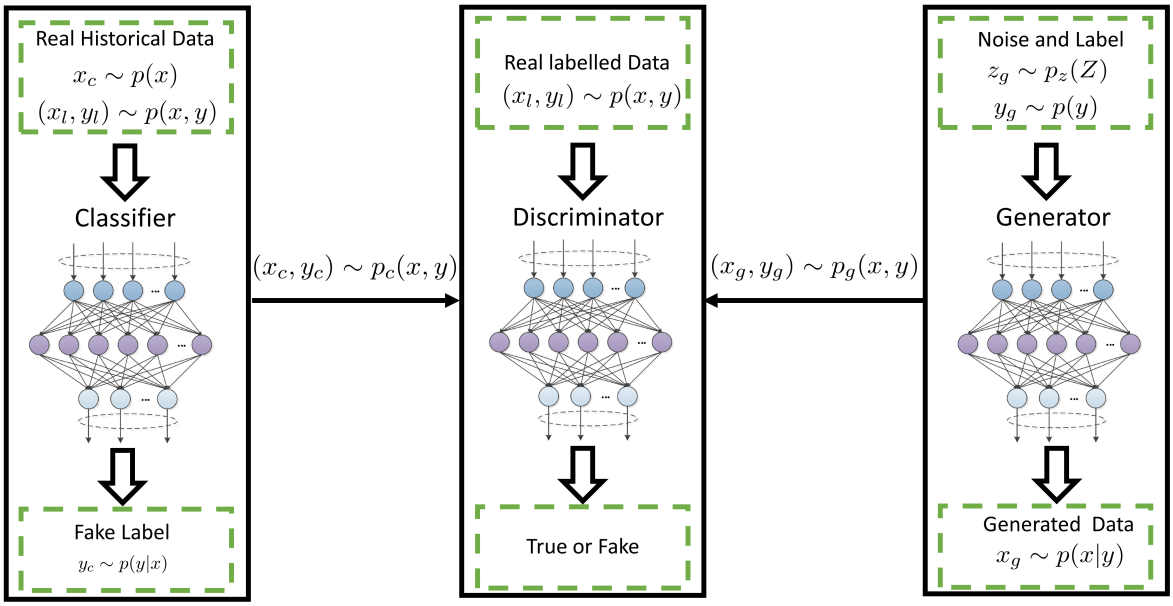

Figure 4.2: Structure of the TripleGAN

4.5.2 Triple Generative Adversarial Network

Although the original GAN is a powerful tool in computer vision and image processing areas, there are several problems for its application in our problem: (1) the generator and the classifier (discriminator) may not align to the data distribution at the same time; (2) the generator is not able to control the feature of the generated data. For this case, we use the triple generative adversarial network (TripleGAN) developed in [108], which performs excellent classification and generation at the same time. TripleGAN can be considered as an extension of the standard GAN based on the three-player game to characterize the process of classification and generation. Here are some notations used for this model. The true measurement input is denoted as x and the label is denoted as y indicating if the corresponding measurement is normal or is compromised by the attacker, with the distribution denoted by $p(x, y)$. The true marginal distributions of x and y are denoted as $p(x)$ and $p(y)$, respectively.

The structure of TripleGAN is shown in Fig. 4.2. TripleGAN consists of three compo-

nents: a generator (G), a classifier (C) and a discriminator (D). Specifically, the generator is to produce data following the distribution of the historical true measurement data, and the classifier is to label the data that follow the true distribution. The discriminator plays the sole role to distinguish the generated data-label pair and true historical data pair. By training them iteratively, the TripleGAN achieves an equilibrium such that the generator and the classifier both converge to the true data distribution. All the three components are characterized as deep neural networks. For the attack detection and mitigation problem in this research, the trained classifier is to determine whether the collected measurement data is tampered with or not, and the bad data can be replaced by the generated data produced from the generator, thus mitigating the impact of FDI attacks.

Generator (G): The generator aims to characterize the conditional distribution $p_g(x|y) \approx p(x|y)$. For $p_g(x|y)$, x is transformed from noise z given the label y, that is, $x = G(z, y)$. z is defined that can be easily sampled from a simple distribution $p_z(z)$. Thus a generated data-label can be synthesized from G by extracting $y_g \sim p(y)$ and $x_g \sim p_g(x|y)$. The synthesized pair is from the joint distribution $p_g(x, y) = p(y)p_g(x|y)$.

Classifier (C): The classifier C aims to characterize the distribution $p_c(y|x) \approx p(y|x)$. After sampling x_c from $p(x)$, C produces a label y_c that follows the conditional distribution $p_c(y|x)$. Similarly, this data-label pair is from the joint distribution $p_c(x, y) = p(x)p_c(y|x)$.

Discriminator (D): The discriminator aims to identify if a data-label pair comes from the true data or the synthesized fake data. Clearly, the fake data-label pairs (x_c, y_c) and (x_g, y_g) are sent to the discriminator. The data-label pairs from the true data distribution are also delivered to D as positive samples.

The loss functions from the G, C and D can be formulated as:

$$\min_G V(G, C, D) = \mathbb{E}_{p_g(x,y)}[\log(1 - D(x_g, y_g)], \tag{4.9}$$

$$\min_C V(G, C, D) = \mathbb{E}_{p_c(y|x)}[\log(1 - D(x_c, y_c))], \tag{4.10}$$

$$\begin{aligned}\max_{D} V(G,C,D) = \mathbb{E}_{p_{x,y}}[\log D(x_l, y_l)] \\ + \gamma \mathbb{E}_{p_c(y,x)}[\log(1 - D(x_c, y_c))] \\ + (1-\gamma)\mathbb{E}_{p_g(x,y)}[\log(1 - D(x_g, y_g)].\end{aligned} \tag{4.11}$$

Combining the objective functions above, a three-player minimax game with value function $V(G, C, D)$ can be formulated as:

$$\begin{aligned}\min_{C,G}\max_{D} V(G,C,D) = \mathbb{E}_{p(x,y)}[\log D(x,y)] \\ + \gamma \mathbb{E}_{p_c(x,y)}[\log(1 - D(x,y))] \\ + (1-\gamma)\mathbb{E}_{p_g(x,y)}[\log(1 - D(G(y,z),y))] + \mathcal{R}_{\mathcal{C}},\end{aligned} \tag{4.12}$$

where $\gamma \in (0, 1)$ is a coefficient constant to control the weights of generation and classification. $\mathcal{R}_{\mathcal{C}} = \mathbb{E}_{p(x,y)}[-\log p_c(y|x)]$ is the cross-entropy loss on the labeled data to C, which guarantees the uniqueness of the equilibrium $p(x, y) = p_g(x, y) = p_c(x, y)$. The equilibrium proof is provided in [108]. In this minimax game, the discriminator competes with the generator and classifier while G and C do not compete with each other. The generator aims to produce high-fidelity data, and the classifier attempts to accurately predict the unlabelled data.

4.5.3 Feature Matching

Though the basic theory of TripleGAN has been proved and demonstrated, there is room for further improvement by tailoring it to our problem. The main concern here is about the stability of the TripleGAN training with the measurement data collected from the power system. There is a certain chance that the generator does not converge to true data distribution resulting that the reconstructed state estimation is far away from the true states. This motivates us to improve the method for a more stable convergence of the generator. Inspired by [107], we adopt the feature matching technique and modify the TripleGAN loss functions.

The fundamental idea is to add a new objective to the generator that enforces the generated data matching with the statistics of the expected distribution, instead of only using the discriminator to distinguish the data-label pairs from the true data and the synthesized data. Reference [107] indicates the new objective for the generator as:

$$\left\|\mathbb{E}_{x\sim p(x,y)}f(x)-\mathbb{E}_{z\sim p_z(z)}f(G(z))\right\|_2^2, \tag{4.13}$$

where $f(\cdot)$ denotes the feature associated with a hidden layer of the discriminator. That is, the generator is forced to match the feature on a hidden layer of the discriminator. To implement it in the TripleGAN, the categories of the samples in (4.13) are taken into account. The objective is modified as follows:

$$\begin{aligned}\mathcal{R}_l = \sum_i ||&\mathbb{E}_{(x,y)\sim p(x,y)}\Omega(y,i)f_C(x)\\ &-\mathbb{E}_{(z,y_g)\sim p_z(z)}\Omega(y_g,i)f_C(G(z,y_g))||_2^2,\end{aligned} \tag{4.14}$$

where $f_C(\cdot)$ denotes the features on the hidden layer of the classifier C; $\Omega(\cdot)$ denotes a binary function which returns 1 if the inputs are equal and 0 otherwise. $i \in (0,1)$ is the class index representing normal and tampered measurement data. Further, to enforce the consistency between the classifier and generator, a semantic matching is adopted to regularize the generator:

$$\mathcal{R}_s = \mathbb{E}_{(z,y_g)\sim p_z(z)}[-y_g \log p_c(x,y)(G(z,y_g))]. \tag{4.15}$$

The advantage of $\mathcal{R}_s$ is it can directly improve the classification through utilizing the generated data which has been regulated by $\mathcal{R}_l$. Thus, by including the feature matching objective terms, the optimization of the generator as (4.9) can be modified as:

$$\min_G V(G,C,D) = \mathbb{E}_{p_g(x,y)}[\log(1-D(x_g,y_g)] + \mathcal{R}_s + \mathcal{R}_l. \tag{4.16}$$

The complete training process of the TripleGAN is shown in Algorithm 3. For the

objective function given above, the three components are trained alternately by the stochastic gradient descent method. The neural network weights θ_d, θ_c and θ_g are updated based on the loss functions in each iteration in steps 4, 6 and 7, respectively.

Algorithm 2 TripleGAN Training Process

Input: Historical measurement data; minibatch size m_g, m_c and m_d; learning rate α

1: Initialize generator, classifier, and discriminator weight as θ_g, θ_c and θ_d

2: **for** number of training iterations **do**

3: Sample minibatch of m_d instances from historical data $(x_l, y_l) \sim p(x, y)$; sample minibatch of m_c instances from data pair $(x_c, y_c) \sim p_c(x, y)$; sample minibatch of m_g instances from data $(x_g, y_g) \sim p_g(x, y)$

4: Update D by stochastic gradient descent:
$g_{\theta_d} \leftarrow \nabla_{\theta_d}[\frac{1}{m_d}(\sum_{(x_l,y_l)} \log D(x_l, y_l))$
$+\frac{\alpha}{m_c}\sum_{(x_c,y_c)} \log(1 - D(x_c, y_c))$
$+\frac{1-\alpha}{m_g}\sum_{(x_g,y_g)} \log(1 - D(x_g, y_g))]$
$\theta_d \leftarrow \theta_d \cdot Adam(\theta_d, g_{\theta_d})$

5: Compute $\mathcal{R}_c$:
$\mathcal{R}_\mathcal{C} = -\frac{1}{m_d}\sum_{(x_d,y_d)} \log p_c(y_d|x_d)$

6: Update C by Adam gradient descent:
$g_{\theta_c} \leftarrow \nabla_{\theta_c}[\frac{-\alpha}{m_c}\sum_{(x_c,y_c)} \log p_c(y_c|x_c) \log(1 - D(x_c, y_c)) + \hat{\mathcal{R}}_c]$
$\theta_c \leftarrow \theta_c \cdot Adam(\theta_c, g_{\theta_c})$

7: Update G by Adam gradient descent:
$g_{\theta_g} \leftarrow \nabla_{\theta_g}[-\frac{1-\alpha}{m_g}\sum_{(x_g,y_g)} \log(1 - D(x_g, y_g)) + \mathcal{R}_s + \mathcal{R}_l]$
$\theta_g \leftarrow \theta_g \cdot Adam(\theta_g, g_{\theta_g})$

8: **end for**

In Algorithm 3, the *Adam* optimization algorithm is used for the stochastic gradient descent. This algorithm, proposed by Kingma and Ba in [112], is based on adaptive estimates of lower-order moments, which aims to address the gradient-based stochastic optimization problem with large datasets and high dimensional spaces. The main advantage of the *Adam* algorithm lies in its higher computational efficiency which renders it to be well suited for tackling problems with a large set of data and parameters.

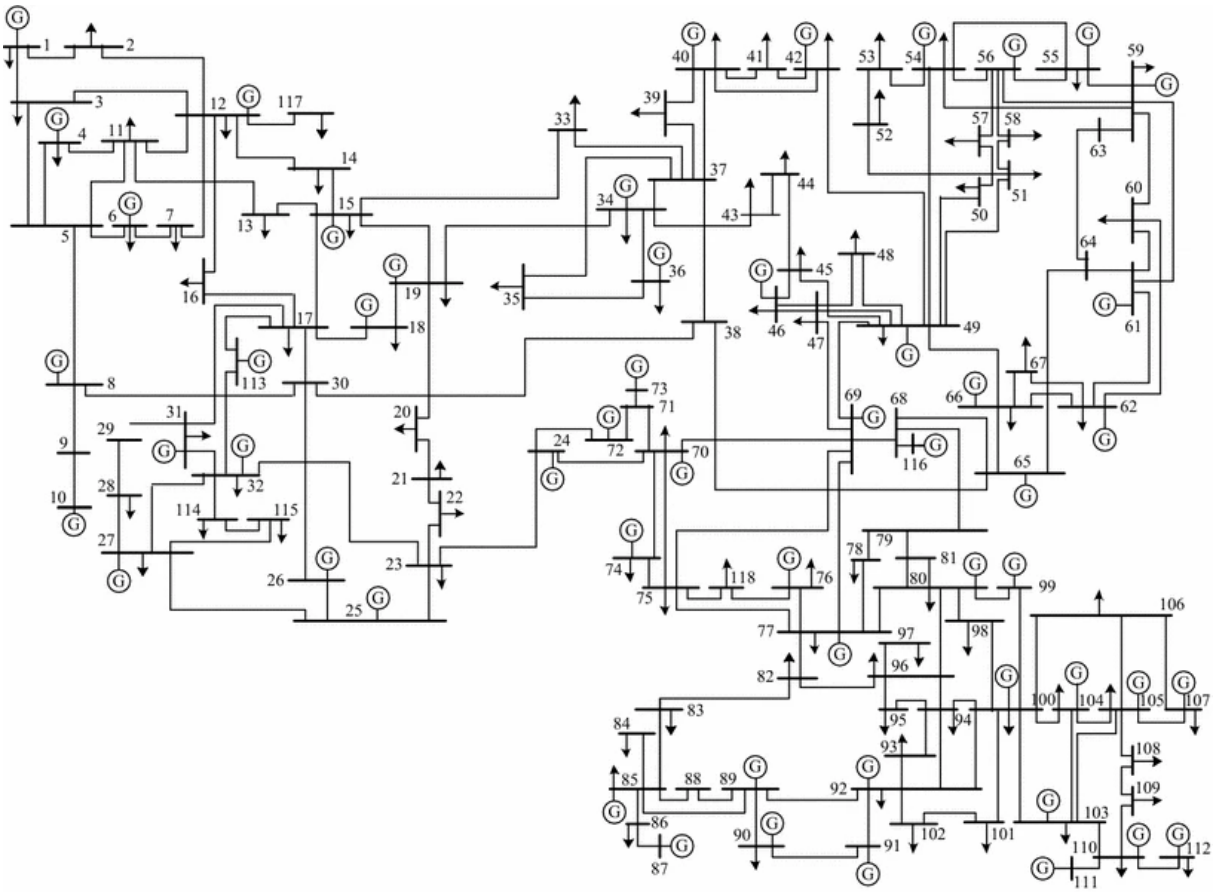

Figure 4.3: IEEE 118-bus system

4.6 Simulation Results and Analysis

Simulations were conducted on the IEEE 118-bus system to evaluate the performance of the proposed model against FDI attacks as Fig. 4.3 [113]. The tampered system data by FDI attacks were simulated on the MATLAB/MATPOWER platform using DC state estimation model discussed in Section II. In total there are 30,000 training samples (which are normal and attacked scenarios) including a certain portion of labeled data. The proposed TripleGAN is trained on the nVidia 1080Ti GPU with an Intel Core i7 CPU of 2.8 Ghz and 16-GB RAM. Each result is the average of 3 runs.

4.6.1 Parameter Selection

- The test system contains 186 measurements which are the line active power flows of each branch ($z \in \mathbb{R}^{186}$), and 117 voltage phase angles of individual buses as the system states ($x \in \mathbb{R}^{117}$). The phase angle on the reference bus is excluded since it is fixed and known. Measurement noise is modeled as a random Gaussian vector $\mathbf{e} \sim \mathcal{N}(0, \Sigma)$ with the standard deviation $\Sigma = 0.001$p.u.

- In this study, the F_1 score is used to measure the accuracy:

$$F_1 = 2\frac{P_r \times R_e}{P_r + R_e}, \tag{4.17}$$

  where P_r and R_e are the precision and the recall, respectively, and they are calculated by the following equations:

$$P_r = \frac{\text{True Positive}}{\text{Predicted Positive}},\ R_e = \frac{\text{True Positive}}{\text{Actual Positive}}.$$

  The precision is the ratio of the correctly predicted positive (normal) data to the total predicted normal data and the recall is the ratio of the correctly predicted positive data to all the normal data in the actual class. In general, the bigger the value of F1 score is, the more accurate the classifier is.

- Dataset: In total, the measurements consist of 30,000 training samples and 6,000 testing samples. There are 500 randomly selected labeled instances among the training data.

- Learning Rate α: The learning rate is a hyper-parameter that controls the step of adjusting the weights of the network with respect to the loss gradient. In this simulation, the learning rate α is set to be 0.001.

- γ in equation (4.12) determines the relative importance of generation and classification.

The higher γ is, the more the weights are placed on classification. In the following simulations, γ is fixed at 0.5 for the balance of the classifier and generator.

- For a fair comparison, all detection results are averaged by 3 runs with different initialization settings.
- All parameters in the *Adam* algorithm follow [112].
- The mean relative errors (MRE) are used to evaluate the performance of the generated data. The state estimation (the phase angle at each bus) computed from the generated data is compared with the normal operation data as:

$$MRE = \frac{1}{N}\sum_{i=1}^{N}|\frac{x_{gi} - x_i}{x_i}|. \tag{4.18}$$

- The detailed structure of our TripleGAN is presented in Table 4.1.

Table 4.1: TripleGAN architecture

Generator G	Discriminator D	Classifier C
Input Class y, Noise z	Input 1x186 measurement Class y	Input 1x186 measurement
MLP 50 units, ReLU MLP 50 units, ReLU MLP 186 units	MLP 120 units,ReLU MLP 120 units, ReLU MLP 186 units, ReLU MLP 1 unit,sigmoid	MLP 80 units, ReLU MLP 80 units, ReLU MLP 80 units, ReLU 2-class sigmoid

4.6.2 Convergence of Proposed TripleGAN-based model

The convergence curve of error rate for the test data is shown as 4.4. The model is trained using the dataset defined in Section IV-A and can reach good detection result in 50 epochs.

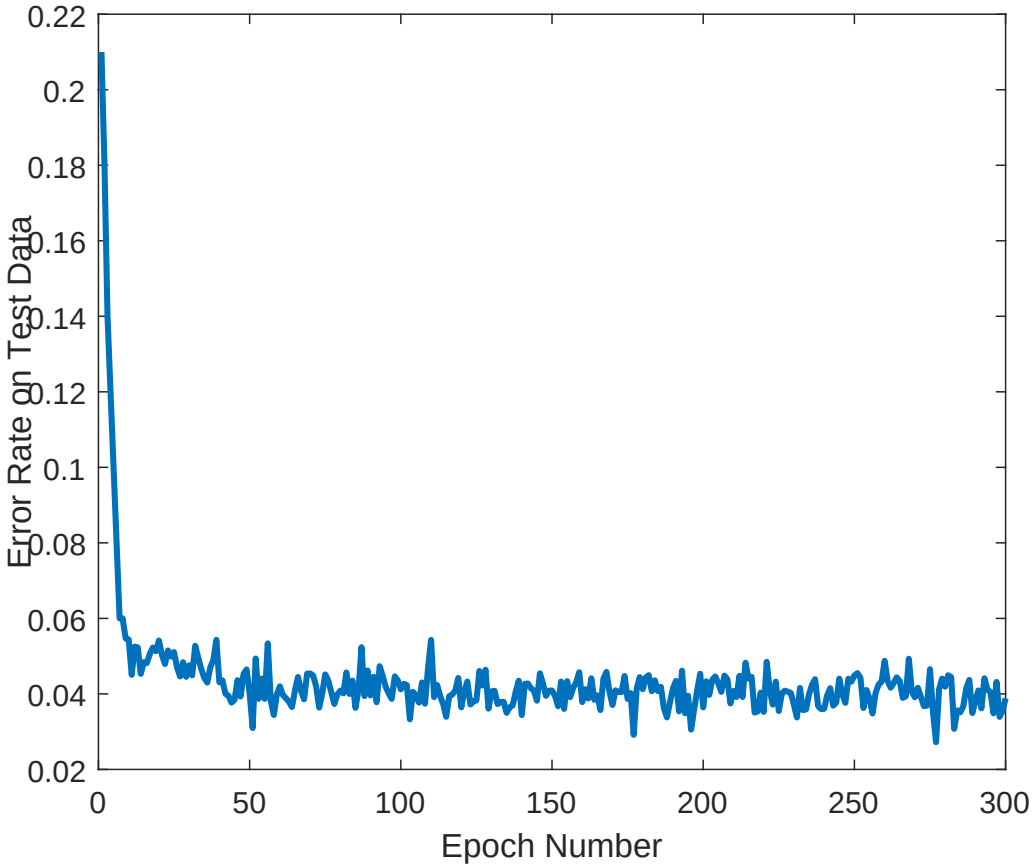

Figure 4.4: Convergence of proposed TripleGAN-based model

4.6.3 FDI Attack Detection

Firstly, the proposed method is compared with two semi-supervised learning algorithms, S3VM [89] and Improved-GAN [107], with different numbers of compromised measurements k. The attack vector is formed following the method in Section II-C to guarantee that the compromised measurements can bypass the residual-based BDD. The selection of targeted measurement sensors are chosen randomly, and the attack intensity is fixed as 30%. The attack intensity is used to define the percentage of compromised measurements of the magni-

tude of normal measurements. In general, the state estimation results deviate further away from the normal values with higher attack intensity. Note that $k \geq m-n+1$ to keep the attack unobservable by BDD [17]. Fig. 4.5 shows the quantitative results. Among these three semi-supervised classifiers, the proposed TripleGAN-based framework outperforms others substantially, especially for the situation where fewer measurement sensors are injected by bad data. The F1 score of our model is 4% to 12% higher than other two methods when k is less than 120. Note that the more measurements the attacker compromises, the more accurate the detection is, which is reasonable from the data correlation perspective. On the other hand, the attacker may cause severer damage to the system with more compromised measurements but under higher risk of being detected. Thus, the attacker tends to seek a trade-off between the attack impact and the risk of exposure.

Then, the robustness of the proposed TripleGAN detection to the noise in the data acquisition environment is evaluated. In this case, the number of compromised measurements k is fixed at 160 and the environmental noise $\mathbf{e} \sim \mathcal{N}(0, \Sigma)$ with the standard deviation Σ changing from 0.001 p.u. to 0.02 p.u. [110]. Intuitively, the detection accuracy decreases with a higher level of the environmental noise because the corrupted and normal data are less distinguishable. However, a detection structure with good robustness can mitigate this trend. Fig. 4.6 demonstrates that the proposed detection scheme is more robust to the environmental noise compared with S3VM and Improved-GAN. It is found from Fig. 4.6 that even with $\Sigma = 0.02$ p.u. environmental noise, the proposed TripleGAN-based detection still achieves an F1 score of 0.9 which is higher than the others.

Lastly, the proposed TripleGAN detection method is compared with two supervised methods, SVM [87] and kNN [114], on a small labeled dataset including 500 normal and compromised samples. To perform the supervised learning of TripleGAN, the losses on unlabeled data as presented in Section III are removed and all the networks are the same as the semi-supervised learning. The comparison is still based on the number of compromised measurements and the environmental noise as shown in Fig. 4.7 and 4.8. Because of a

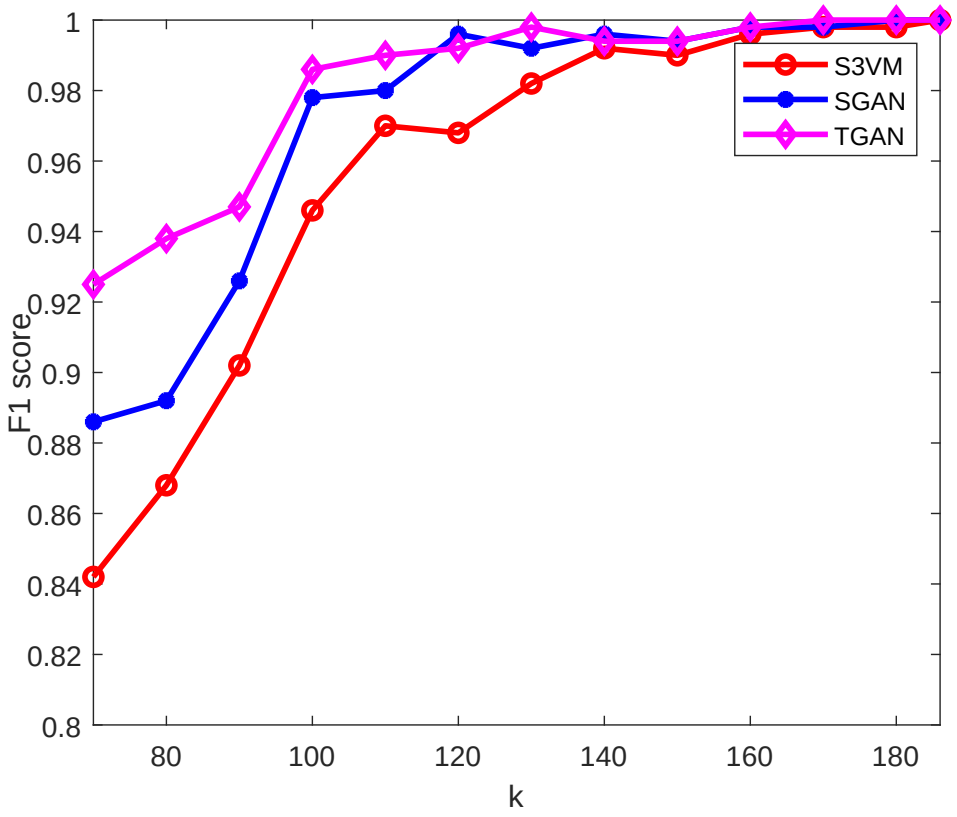

Figure 4.5: Detection performance of S3VM, SGAN and TripleGAN with different numbers of compromised measurements

smaller set of training data which is reasonable in the real world, all the F1 scores are worse than the semi-supervised learning cases in Fig. 4.5 and 4.6. Nevertheless, the proposed TripleGAN-based detection framework exhibits a better detection performance and higher robustness to the environmental noise.

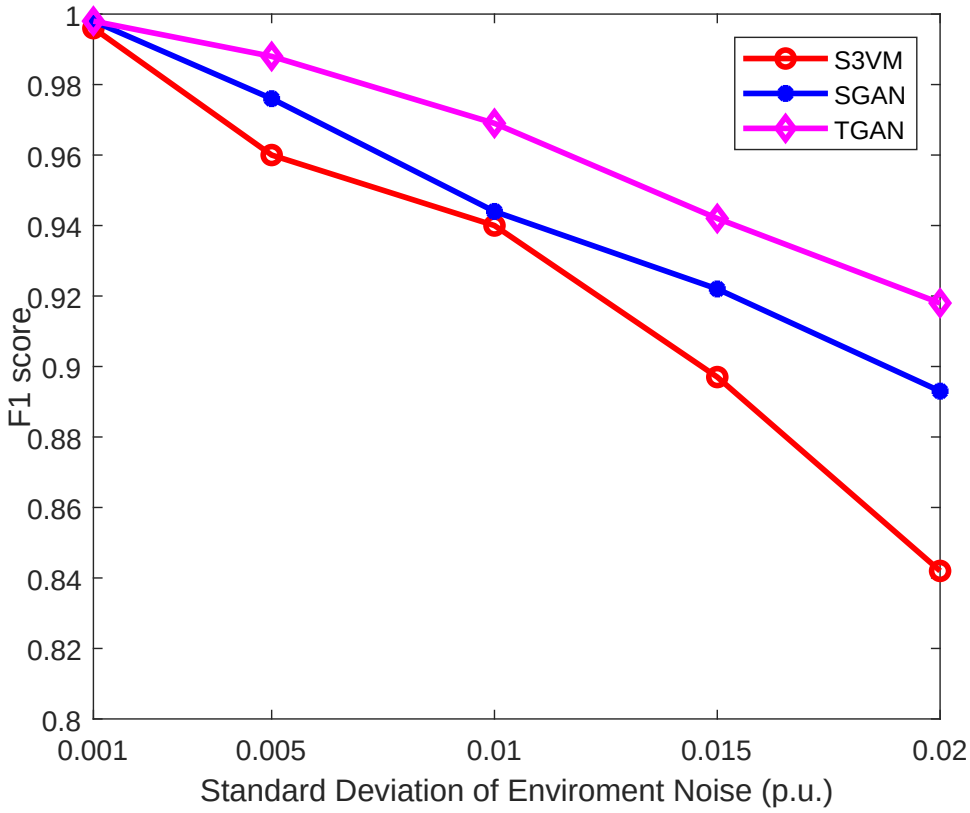

Figure 4.6: Detection performance of S3VM, SGAN and TripleGAN with different StD of environmental noise

4.6.4 Attack Mitigation

As discussed in Section III, the proposed mitigation principle is to generate new measurement data which are sufficiently close to the true measurements to replace the tampered data. The generated data corrects the state estimator. Thus, to evaluate the proposed defense model, we firstly compare the reconstructed data with the normal data from the state estimator as shown in Fig. 4.9, 4.10 and 4.11. Due to the space limitation, only three attack scenarios are presented here: 80, 100 and 140 tampered measurements which are randomly selected. The results demonstrate that state estimation is robust to the stealthy FDI attacks with the

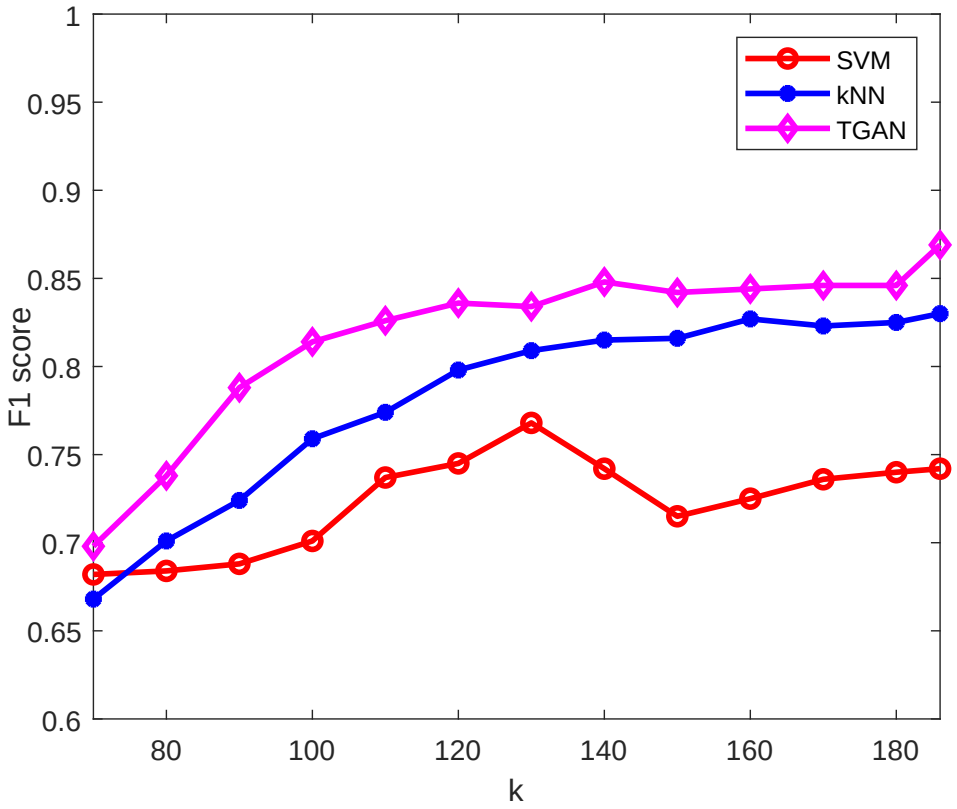

Figure 4.7: Detection performance of SVM, kNN and TripleGAN with different numbers of compromised measurements on small labeled dataset

proposed defense model. The phase angles are sufficiently close to the true values, where the impact of FDI attacks is mitigated in the power grid.

Next, the results are evaluated quantitatively compared with the data generated by the semi-supervised GAN (SGAN) [107] and Conditional GAN (CGAN) [115]. Table 4.2 summarizes the MRE of three GAN models. In all three attack scenarios, the recovered data obtained by the proposed TripleGAN has smaller MRE than SGAN and CGAN, which means the data is closer to the states calculated by true measurements. The reason why CGAN has the relatively worst performance is that CGAN has to be trained with fully labeled data.

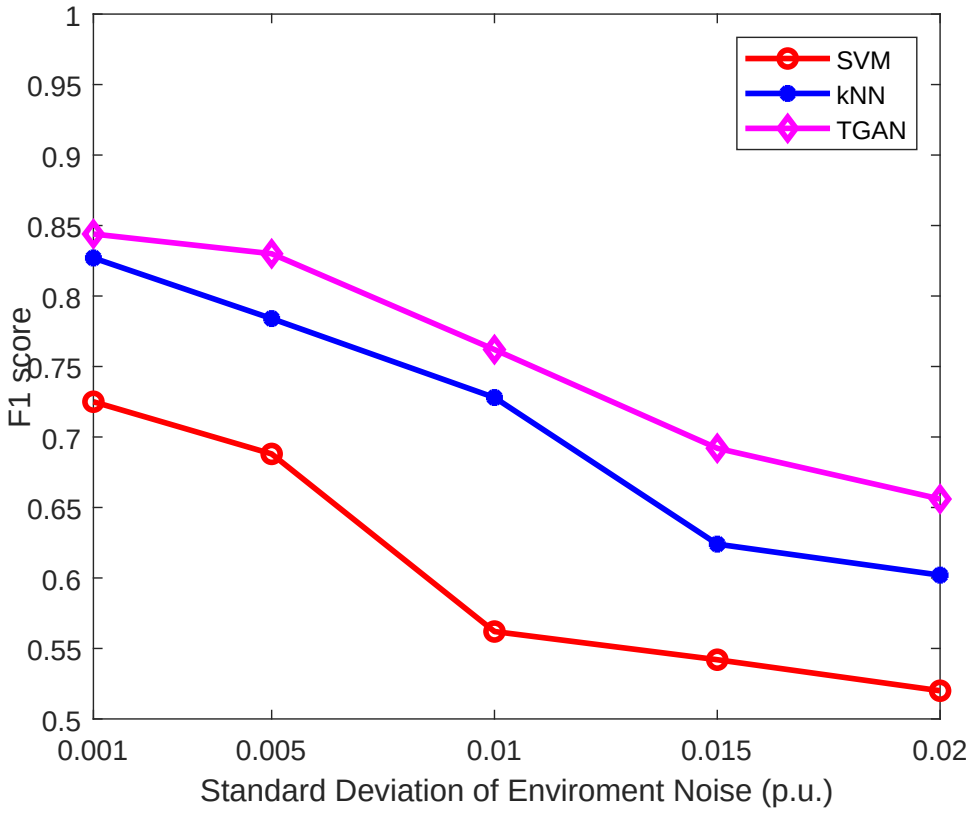

Figure 4.8: Detection performance of SVM, kNN and TripleGAN with different StD of environmental noise on small labeled dataset

In our case, there are only 500 labels in the 30,000 training samples and CGAN does not leverage the large amount of unlabeled data. Overall, the mitigation results confirm that the proposed TripleGAN-based model can recover the tampered data caused by FDI attacks and thus lead to a better performance than the others.

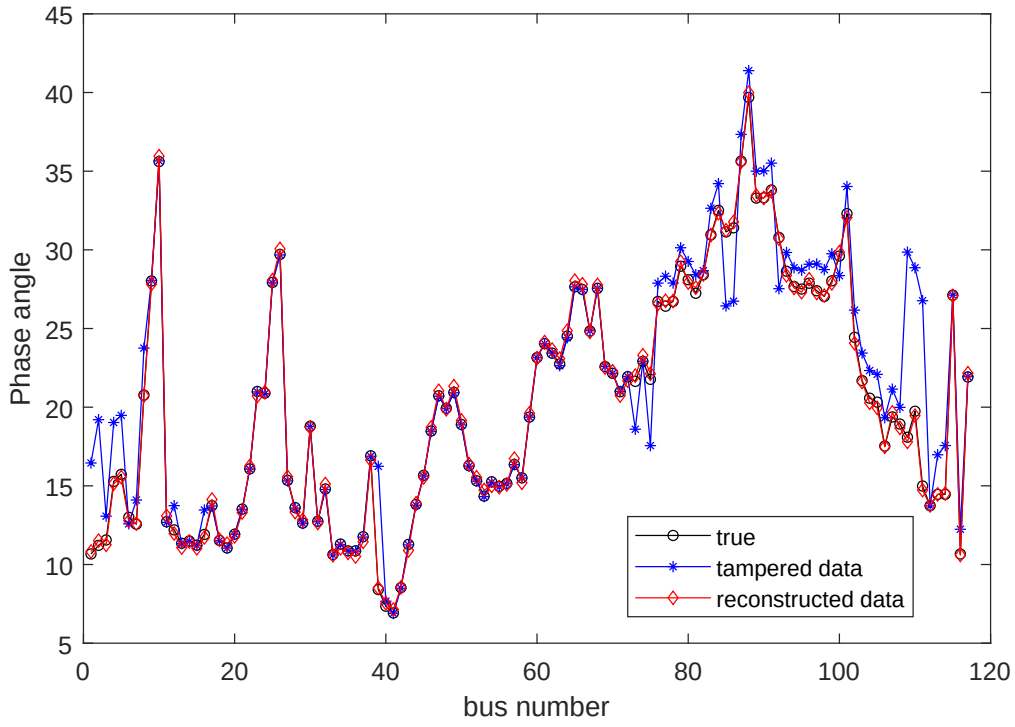

Figure 4.9: Phase angle in state estimation with 80 corrupted measurements

Table 4.2: MRE of TripleGAN and standard GAN

Number of corrupted measurements	80	100	140
SGAN [107]	2.48%	2.86%	3.05%
CGAN [115]	3.69%	4.15%	6.73%
TripleGAN	1.42%	2.04%	2.21%

Table 4.3: Performance of different numbers of labeled samples for TripleGAN-based framework

Number of Labeled Samples	200	500	900	1400
Detection F1 Score	0.927	0.938	0.941	0.945
Generation MRE	2.40%	1.95%	1.84%	1.78%

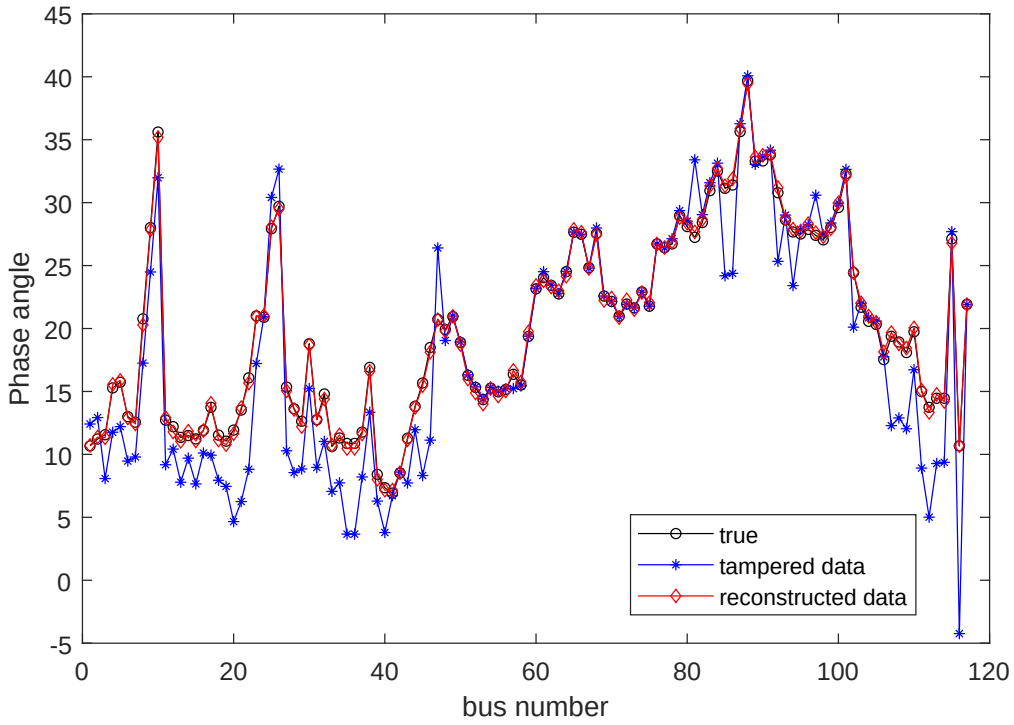

Figure 4.10: Phase angle in state estimation with 100 corrupted measurements

4.6.5 Impact of the number of labeled data on performance

Table 4.3 compares the performance of the proposed model using different numbers of labeled samples. Apparently, the larger the portion of labeled data is, the better the performance of the model has. However, we emphasize that there are limited labeled samples in our dataset in practice because detected and observed cyber-attacks records for power system are rare (but the damage is severe once it happens). Since the operation data are highly confidential for electric utility companies, we assume that 1% to 2% compromised data are labeled. Furthermore, for the cases when the number of labeled samples is larger than 500, the performance is improved finitely. The results prove that our proposed model performs satisfactorily on 30,000 training samples with 500 labeled samples.

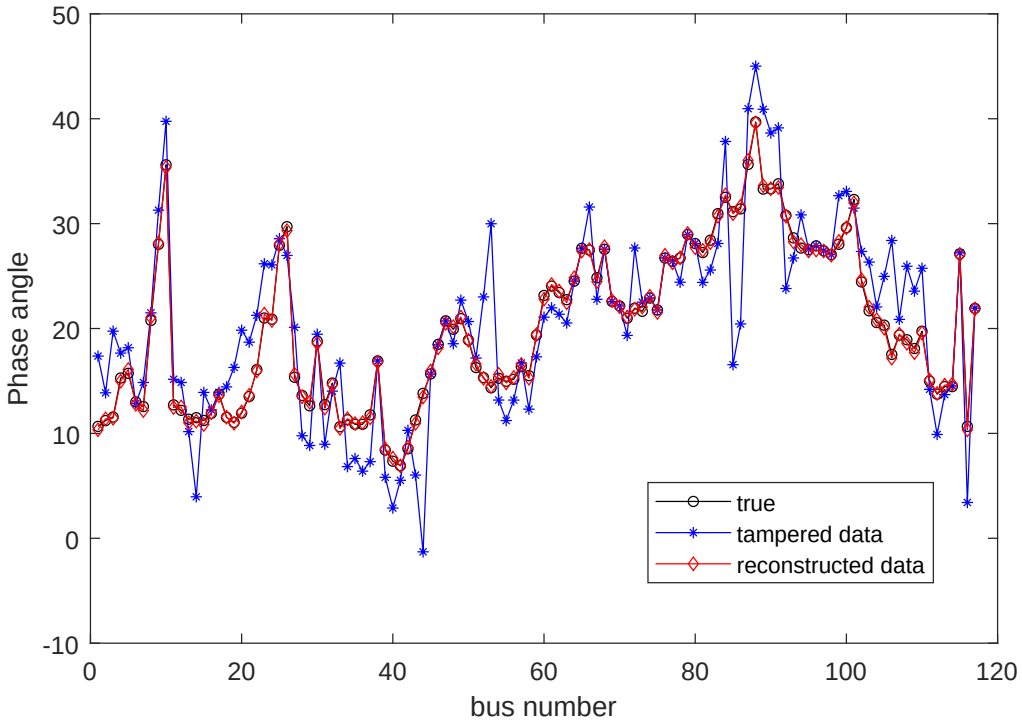

Figure 4.11: Phase angle in state estimation with 140 corrupted measurements

4.7 Summary

This chapter proposes a novel TripleGAN-based defense framework against the stealthy FDI attacks, which aims to accurately detect the attack and effectively mitigate its impact at the same time with a few labeled historical measurements. The state estimation and stealthy FDI cyber-attack model are first introduced in this chapter. Then, the detection and mitigation defense framework is proposed including offline training and online processing. In the offline training stage, the TripleGAN model is trained by the collected historical data including a small amount of labelled and a large amount of unlabeled normal/tampered measurements. The detection (performed by the classifier) is integrated and trained with the mitigation (performed by the generator) together and they reinforce each other. To

enhance the detection accuracy and recovery efficiency, an extended loss function is proposed by including feature matching. In the online processing stage, the classifier is used to detect if the measurements are tampered by FDI attack. If the attack alarm is triggered, the tampered measurements are replaced by the generated data from the generator and sent back to control center for recovering the sate estimation and informing further decisions. F1 Score and MRE are used to evaluate the detection and mitigation performance.

The simulation results demonstrate that the proposed defense model is able to accurately detect the stealthy FDI attacks and the recovered state estimation is sufficiently close to the normal operation status, which thus improves the power system resilience. Furthermore, under various circumstances (with different numbers of targeted measurements, different intensities of environmental noise, and fewer historical data), the obtained results confirm that the proposed techniques exhibit some advantages over other machine learning based detection and recovery methods.

Chapter 5

DATA-DRIVEN FDIA MODEL AGAINST LOAD FREQUENCY CONTROL

5.1 Overview

With the tighter integration of physical power system and cyber system, the security of Load Frequency Control (LFC) system has attracted wide attention recently. As a major class of cyber threats to power systems, False Data Injection (FDI) attacks inject well-crafted attack data into the data collected by the LFC system, which may mislead the decision-making of control center and compromise the secure operation of power systems. In this chapter, a data-driven FDI attack model based on multi-agent deep reinforcement learning (MA-DRL) against LFC system is proposed. Instead of using legacy linearized LFC model, AC state estimation (ACSE) is integrated with LFC to reduce measurement noise and perform bad data detection. Thus, the system environment becomes more practical and complex, and more requirements need to be satisfied for the attacker to perform successful FDI attacks. In order to achieve two attack objectives simultaneously, i.e., stealthily maximizing the frequency deviation and minimizing the number of compromised measurements, a modified Multi-Agent Deep Deterministic Policy Gradient (MA-DDPG) algorithm is devised in this study, which treats the two objectives separately by global and local individual critic networks other than

a simple linear combination. The impact of FDI attack on the LFC with ACSE is also analytically derived. The obtained Laplace-domain transfer function can be used to replace the global critic network to accelerate offline training process. During the online execution, each attacker of LFC control area cooperatively injects the optimal attack vector only by its own local measurements. Simulation results on the New-England 39-bus system verify the effectiveness of the proposed FDI attack model compared with other methods. In addition, corresponding countermeasures based on the critical measurements are discussed.

5.2 Related Work

The load frequency control (LFC) system aims to restore system frequency and eliminate the imbalance between power generation and load demand, where the imbalance is generally a consequence of scheduled or unscheduled fluctuation of load [85, 116]. As the basic function of Automatic Generation Control (AGC), the LFC system computes control command based on Area Control Error (ACE) collected from the distributed sensors. The control command is then sent to the generator to adjust the power output to balance the load demand change, thereby the deviated grid frequency is brought back to the nominal value and can be kept within a specific fluctuation range.

However, as deeper integration and coupling with cyber system, the LFC system faces severe security challenges. LFC relies on modern communication technologies for data collection and feedback, and requires high integrity of the collected data, so it is vulnerable to cyber attacks [104, 117]. False Data Injection (FDI) attack is one of the most threatening cyber attacks to LFC system, since the data collection devices usually are not deployed with complicated encryption technology due to real-time performance requirements and the field data being transmitted by susceptible communication channels to the LFC control center [118]. The vulnerability of data collection and transmission can be exploited by malicious

adversaries to perform FDI attacks in order to cause shocks in power system and severe grid accidents. Thus, it is imperative to investigate the threat of FDI attacks to LFC system and the corresponding defense strategy.

Presently, there are many model-based FDI attack studies to LFC system. In [18], the impact of data integrity attacks on AGC on power system frequency is proposed and demonstrated. In [119], the authors propose a coordinated FDI attack which is a combination of several attack templates. This coordinated attack is verified to cause wider parameter deviation in shorter execution time. In Ref. [120], the resonance attacks to LFC is studied. The attacker can craftily modify the input of generator based on resonance source and quickly cause the power plant unstable. The authors in [121] propose the FDI attack and defense model on LFC based on Generative Adversarial Network (GAN). In [122], the FDI attack that targets on frequency sensor and the corresponding decentralized detection framework are studied. Whereas, the FDI attack models [18, 119–122] directly target on ACE signal or frequency sensor, which are usually not easy and practical to perform because the data link of ACE is always well-protected by isolated cable instead of transmitting signals by wireless networks, and as a global parameter, grid frequency can be easily monitored and verified by sensors deployed in the control center.

On the other hand, with the development of high-performance computing technologies, State Estimation (SE) is integrated to LFC to reduce measurement noise and perform faulty data detection [123]. SE system ensures the measurements are free of environmental noise and bad data and then the improved measurements without noise are transmitted to LFC control center, leading to more accurate control signal. The SE is also vulnerable to FDI attack that targets on distributed measurements [124], yet more careful attack structure design needs to be considered. In order to persistently impact the operation of SE and then interfere with LFC control, the injected false data have to be stealthy to avoid being detected by Bad Data Detector (BDD), such as l_2-norm detection. In addition, attackers attempt to minimize the number of tampered measurements due to limited computing and

communication resources. In the literature, crafting stealthy FDI has been widely studied on linear power flow model for DC state estimation (DCSE). In [125], a stealthy FDI attack against DCSE is proposed, where the minimum set of attacked meters can be obtained by the observability of the system. A coordinated FDI attack is designed in [126], where the physical topology can be coordinated modified to mislead the results of SE. In [30,127–129], with only incomplete system information, the stealthy FDI attack is crafted and system matrix can be estimated by model-based or data-driven methods. As for more practical AC state estimation (ACSE) based on nonlinear power flow, reference [130] analyzes the vulnerability of ACSE and generalizes a common FDI attack policy. In [131], a nonlinear optimization-based attack policy which can be computed by semi-definite programming is designed. In [25], the authors propose and analyze the uncertainties for launching successful FDI attacks with the upper bound under the condition of incomplete system information. In [132, 133], FDI attacks without network information are derived, where the network parameters are extracted by the analytical algorithm or GAN-based machine learning method.

Although FDI attacks against SE have been widely studied in the aforementioned research, the impact of FDI attacks on LFC system integrated with SE is an underexplored domain in the power system security area. In [134], the optimal FDI attack sequence against AGC is derived by analyzing the impact of FDIA on AGC. Although the SE is included in this research, the linearized LFC and DCSE system is adopted such that the derived optimal attack could be easily detected by widely used ACSE in practical power systems. In addition, the computational burden for each step might be overlooked especially for the attackers with limited resources. In [135], a framework of FDI attack along with denial-of-service (DoS) attack is proposed, where the false data is injected at frequency and power flow sensors. However, SE is not considered in this research, so the attack lacks stealthiness which could be easily detected.

The above presented work motivates us to design a data-driven stealthy FDI attack scheme against LFC system applied in centralized learning and distributed execution way. To

fill the research gaps mentioned in the above literature review, this research proposes a novel MA-DRL based FDI attack model against LFC system integrated with AC state estimation. Considering the different characteristics of LFC control areas in the multi-area power system, the attack vector injected in power measurements for each control area adjusts its value at each time step and cooperatively maximizes the grid frequency deviation and minimizes the number of tampered measurements at the same time. The principle of the proposed method consists of offline centralized training and online distributed implementation stages. At offline stage, attack of each LFC control area design is formulated as MA-DRL to maximize their objectives. During online execution, the attacker of each area cooperatively adjusts its attack vector only by its own local measurements. The main contributions of this article are summarized as follows.

- A data-driven FDI attack model against LFC system is proposed. Instead of legacy linearized LFC model, AC state estimation is coupled with LFC to reduce environment noise and filter bad data before sending measurements to LFC controller, which is more precise and practical for the analysis of interactions among control areas.
- MA-DRL method is used to solve the problem of increasing frequency deviation and minimizing the number of tampered measurements. In order to achieve these two objectives simultaneously, a modified MA-DDPG algorithm is proposed in this chapter. Unlike a simple linear combination of the two objectives in [63], the proposed algorithm handles them separately by global and local individual critic networks compared to standard MA-DDPG.
- The impact of FDI attack on the LFC system is analytically derived. A closed-form Laplace-domain model is obtained and then utilized to replace global critic network to accelerate the training process and reduce the performance variance.
- During online execution, the trained FDI attack model is verified in a more complex and practical system in Matlab/simulink. The detailed dynamic transient model of

three-phase synchronous machines and resistance of all the lines and transformers are taken into account. The simulation results validate the good performance of proposed FDI attack model compared with other methods. Furthermore, corresponding countermeasures based on checking out the critical measurements are discussed and verified.

5.3 System Models

In this section, Load Frequency Control (LFC) integrating State Estimation (SE) and FDI attack against LFC model in power grids are introduced.

5.3.1 Load Frequency Control and State Estimation

Frequency status has to be monitored all the time and regulated periodically in power system, since frequency deviation from the nominal value caused by imbalance between power generation and load demand can directly impact power grid security, reliability and operation. Thus, in order to ensure the balance between load and generation, LFC is enforced to bring the deviated frequency back to the nominal value. Each frequency controller has the area control error (ACE) as its control input:

$$ACE_i = \Delta P_{tie,i} + B_i \Delta f_i, \tag{5.1}$$

where $\Delta P_{tie,i}$ and Δf_i are tie-line power flow deviation and frequency deviation for control area i, B_i is constant. However, tie-line power flow sensors are usually noisy and faulty, which may mislead the LFC controller and impact frequency stability. State Estimation (SE) can reduce measurement noise and perform faulty sensor detection at the same time, which yields more accurate control signal to LFC. Especially in recent years, with the development of high-performance computing, the execution time of SE is significantly reduced and integrating

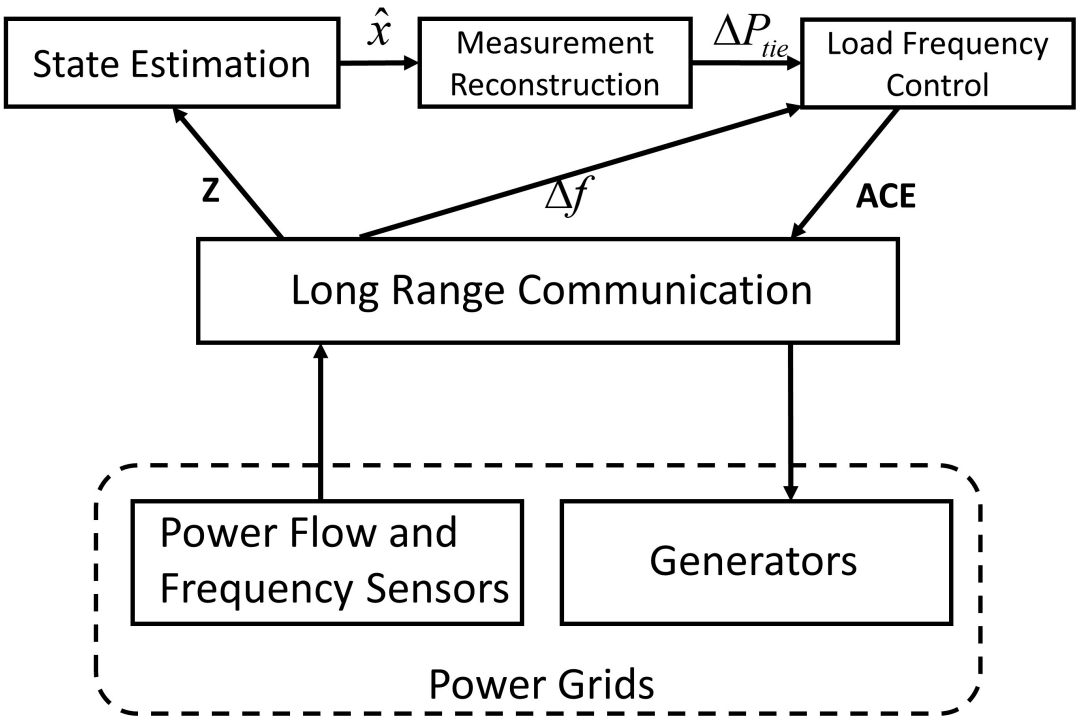

Figure 5.1: Overview of LFC with State Estimation

SE to LFC becomes feasible [123, 136].

Fig. 5.1 depicts the LFC pipelined with SE. The control center collects measurements $\mathbf{z}$ from distributed power flow and frequency sensors and calculates the grid state $\hat{x}$ through state estimation. ACE signal is a linear combination of tie-line's power deviation of each area ΔP_{tie} and measured grid frequency deviation Δf, where ΔP_{tie} is computed based on state estimation $\hat{x}$. Then ACE is transmitted to the control unit of generators to specify their setpoints.

For the state estimation problem, considering a system with m measurements and n state variables, the relation between measured value z and state vector x can be constructed by the nonlinear measurement model, i.e., AC state estimation (ACSE):

$$\boldsymbol{z} = \boldsymbol{h}(\boldsymbol{x}) + \boldsymbol{e}, \tag{5.2}$$

where $\boldsymbol{z} \in \mathbb{R}^{mx1}$ is the measurement vector; $\boldsymbol{x} \in \mathbb{R}^{n\times 1}$ is system state vector; $\boldsymbol{e} \in \mathbb{R}^{m\times 1}$ is the measurement noise vector which is assumed to follow Gaussian distribution with zero mean and covariance matrix $\boldsymbol{R} \in \mathbb{R}^{m\times m}$; and $\boldsymbol{h}(\cdot)$ is a nonlinear function. The state estimator can be solved by minimizing the weighted least squares (WLS) criterion, yielding

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} [\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x})]^T \boldsymbol{R}^{-1} [\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x})]. \tag{5.3}$$

The solution to (5.3) can be obtained by applying the Gauss-Newton iterative algorithm:

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + (\boldsymbol{U}(\boldsymbol{x}))^{-1} \boldsymbol{H}\left(\boldsymbol{x}^k\right)^T \boldsymbol{R}^{-1} \left(\boldsymbol{z} - \boldsymbol{h}\left(\boldsymbol{x}^k\right)\right), \tag{5.4}$$

$$\boldsymbol{U}(\boldsymbol{x}) = \boldsymbol{H}\left(\boldsymbol{x}^k\right)^T \boldsymbol{R}^{-1} \boldsymbol{H}\left(\boldsymbol{x}^k\right), \tag{5.5}$$

where k is the iteration index; and $\boldsymbol{H}$ is the Jacobian matrix. After conducting estimation, l_2-norm Bad Data Detector (BDD) is performed to detect the existence of bad data by checking if the following inequality holds:

$$\|\boldsymbol{r}\| = \|\boldsymbol{z} - \boldsymbol{h}(\widehat{\boldsymbol{x}})\| \geq \tau \tag{5.6}$$

where τ is a detection threshold of the l_2-norm detector.

5.3.2 Attack Model of LFC and SE

For a typical multi-area power grid, LFC model of the ith area is shown in Fig. 5.2. Each area consists of the generator, governor, steam turbine and load. As widely reported in the literature, system dynamics of the ith area can be represented by the differential equations

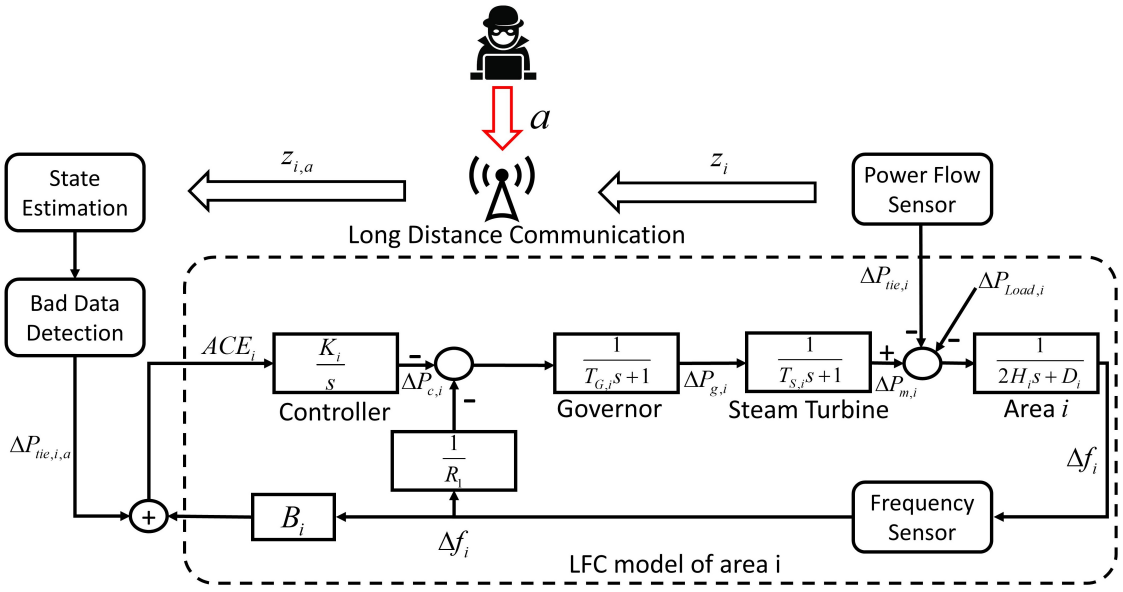

Figure 5.2: FDI Attack against LFC

as follows:

$$\Delta \dot{f}_i = \frac{1}{2H_i}\left(\Delta P_{m,i} - \Delta P_{Load,i} - \Delta P_{tie,i}\right) - \frac{D}{2H_i}\Delta f_i, \tag{5.7}$$

$$\Delta \dot{P}_{m,i} = \frac{1}{T_{S,i}}\Delta P_{g,i} - \frac{1}{T_{S,i}}\Delta P_{m,i}, \tag{5.8}$$

$$\Delta \dot{P}_{g,i} = \frac{1}{T_{G,i}}\Delta P_{c,i} - \frac{1}{R_i T_{G,i}}\Delta f_i - \frac{1}{T_{G,i}}\Delta P_{g,i}. \tag{5.9}$$

$$\Delta \dot{P}_{c,i} = K_i ACE_i. \tag{5.10}$$

The variables' notations are summarized in nomenclature.

Due to the state estimation unit being integrated, equation (5.1) becomes

$$ACE_i = \Delta \hat{P}_{tie,i} + B_i \Delta f_i, \tag{5.11}$$

where $\Delta \hat{P}_{tie,i}$ is tie-line power flow deviation calculated by estimated states for the ith area.

As can be seen in Fig. 5.2, FDI attack may be achieved by exploiting ACE signals, fre-

quency sensors or power flow sensors. Among the three potential compromising approaches, the data streams that transmitting ACE signals from the control center to the generators are usually hard to attack, because they are well-protected by isolated cable instead of transmitting signals by wireless networks. Since the grid frequency is a global parameter of power system, the measurements from remote sensors can be easily monitored and verified by sensors deployed in the control center. Therefore, in this research we focus on FDI attacks on the power flow sensor measurements. Note that other than attacking geographically distributed power flow sensors, targeting the sensor data links is more practical for attackers to avoid the tedious hacking process.

The goal of the attacker in performing FDI attacks is to inject a false data sequence $\boldsymbol{a}$ into measurements without being detected by bad data detector (BDD). The tampered measurement matrix can be represented as:

$$\boldsymbol{z}_a = \boldsymbol{z} + \boldsymbol{a}. \tag{5.12}$$

In order to perform FDI attack stealthily, attack sequence a has to bypass the l_2-norm detector. Based on the aforementioned detector methodology, the attacked measurement residual $\boldsymbol{r}_a$ is given by:

$$\begin{aligned} \|\boldsymbol{r}_a\| &= \|\boldsymbol{z}_a - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}_a\right)\| = \|\boldsymbol{z} + \boldsymbol{a} - \boldsymbol{h}(\widehat{\boldsymbol{x}} + \boldsymbol{c})\| \\ &= \|\boldsymbol{z} + \boldsymbol{a} - \boldsymbol{h}(\widehat{\boldsymbol{x}} + \boldsymbol{c}) + \boldsymbol{h}(\widehat{\boldsymbol{x}}) - \boldsymbol{h}(\widehat{\boldsymbol{x}})\| \\ &= \|\boldsymbol{r} + \boldsymbol{a} - \boldsymbol{h}(\widehat{\boldsymbol{x}} + \boldsymbol{c}) + \boldsymbol{h}(\widehat{\boldsymbol{x}})\|, \end{aligned} \tag{5.13}$$

where $\boldsymbol{c}$ represents the changes of state variable. From equation (5.13), it is observed that if $\boldsymbol{a} = \boldsymbol{h}(\widehat{\boldsymbol{x}} + \boldsymbol{c}) + \boldsymbol{h}(\widehat{\boldsymbol{x}})$, attacked measurement residual $\|\boldsymbol{r}_a\| = \|\boldsymbol{r}\| \leq \tau$, which means that the tampered measurement could avoid being detected by l_2-norm detector.

Moreover, the grid operator may apply other data quality check on $\boldsymbol{z}$. For example, $\boldsymbol{z}$ should not significantly change in a short time period. Therefore, each element of attack

sequence $\boldsymbol{a}$ has to be bounded in a small range around zero:

$$\boldsymbol{a}_{min} \leq \boldsymbol{a} \leq \boldsymbol{a}_{max}. \tag{5.14}$$

Under the premise of satisfying (5.14), the number of compromised sensors for each LFC area should be as fewer as possible. Intuitively, if fewer sensors are altered, the attack is harder to be recognized by the operator. In this way, those detection mechanisms which are designed to be insensitive to natural random noises will not be altered.

For such FDI attacks, the goal is to maximize the frequency deviation to sabotage the stability of power systems. However, due to the constraints in (5.13), (5.14) and number of accessible sensors, the hacker might not be able to cause enough frequency deviation in one shot. Instead, the hacker can inject a series attack vectors to create the insecure frequency deviation.

In this research, the following assumptions are made to perform the mentioned stealthy FDI attacks: 1) the hacker has priori knowledge of the power system topology, including the line parameters, $\boldsymbol{a}_{min}$ and $\boldsymbol{a}_{max}$; 2) The hacker could only alter real-time measurements in sub-area $\mathcal{S}$ but would have access to reading the whole power flow measurements. Thus, the attack vector $\boldsymbol{a}$ is subject to $\boldsymbol{a}[j] = 0$ if $j \notin \mathcal{S}$, where $a[j]$ is the jth element of the attack vector; and 3) the hacker is able to change all the measurements $\mathcal{S}$.

5.4 Multi-Agent Deep Reinforcement Learning FDI Attack

5.4.1 Framework Overview

In this study, the stealthy FDI attack against LFC problem is formulated in a MA-DRL framework. Each LFC area's attack of multi-area power system can be viewed as a DRL agent. Reinforcement learning explores the optimal actions policy to maximize the agents' long-term performance by interacting with the environment that is the LFC model defined in equations (5.7)-(5.10). As such, the FDI attack problem can be formulated in a distributed manner by utilizing MA-DRL.

The schematic of the proposed framework is illustrated in Fig. 5.3. The method is divided into offline centralized learning and online decentralized execution. At the offline stage, each agent is modeled based on artificial neural network (ANN). The agents try out different actions, i.e. attack vectors, interacting with the LFC system. The obtained action-states and reward are employed to iteratively update the ANN parameters of agents with global objective. At the online stage, each agent collects its own states, i.e., local observation, and calculate attack vector for each LFC area based on trained ANN at online stage. In this manner, agents in different areas will cooperatively adjust the attack vector according to their state observations to maximize frequency deviation in a fully distributed way.

5.4.2 Markov Decision Process

The MA-DRL generally formulates its environment as a multi-agent extension Markov Decision Process (MDP). At time step t, MDP with N agents is defined as a set of states $s^t \in \mathcal{S}$ that describes the environment for all agents, a set of actions $a_i^t \in \mathcal{A}_i$ and a set

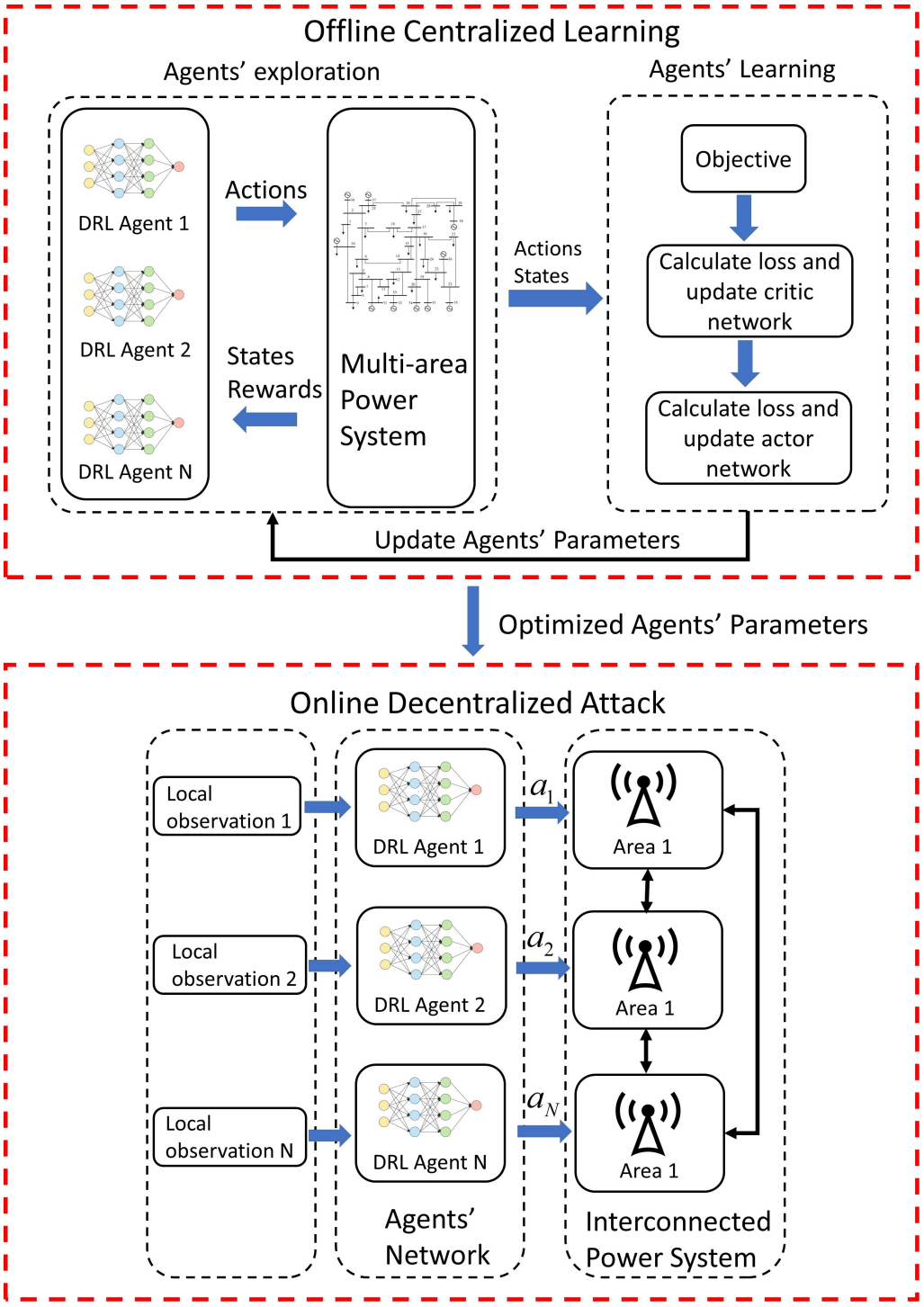

Figure 5.3: Framework of the Proposed Approach

of observations $o_i^t \in \mathcal{O}_i$. Action space $\mathcal{A}$ consists of $\{\mathcal{A}_1,...,\mathcal{A}_N\}$ and observation space $\mathcal{O}$ contains $\{\mathcal{O}_1,...,\mathcal{O}_N\}$. Each agent uses its own policy $\pi_i(o_i^t)$ to select actions and then the next state is produced according to state transition model $T(s^{t+1}|s^t, a^t)$. State transition defines the dynamics of the environment depending on the equations (5.7)-(5.10) in this paper. Each agent receives new observation o_i^{t+1} reward r_i^t as a function of the state and the agent's action. The aim of each agent is to find out policy π_i^* which maximizes its own total expected discounted return $\mathbb{E}[\sum_{t=0}^{T} \gamma^t r_i^t]$, where $\gamma \in [0, 1]$ is the discount factor and T is the time horizon.

Definitions of agents, state, action and reward in this chapter are discussed as follows.

- **Agents**: In a multi-area power system, each agent represents the attacker of each LFC area.
- **Action**: The actions are defined as attack vectors. Each agent is able to continually adjust its attack vector a_i to modify the state of the environment.
- **State and Observation**: The states are defined as the system-wide frequency deviation $\{\Delta f_1^t ... \Delta f_N^t\}$, sensor measurements $\{z_1^t ... z_N^t\}$ and number of attacked sensors $\{k_1^t ... k_N^t\}$. These states reflect the impact caused by FDI attack and the magnitude of these injected attack vectors. The observation of each agent is defined as its own frequency deviation, derivative of frequency deviation, sensory measurements and number of attacked sensors, i.e., local measurements Δf_i^t, $d\Delta f_i^t/dt$, z_i and k_i.
- **Reward**: The design of reward function is crucial because it evaluates the effectiveness of actions and determines the pattern that agents learn. Rewards are usually stipulated based on agents' objectives. For the case in this research, two components need to be taken into account: causing more frequency variation and not being detected by BDD.

Thus, two separate rewards are designed as:

$$R1_i^t = \sum_{i=0}^{N} (\Delta f_i)^2 / N \tag{5.15}$$

$$R2_i^t = \begin{cases} -R_a k_i, & \text{if } k_i < k_{max,i} \ , \\ -R_b, & \text{otherwise,} \end{cases} \tag{5.16}$$

where k_i is the number of altered sensors for the ith LFC area; R_a represents the positive constant for scaling the reward; R_b is large positive constants for harsh penalty if k_i is larger than a constant threshold $k_{max,i}$, e.g., 90% sensors of the ith area. Reward (5.15) represents the average frequency deviation of all LFC areas, where each agent will be awarded larger rewards with larger average frequency deviation. Reward (5.16) is designed to minimize the number of attacked sensors to not trigger detection alert. It is worth mentioning that all agents have the same reward in (5.15) given the system-wide information, while in (5.16) each agent is awarded different reward. Thus, $R1_i^t$ can be considered as global reward denoted as Rg_i^t and $R2_i^t$ is denoted as Rl_i^t representing local individual reward. The total reward r_i^t can be expressed as a function of Rg_i^t and Rl_i^t,

$$r_i^t = F(Rg_i^t, Rl_i^t) \tag{5.17}$$

How to handle function F will be discussed later.

5.4.3 Multi-Agent Deep Deterministic Policy Gradient

The essential problem for the MDP mentioned above is to learn an optimal policy to maximize the long-term expected discounted return. One of the state-of-the-art solving process is MADDPG [137], which is a multi-agent and actor-critic DRL algorithm.

In this algorithm, each agent is characterized by ANN denoted as $\mu_i(\cdot|\theta_i)$, i.e., actor,

mapping from observation to action, where θ_i is the weight parameters of the actor. The offline centralized training aims to learn the optimal weights for each agent. Policy Gradient method allows that parameter θ_i is adjusted directly to maximize the objective function $J(\theta_i)$ by moving the policy in the direction of the gradient of $J(\theta_i)$:

$$J(\theta_i) = \mathop{\mathbb{E}}_{s^{t+1}\sim T, a_i^t \sim \mu_i} \sum_{t=0}^{T} \gamma^t r_i^t. \tag{5.18}$$

The gradient for agent i is given by

$$\nabla J(\theta_i) = \mathop{\mathbb{E}}_{s^t \sim \mathcal{D}} \left[\nabla_{\theta_i} Q\left(s^t, a_1^t, \ldots, a_N^t\right) |_{a_i^t = \mu_i(o_i^t|\theta_i)} \right]. \tag{5.19}$$

where $\mathcal{D}$ is the replay buffer that contains a collection of experience; $Q(\cdot)$ is the centralized action-value function that evaluates the performance of actor. By taking the chain rule, the equation (5.19) can be divided into the gradient of Q with respect to a_i and the gradient of the policy with respect to θ,

$$\nabla J(\theta_i) = \mathop{\mathbb{E}}_{s^t \sim \mathcal{D}} \left[\nabla_{\theta_i} \mu_i(o_i^t|\theta_i) \nabla_{a_i^t} Q\left(s^t, a_1^t, \ldots, a_N^t\right) |_{a_i^t = \mu_i(o_i^t|\theta_i)} \right], \tag{5.20}$$

Next, the action-value function $Q\left(s^t, a_1^t, \ldots, a_N^t\right)$ is parameterized by ϕ, i.e., critic, denoted by $Q(\cdot|\phi)$ for agent i. It can be updated by minimizing the following loss:

$$\mathcal{L}(\phi) = \mathop{\mathbb{E}}_{s^t \sim \mathcal{D}} \left[(Q(s^t, a_1^t, \ldots, a_N^t \mid \phi) - y_i^t)^2 \right], \tag{5.21}$$

where

$$y^t = r^t + \gamma Q'(s^{t+1}, a_1^{t+1\prime}, \ldots, a_N^{t+1\prime} \mid \phi_i')|_{a_i^{t+1\prime}} = \mu_i'\left(o_i^{t+1}\right). \tag{5.22}$$

$\mu_i'(\cdot|\theta_i')$ and $Q'(\cdot|\phi')$ are the target actor and critic networks with delayed parameters θ_i' and ϕ', respectively. They are used to estimate y^t that is the reference value for Q. The weights

are updated by slowly tracking the actor and critic network.

$$\begin{aligned} \theta_i' &\leftarrow \tau\theta_i + (1-\tau)\theta_i' \\ \phi' &\leftarrow \tau\phi + (1-\tau)\phi' \end{aligned} \tag{5.23}$$

5.5 The Proposed Solution Process

In this section, a novel modified training process is presented as the extension of MADDPG. The basic idea is to combine MADDPG with single agent DDPG together and optimize the global objective.

5.5.1 Local Critic

Hitherto, the only remaining problem is how to deal with the two different reward or control goals described in (5.15) and (5.16). To achieve both objectives, generally the reward function (5.17) for each agent i can be designed as:

$$r_i^t = (\alpha)Rg_i^t + (1-\alpha)Rl_i^t \tag{5.24}$$

where $\alpha \in (0,1)$ is the weight assigned to both rewards. The main concern on equation (5.24) is the stability issue. Since there are two reward components, oscillation of the policy between the local reward Rl_i^t and global reward Rg_i^t may ensue. In addition, α can be either set as a constant which may not balance the two rewards, or trained as a parameter that is very time-consuming. Thus, the local individual critic denoted as $Q_i^l(s, a_i|\phi_i^l)$ approximated by a neural network with weights ϕ_i^l for agent i is introduced besides global critic $Q(s, a_1, ..., a_N)$ to handle the local reward. Note that comparing to $Q(s, a_1, ...a_N)$ in equation (5.19) that receives the actions of all agents and the environment states, the local critic $Q_i^l(s, a_i)$ only considers its own actions for agent i. Moreover, each agent has its individual local critic

$Q_i^l(\cdot|\phi_i^l)$ while the global critic $Q(\cdot|\phi)$ is a shared network between all N agents during learning process. By adopting the local critic, the gradient equation (5.20) becomes

$$\begin{aligned}\nabla J(\theta_i) = &\mathop{\mathbb{E}}_{s^t\sim\mathcal{D}}\left[\nabla_{\theta_i}\mu_i(o_i^t|\theta_i)\nabla_{a_i^t}Q\left(s^t,a_1^t,\ldots,a_N^t\right)\right]\\ &+\mathop{\mathbb{E}}_{o_i^t,a_i^t\sim\mathcal{D}}\left[\nabla_{\theta_i}\mu_i(o_i^t|\theta_i)\nabla_{a_i^t}Q_i^l\left(s^t,a_i^t\right)\right]\end{aligned} \tag{5.25}$$

where $a_i^t = \mu_i(o_i^t|\theta_i)$. MADDPG is still used to update the first term of equation (5.25) maximizing the global reward Rg:

$$\mathcal{L}(\phi) = \mathop{\mathbb{E}}_{s^t\sim\mathcal{D}}\left[(Q(s^t,a_1^t,\ldots,a_N^t \mid \phi) - y_g^t)^2\right], \tag{5.26}$$

where

$$y_g^t = Rg^t + \gamma Q'(s^{t+1},a_1^{t+1\prime},\ldots,a_N^{t+1\prime} \mid \phi')|_{a_i^{t+1\prime}} = \mu_i'\left(o^{t+1}\right). \tag{5.27}$$

The slight difference between equations (5.26)-(5.27) and (5.21)-(5.22) is that average reward r_i^t is replaced by global reward Rg_i^t.

The second term of equation (5.25) is updated by the single-agent DDPG maximizing local reward Rl,

$$\mathcal{L}(\phi_i) = \mathop{\mathbb{E}}_{s^t,a_i^t\sim\mathcal{D}}\left[(Q_i^l(s^t,a_i^t \mid \phi_i) - y_{li}^t)^2\right], \tag{5.28}$$

where

$$y_{li}^t = Rl_i^t + \gamma Q_i^{l\prime}(s^{t+1},a_i^{t+1\prime} \mid \phi_i^{l'})|_{a_i^{t+1\prime}} = \mu_i'\left(o_i^{t+1}\right). \tag{5.29}$$

Similarly, target local critic $Q_i^{l'}(\cdot|\phi_i^{l'})$ is also adopted to improve the stability of learning. In this way, by introducing the local critic, the agent learns to maximize the global and local rewards simultaneously.

5.5.2 DDPG based Solution Method

Although the problem of two separate rewards is solved by the aforementioned local critic technique, the offline training program for such multi-area power systems could be very time-consuming. In addition, the action gradients estimated by critic may cause high variance, especially for the global critic regarding the reward (5.15). Considering the frequency dynamics is differentiable in power systems, the mentioned DDPG method can be derived without the critic approximations.

Firstly, since the global objective is to maximize the average frequency deviation, the action-value Q function can be modeled as:

$$Q\left(s^t, a_1^t, \ldots, a_N^t\right) = \sum_{i=0}^{N} (\Delta f_i)^2 / N, \tag{5.30}$$

where, Q is dependent on the all agents' actions, i.e., $(a_1^t, \ldots, a_N^t)$, and the state information s. Next gradient of action-value Q function with respect to attack vector a_i needs to be quantified. For agent i in the ith area, taking gradient with respect to a_i on both sides of equation (5.30), it becomes:

$$\nabla_{a_i} Q = 2\Delta f_i \frac{\partial \Delta f_i}{\partial a_i}. \tag{5.31}$$

Now, the gradient of frequency deviation in different areas with respect to the attack vector a_i needs to be estimated. Using the diagram in Fig. 5.3 and equations (5.7)-(5.13), the transfer function between frequency deviation and attack can be derived. Note that generally the load disturbance ΔP_{load} is not affected by attack and thus $\nabla_a \Delta P_{load} = 0$. The relationship between the frequency deviation and the attack vector can be obtained as follows:

$$\Delta F_i(s) = \frac{1}{\xi_4 s^4 + \xi_3 s^3 + \xi_2 s^2 + \xi_1 s + \xi_0 s} A_i(s) \tag{5.32}$$

where the s represents s-domain for Laplace transformation; the coefficients

$$\begin{aligned}
\xi_4 &= 2HT_gT_s, && (5.33)\\
\xi_3 &= \frac{2H(T_g+T_s)+DRT_gT_s}{DT_gT_sV\psi}, && (5.34)\\
\xi_2 &= \frac{2HT_gT_s(DT_gT_s+2HT_g+2HT_s)}{R\psi(T_g+T_s)}, && (5.35)\\
\xi_1 &= \frac{2H+DT_g+DT_s}{T_gT_sKV\psi}, && (5.36)\\
\xi_0 &= BK/R. && (5.37)
\end{aligned}$$

Note that the index i of the parameters that stands for LFC area is omitted for simplicity. Here, ψ_2 is defined as a column vector that selects relevant elements in power flow measurement and aggregates them as P_{tie}; V is the improved measurement after state estimation when no attack vector is injected, i.e., $\boldsymbol{h}(\widehat{\boldsymbol{x}})$.

Then, the relationship between $\Delta f(t)$ and $a(t)$ can be derived by multiplying $\xi_4 s^4+\xi_3 s^3+\xi_2 s^2+\xi_1 s+\xi_0 s$ on both sides of equation (5.32) and taking Laplace inverse transformation. Next, the gradient of variables with respect to $a(t)$ can be calculated as:

$$\nabla_a \Delta f(t) = \frac{1}{\xi_0}\left(1-\xi_1\frac{d\Delta f(t)}{dt}\right) - \frac{\xi_2}{\xi_0}\nabla_a\frac{d^2\Delta f(t)}{dt^2} \quad (5.38)$$

$$-\frac{\xi_3}{\xi_0}\nabla_a\frac{d^3\Delta f(t)}{dt^3} - \frac{\xi_4}{\xi_0}\nabla_a\frac{d^4\Delta f(t)}{dt^4}. \quad (5.39)$$

If high-order terms of Δf are neglected in (5.38), by substituting the frequency gradient into (5.31), the estimation for the $\nabla_{a_i} Q$ is obtained:

$$\nabla_{a_i} Q \approx 2\Delta f_i \frac{1}{\xi_0}\left(1-\xi_1\frac{d\Delta f_i}{dt}\right). \quad (5.40)$$

Hitherto, the global critic ANN $Q(\cdot|\phi)$ and loss calculation (5.26) are not needed anymore because we already derive the estimation of $\nabla_{a_i^t} Q$ component in the first term of (5.25). Then, the gradient of actions with respect to the parameters of each agent $\nabla_{\theta_i}\mu_i(o_i^t|\theta_i)$ in

(5.25) can be computed as follows,

$$\nabla_{\theta_i^{(k)}} \mu_i(o_i^t|\theta_i) = \nabla_{\boldsymbol{\theta}_i^k} \left(f_{\boldsymbol{\theta}_i^k}^{(n)} \left(\dots f_{\boldsymbol{\theta}_i^k}^{(1)}(\mathbf{X}) \right) \right), \tag{5.41}$$

where $f_\theta^{(l)(\cdot)}$ is the activation function in the lth layer, and $\mathbf{X}$ is the input vector of ANN. Note that the local critic $Q_i^l(s, a_i|\phi_i^l)$ and equations (5.28)(5.29) still remain for optimizing the local objective for each agent. The detailed training algorithm is summarized in Algorithm 3.

5.5.3 Training and Execution Processes

At the offline training stage, each agent has individual actor and local critic network. As can be seen in Algorithm (3), the training process starts with the initialization of each neural network. The episodes loop is the main loop, in which each agent interacts with the power system environment to learn the optimal control policy for the generator. As the number of episodes increases, each attack agent tends to approach the optimal policy that maximizes the frequency deviation and probability of not being detected.

Each episode consists of a number of iterations. The number of iterations should not be too large for reducing negative effects of ineffective actions. For each iteration time step t, attack actions are determined by exploration $a^t = \mu(s_t|\theta) + \mathcal{N}_r^t$, where $\mathcal{N}_r \sim (0, \sigma^2)$ is Gaussian noise. Along with the training, the exploration rate is decreased by reducing the noise. Then according to grid dynamics (5.7)-(5.13), system's next state and rewards are observed as $(s^t, a^t, Rl^t, s^{(t+1)})$. Note that Rl is a vector instead of a scalar number since each agent is awarded with different local rewards. These observations are stored in the replay buffer $\mathcal{D}$. All networks at each time step t are updated by uniformly sampling minibatch from the replay buffer, which avoid the divergence of learning caused by using sequential samples. At the end of the episodes, each attack agent learns an optimal policy. At the

Algorithm 3 The Proposed MA-DRL Training Algorithm

Randomly initialize local critic network $Q_i^l(o_i, a_i|\phi_i^l)$ and actor network $\mu_i(s, a|\theta_i)$
Initialize target network $Q_i^{l'}$ and μ_i' by $\phi_i^{l'} \leftarrow \phi_i^l$ and $\theta_i' \leftarrow \theta_i$
for episode =1 to M **do**
  Initialize power system environment
  Initialize a random process $\mathcal{N}_r$ for action exploration
  Receive initial state s_1
  for t=1 to T **do**
    Determine actions by $a^t = \mu(s_t|\theta) + \mathcal{N}_r^t$
    Execute actions $a^t = (a_1, ...a_N)$ and observe new state $S_{(t+1)}$ and obtained reward Rl^t
    Store $(s^t, a^t, Rg^t, Rl^t, s^{(t+1)})$ to replay buffer $\mathcal{D}$
    Sample a random minibatch of m samples $(s^k, a^k, Rl^k, s^{(k+1)})$ from $\mathcal{D}$
    Calculate the policy gradient
    $\nabla_{a_i^k} Q \approx 2\Delta f_i \frac{1}{\xi_0}\left(1 - \xi_1 \frac{d\Delta f_i}{dt}\right)$
    Calculate the policy gradient
    $\nabla_{\theta_i^{(k)}} \mu_i(o_i^k|\theta_i) = \nabla_{\boldsymbol{\theta}_i^k}\left(f_{\boldsymbol{\theta}_i^k}^{(n)}\left[\ldots f_{\boldsymbol{\theta}_i^k}^{(1)}(\mathbf{X})\right]\right)$
    for Agent i to N **do**
      Set $y_{li}^k = Rl_i^k + \gamma Q_i^{l\prime}(o^{k+1}, a_i^{k+1\prime} \mid \phi_i^{l'})|_{a_i^{k+1\prime}=\mu_i'(o_i^{k+1})}$
      Update local critic by minimizing the loss
      $\mathcal{L}(\phi_i) = \frac{1}{m}\sum_k \underset{o_i^k, a_i^k \sim \mathcal{D}}{\mathbb{E}}\left[(Q_i^l(s^k, a_i^k \mid \phi_i) - y_{li}^k)^2\right]$
      Update actor policy by $\frac{1}{m}\sum_k (\nabla_{\theta_i}\mu_i(o_i^k|\theta_i)\nabla_{a_i^k} Q_i\left(s^k, a_1^k, \ldots, a_N^k\right) + \nabla_{\theta_i}\mu_i(o_i^k|\theta_i)\nabla_{a_i^k} Q_i^l\left(s^k, a_i^k\right))$
    end for
    Update target network parameters for each agent i
    $\theta_i' \leftarrow \tau\theta_i + (1-\tau)\theta_i'$
    $\phi_i^{l'} \leftarrow \tau\phi_i^l + (1-\tau)\phi_i^{l'}$
  end for
end for

online stage, the attack agent will only use the local measurements from the power system to inject attack vector.

5.6 Simulation Results

In this section, comparative simulation studies are carried out on the New England 39-bus system to evaluate the performance of the proposed FDI attack. All tests are performed in the Matlab/simulink environment with deep learning toolbox version 2020b in a computer with 16 GB RAM and 2.8 GHz CPU. The MA-DRL model trained in a Python environment is imported to Matlab/Simulink and tested on a more realistic 39-bus system.

5.6.1 Test Environment

The simulations are performed on a fully-modeled 39-bus system in Matlab/simulink, where the detailed dynamic transient model of synchronous machines is adopted. The 39-bus system is divided into three areas as shown in Fig. (5.4). The tie lines are $line_{1,2}$, $line_{3,4}$, $line_{16,17}$ and $line_{14,15}$. LFC controllers are equipped for the generators on buses 30, 32, 35 and 38, while the remaining generators are under simple primary control. It is assumed that the measurements $\boldsymbol{z}$ in the power system are the following: 1) active power and reactive power flows on all lines at both ends of the line; 2) active power and reactive power injections at buses with loads. Total numbers of measurements and accessible measurements for each area are summarized in Table 5.2. Detailed system parameters of generators, loads, transformers, and buses are presented in [138]. The proposed FDI attack method is compared with standard MADDPG where the reward is designed in equation (5.24), and a common FDI attack method in [130].

The neural network architecture of actor for each agent is a four-layer ANN which has

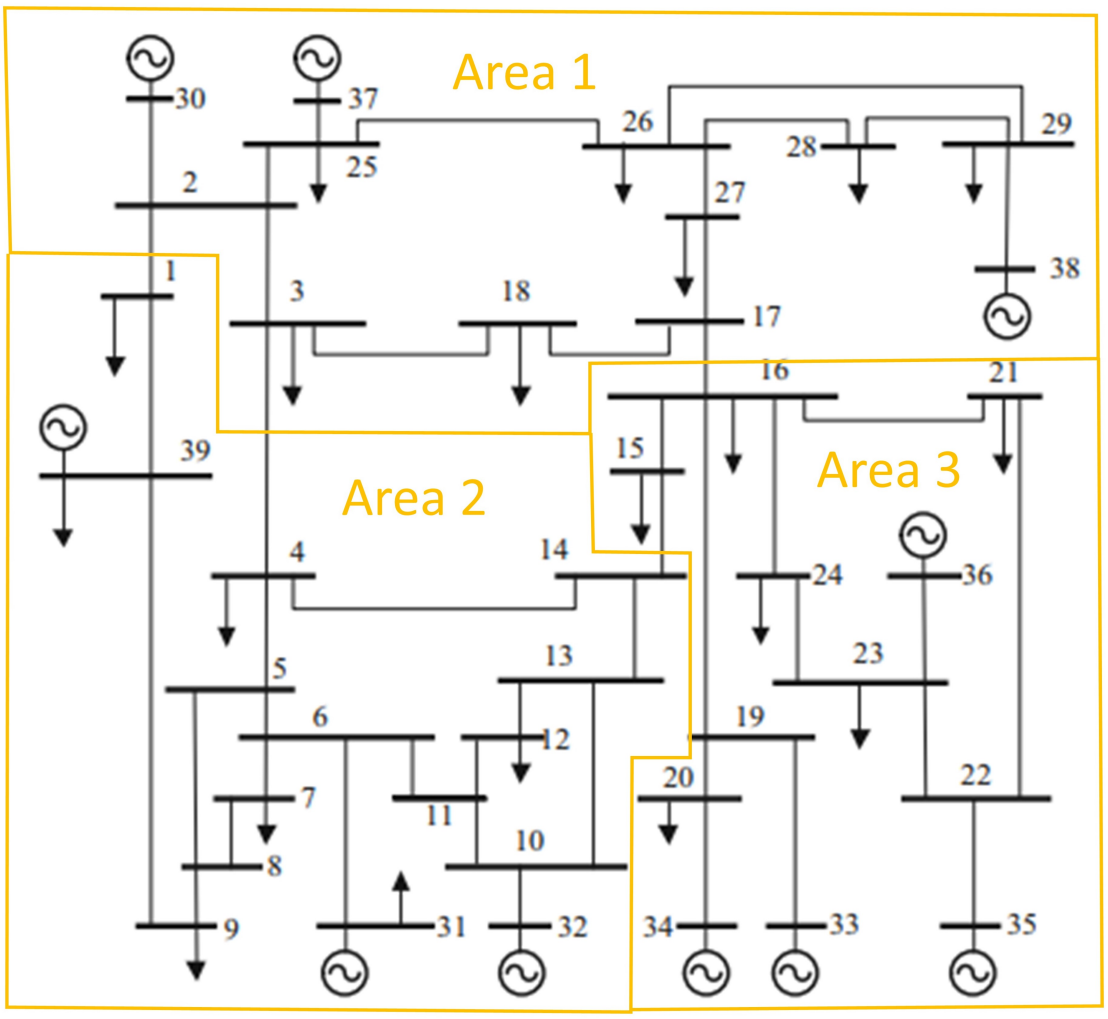

Figure 5.4: New England 39-bus system

one input layer, three fully-connected hidden layers and one output layer. The number of inputs and outputs correspond to the number of measurements for each area. The batch normalization is applied to the input and the Sigmoid function is selected as the activation function for each layer. It is worth mentioning that the attack magnitude constraints (5.14) are applied in the neural network during training. Specifically, the output a_i for each actor can be bounded by the Sigmoid function at the output of the network. For this case, a_{min} and a_{max} are set as -5MW and 5MW, respectively. As such, the agent learns the attack cannot exceed the limits. During the training process, memory replay buffer and mini-batch

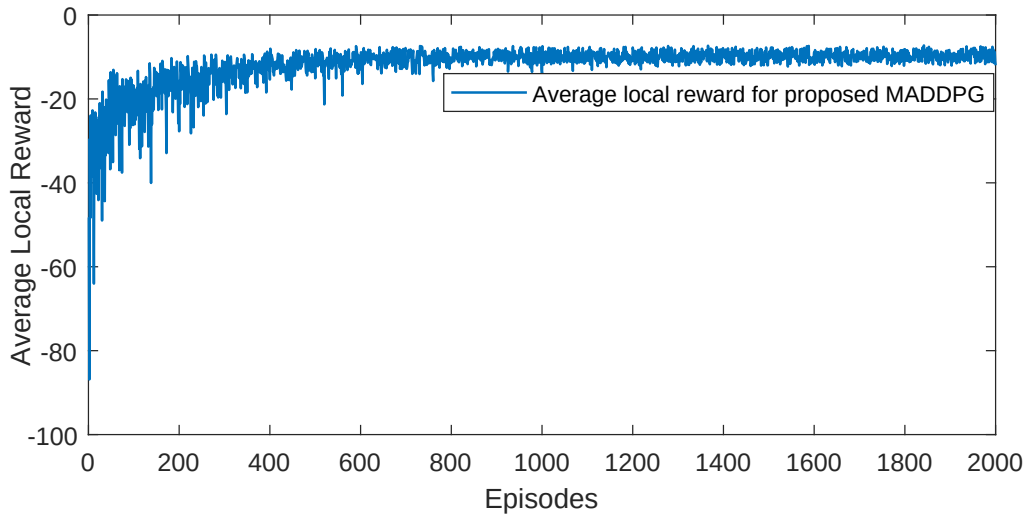

Figure 5.5: Local reward convergence curve

size are set as 1000 and 256, respectively. The network is updated by Adam optimizer to accelerate the training process with a learning rate 0.001. The exploration parameter σ_i decreases by 0.05% per time step.

5.6.2 Simulation Results

Fig. 5.5 delineates the average convergence curve for agents. It can be observed that as the training process proceeds, the average reward increases, that is, each agent learns to minimize the number of corrupt sensors. The learning curve becomes flat after 600 episodes. Note that the local reward converges a little slower than the global reward. One of the reasons is that at around $400th$ episode, the agents' policy fluctuates and is almost trapped by a local optimal point during the exploration process but then the agents climb out and ultimately achieve the goal.

In addition, since all the agents possess identical observation and action space, the number of trainable parameters of Q-function can be represented as $\mathcal{O}(n(o_{dim}+a_{dim}))$, where n is the

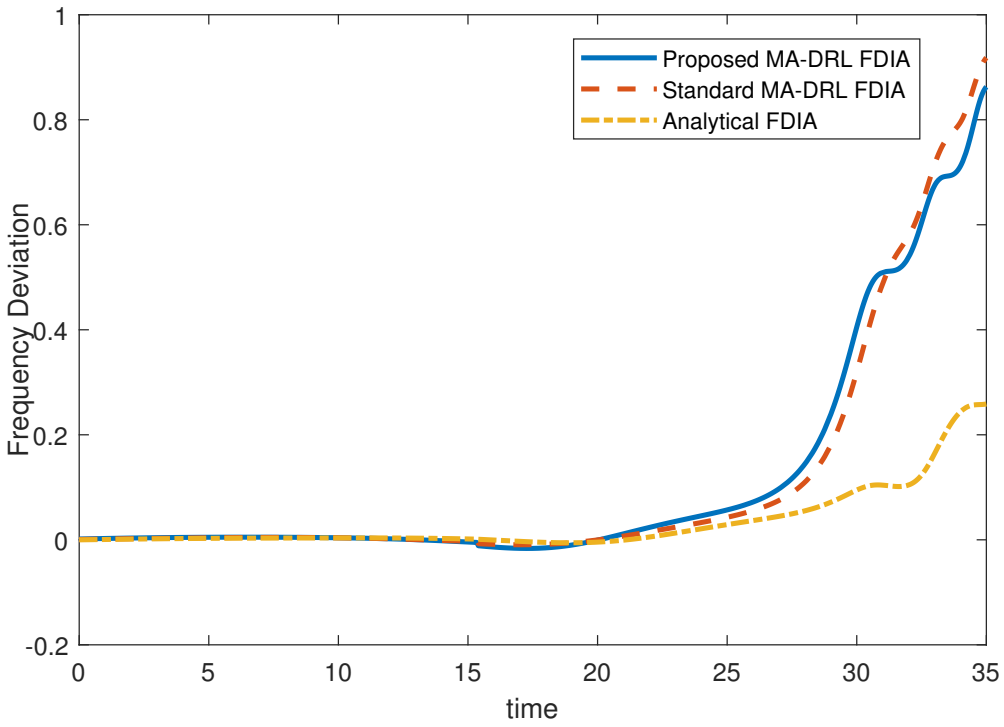

Figure 5.6: The frequency deviation comparison for different methods

number of agents; a_{dim} and o_{dim} are the dimensionality of the action and observation space, respectively. Compared with the standard MADDPG Q-function with $\mathcal{O}(n^2(o_{dim} + a_{dim}))$ parameters, the proposed MADDPG decreases the parametric space linearly.

The frequency response is presented in Fig. 5.6 to compare the attack impact of proposed MA-DRL method with standard MA-DRL and analytical FDI attack method in [25]. Note that the frequency is the average frequency deviation of all areas. The FDI attacks are implemented at 15s. The analytical FDIA method is a one-shot approach, which means the attack is only performed at 15s without subsequent attack injection, while FDI attack vectors for the other two methods are adjusted and injected every time step based on the states for that time.

It can be seen in Fig. 5.6 that the frequency is largely deviated, which implies the stability of power system is compromised. Time-to-emergency (TTE) is used for evaluating the three methods, where TTE is the time from the start of attack to the first time that the average frequency deviation of all the areas is out of a limit range. For this case, the frequency deviation limits are set as [-0.2Hz, 0.2Hz], since if the frequency deviation is out of this range, power system operator would be alerted of the abnormal situation and corresponding remedial actions would be taken, such as load shedding and generator disconnection leading to potential equipment damage.

Table 5.1: TTE comparison

Method	TTE
Proposed MA-DRL FDIA	28.6s
Standard MA-DRL FDIA	29.2s
Analytical FDIA	33.3s

Table 5.1 demonstrates the TTE comparison. The TTE of the proposed attack method is 2.1% and 16.4% faster than other two methods, which means the attacker could reach frequency deviation limits quickly and the probability of being revealed is decreased by adopting the proposed MA-DRL FDI attack.

Table 5.2: Minimum compromised measurements

	Area 1	Area 2	Area 3
Total Meas.	65	81	58
Accessible Meas.	22	27	20
Proposed FDIA	14	23	11
Standard MA-DRL FDIA	19	22	13
Analytical FDIA	18	27	13

The number of measurements that have to be compromised of each attack step is shown in Fig. 5.7. In this comparison, only the proposed FDIA and standard MA-DRL FDIA are presented since the analytical FDIA is a one-shot attack. The minimum number of attacked measurements are summarized in Table 5.2. It can be seen from the results that via

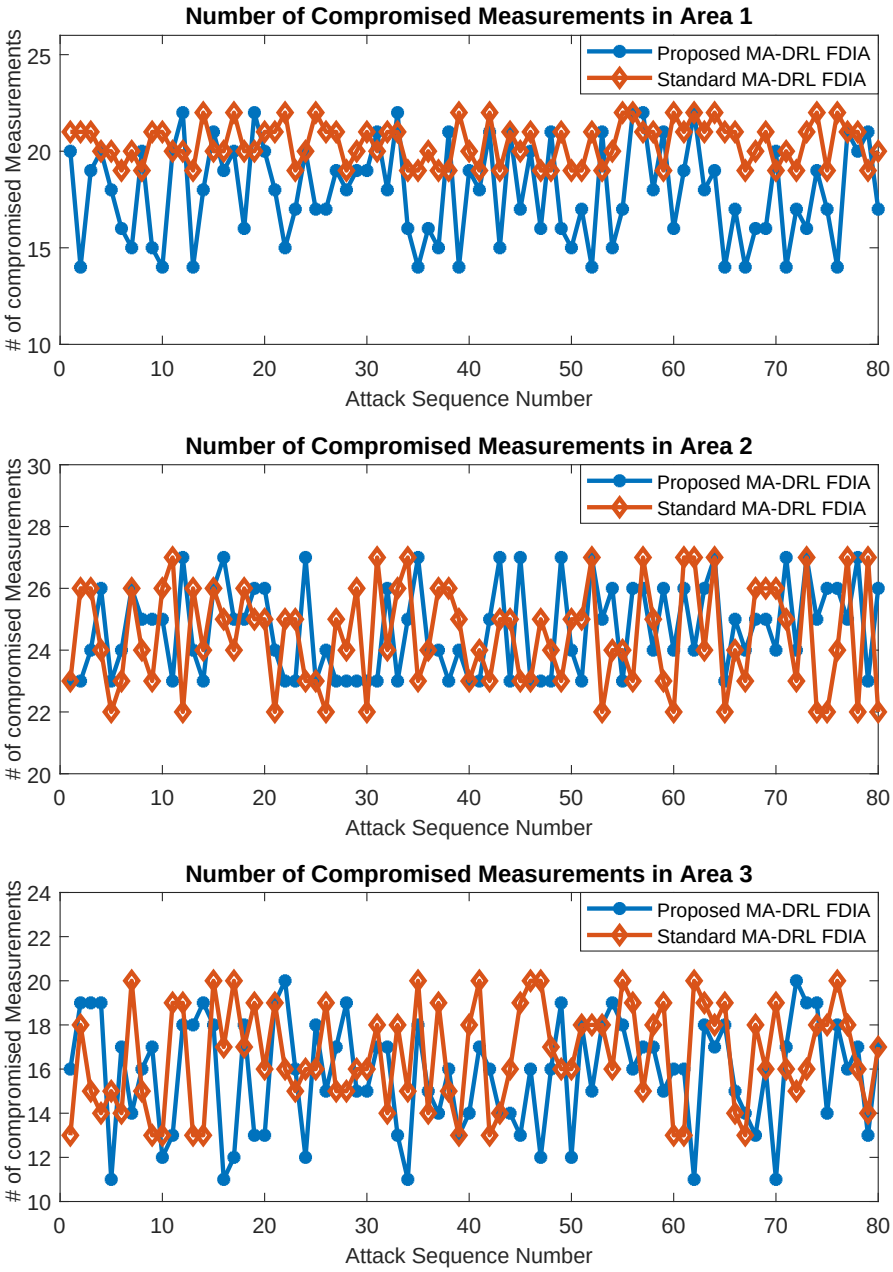

Figure 5.7: Number of compromised measurements

the proposed MA-DRL training process, the attack vectors are injected every step with the minimum number of affected measurements that can cause the largest impact for that step. Compared with other two approaches, our proposed method overall impacts less measurements, which will be more difficult to be detected by power system operator. Consequently, the proposed MA-DRL based FDIA approach may significantly sabotage the power system stability and minimize the compromised measurement to avoid being revealed at the same time.

5.6.3 Countermeasure Discussion

From the analysis of the proposed FDIA, it can be seen that the whole attack sequence is a time series combination of various single attacks. Therefore, it is difficult to verify every single measurement fed into LFC controller; and even the system possesses enough computational power, the result may not be correct because the injected attack vectors keep changing based on the system state at every attack step. One best way to tackle this problem is to analyze the critical measurements that are vulnerable to such attacks. For these vulnerable sensors, extra protection measures or technologies can be deployed such as software-defined networking (SDN) [139], up-to-date firmware [101], and physical isolation.

In order to evaluate the vulnerability of line sensors, Criticality Index (CI) is used in this research, which counts the frequency of targeted measurements in the attack sequence. The CI for the proposed FDIA against IEEE-39 bus system is shown in Fig. 5.8. It can be seen 19 measurements' CI values are larger than 0.9, which means these measurements has high probability to be targeted in the attack sequence. The sensors corresponding to these 19 measurements can be identified as critical sensors which may warrant further protection.

To analyze the effect of critical measurements with 0.9 or higher CI, 19 measurements mentioned above are removed from accessible measurements set and the frequency deviation

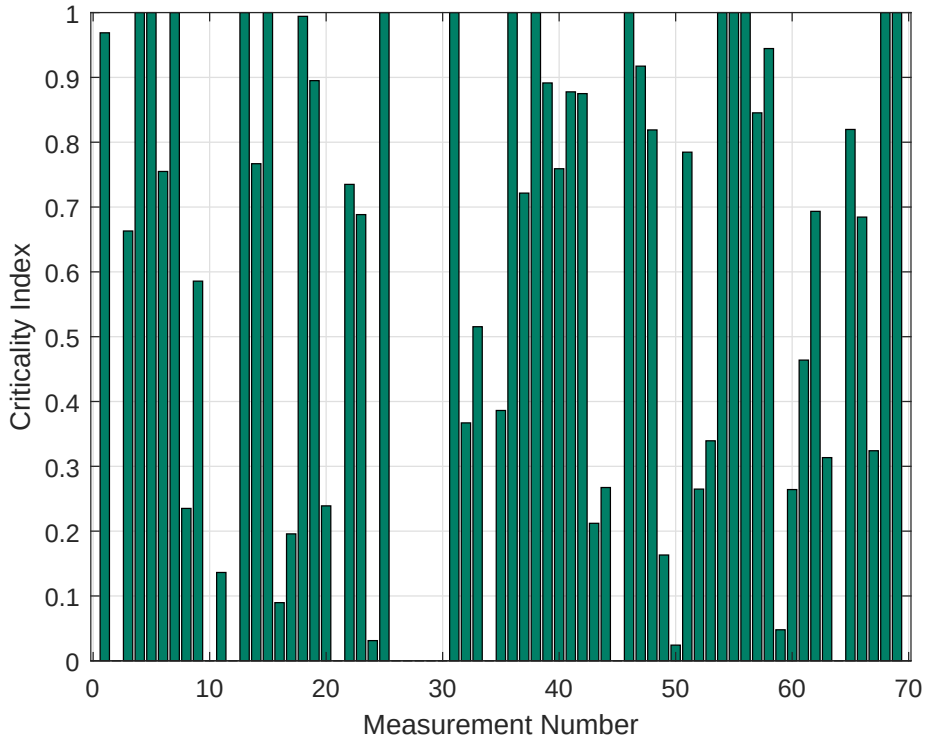

Figure 5.8: Frequency of transmission lines targeted by the proposed FDIA

due to the proposed FDIA is shown in Fig. 5.9. It indicates that deploying higher protection for the critical measurements will significantly decrease the frequency deviation and cause less damage to the power system. The results provide useful insight for power system administrators to address planning and protection issues to improve power system resiliency.

5.7 Summary

In this chapter, a MA-DRL based FDI attack model against the LFC system is proposed. Instead of using legacy linearized LFC model, AC state estimation (ACSE) is coupled with LFC to ensure the collected measurements are free of noise and bad data, thereby the system environment becomes more practical and complex. For this model, there are two attack

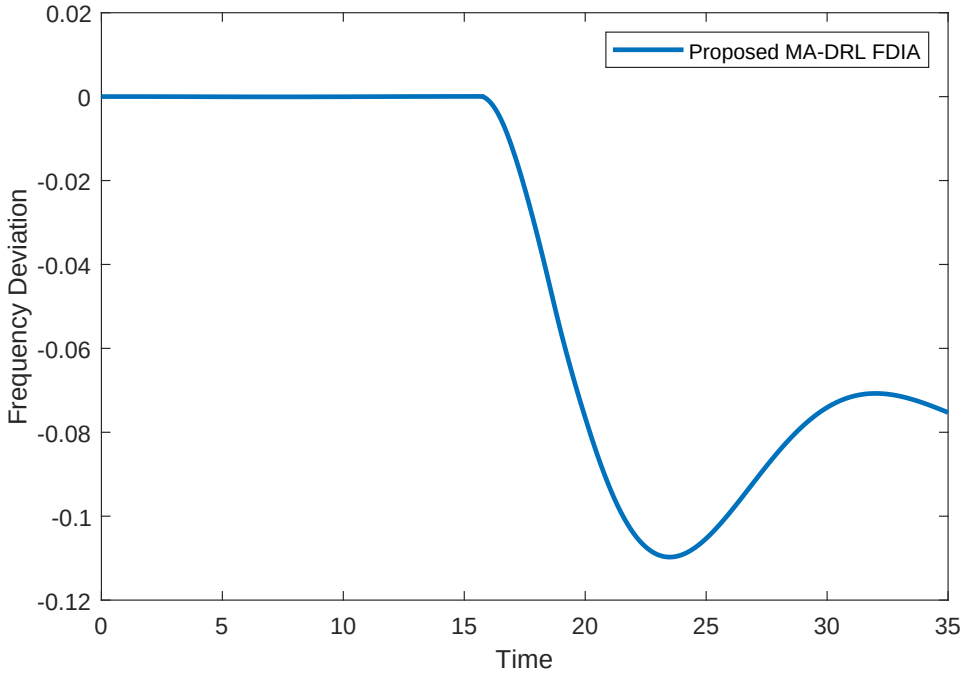

Figure 5.9: The frequency deviation when critical measurements are well-protected

objectives, i.e., stealthily maximizing the frequency deviation and minimizing the number of compromised measurements. A modified MA-DDPG algorithm is devised to achieve the two objectives simultaneously, which treats them separately by global and local individual critic networks other than a simple linear combination. The impact of the FDI attack on the LFC integrated with ACSE is also analytically derived. A Laplace-domain transfer function is obtained and can be used to replace the global critic network to accelerate the offline training process and reduce the variance. During the online execution, each attacker of LFC control area cooperatively injects the optimal attack vector only by its own local measurements. The simulation results on the New-England 39-bus system demonstrate the good performance of the proposed FDI attack model compared with other methods. In addition, corresponding countermeasures based on the critical measurements are discussed and verified. The proposed model and results provide useful insight into system planning for

power system administrators to identify and protect the critical measurements, and thereby power system resiliency is improved.

Chapter 6

CONCLUSIONS

In the previous chapters, all research work in this article are presented. A summary is showed for reviewing the major achievement of this dissertation in this chapter. In the end, some goals which have not been not fully achieved and some future work are suggested.

6.1 Summary of Results

This dissertation is focused on cyber-physical attack detection, defense and mitigation using the state-of-the-art data-driven machine learning technologies. Three research work are conducted and presented in this article. The results demonstrate the effectiveness and efficiency of proposed models on cyber-physical attack detection and mitigation. These studies provide helpful insight to power system administrators against cyber-physical attacks and thereby power system resiliency gets enhanced.

The first research is focused on reinforcement-learning-based dynamic defense strategy against one of the major cyber-physical attacks: dynamic load altering attack (D-LAA). Cascading failures of the chronological D-LAA attack is investigated. A two-player zero-sum Markov game is formulated to analyze the interactions between the attacker and the defender. A minimax-q learning algorithm is also derived to solve Nash equilibrium strategy for the attacker and defender. The performance of the proposed model is evaluated on the IEEE 39-bus system. The results are compared with passive defense strategy and verify the

advantage of the proposed dynamic defense strategy. To improve the power system resiliency, this proposed defense strategy can be deployed in advance when D-LAAs are predicted.

In the second study, a Triple Generative Adversarial Network (TripleGAN) based defense framework is derived against stealthy false data injection (FDI) attacks. The proposed model is used to effectively detect and mitigate the FDI attacks with limited historical measurement data. To improve the detection accuracy and data recovery efficiency, a extended loss function integrating with feature matching is designed. Simulation is conducted on the IEEE 118-bus system and the result verifies that the proposed defense model can accurately detect the stealthy FDI attacks and reconstruct the state estimation modified by the manipulated measurements. Further, the results confirm that the proposed techniques outperform other existing machine learning detection and recovery methods.

The last research proposes a data-driven FDI attack model based on multi-agent deep reinforcement learning (MA-DRL) against load frequency control (LFC) system integrated with AC state estimation (ACSE). Considering the different characteristics of LFC control areas in the multi-area power system, the attack vector injected in power measurements for each control area adjusts its value at each time step and cooperatively maximizes the grid frequency deviation and minimizes the number of tampered measurements at the same time. For this model, there are two attack objectives, i.e., stealthily maximizing the frequency deviation and minimizing the number of compromised measurements. A modified Multi-Agent Deep Deterministic Policy Gradient (MA-DDPG) algorithm is devised to achieve the two objectives simultaneously, which treats them separately by global and local individual critic networks other than a simple linear combination. The impact of the FDI attack on the LFC integrated with ACSE is also analytically derived. A Laplace-domain transfer function is obtained and can be used to replace the global critic network to accelerate the offline training process and reduce the variance. Simulation results on the New-England 39-bus system verify the effectiveness of the proposed FDI attack model compared with other methods. In addition, corresponding countermeasures based on the critical measurements

are discussed.

6.2 Future Research

1. Distributed algorithms will be developed to further enhance the effectiveness of the dynamic defense strategy, such as the learning automata including linear reward-inaction and linear reward-penalty.

2. The proposed TripleGAN based attack and mitigation model in chapter 4 could be generalized to cope with more types of cyber-physical attacks, which requires more elaborated model structure design.

3. For the FDI attack model, it is necessary to investigate the scenario that the attacker only knows part of system information instead of full knowledge. This will make the assumption more practical and realistic.

4. The defense method proposed in chapter 5 is to protect identified critical measurement, which may not be very efficient in cost. More powerful detection algorithm should be developed and directly integrated with LFC.

REFERENCES

[1] J. P. Farwell and R. Rohozinski, "Stuxnet and the future of cyber war," *Survival*, vol. 53, pp. 23–40, Feb. 2011.

[2] P. Shakarian, "Stuxnet: Cyberwar revolution in military affairs," Defense Tech. Inf. Center, Fort Belvoir, VA, USA, Tech. Rep., 2011.

[3] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong, "The 2015 Ukraine blackout: Implications for false data injection attacks," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3317–3318, 2017.

[4] G. N. Ericsson, "Cyber security and power system communication—essential parts of a smart grid infrastructure," *IEEE Transactions on Power Delivery*, vol. 25, no. 3, pp. 1501–1507, 2010.

[5] J. Johnson, "Can nuclear power help save us from climate change?" Chemical & Engineering News, Tech. Rep., 2019.

[6] A. Farraj, E. Hammad, and D. Kundur, "A cyber-physical control framework for transient stability in smart grids," *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 1205–1215, 2018.

[7] H. Lei, B. Chen, K. L. Butler-Purry, and C. Singh, "Security and reliability perspectives in cyber-physical smart grids," in *Proc. IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, 2018, pp. 42–47.

[8] B. Jimada-Ojuolape and J. Teh, "Impact of the integration of information and communication technology on power system reliability: A review," *IEEE Access*, vol. 8, pp. 24 600–24 615, 2020.

[9] W. Wang and Z. Lu, "Cyber security in the smart grid: Survey and challenges," *Computer networks*, vol. 57, no. 5, pp. 1344–1371, 2013.

[10] Ö. Sen, D. Van Der Veldc, P. Linnartz, I. Hacker, M. Henze, M. Andres, and A. Ulbig, "Investigating man-in-the-middle-based false data injection in a smart grid laboratory environment," in *Proc. IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*, 2021, pp. 01–06.

[11] R. V. Yohanandhan, R. M. Elavarasan, P. Manoharan, and L. Mihet-Popa, "Cyber-physical power system (cpps): A review on modeling, simulation, and analysis with cyber security applications," *IEEE Access*, vol. 8, pp. 151 019–151 064, 2020.

[12] R. Santamarta, "Here be backdoors: A journey into the secrets of industrial firmware," in *In Proceedings of the Black Hat USA*, 2012.

[13] Y. Dafalla, B. Liu, D. A. Hahn, H. Wu, R. Ahmadi, and A. G. Bardas, "Prosumer nanogrids: A cybersecurity assessment," *IEEE Access*, vol. 8, pp. 131 150–131 164, 2020.

[14] "National electric sector cybersecurity borganization resource (nescor) final technical report," EPRI, Tech. Rep., 2014. [Online]. Available: https://www.osti.gov/servlets/purl/1163840

[15] M. Zeller, "Common questions and answers addressing the aurora vulnerability," Schweitzer Engineering Laboratories, Inc., Tech. Rep., 2011.

[16] S. Maharjan, Q. Zhu, Y. Zhang, S. Gjessing, and T. Basar, "Dependable demand response management in the smart grid: A stackelberg game approach," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 120–132, 2013.

[17] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security*, vol. 14, May 2011.

[18] S. Sridhar and M. Govindarasu, "Model-based attack detection and mitigation for automatic generation control," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 580–591, 2014.

[19] Y. Yuan, Z. Li, and K. Ren, "Modeling load redistribution attacks in power systems," *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 382–390, 2011.

[20] J. Kim and L. Tong, "On topology attack of a smart grid: Undetectable attacks and countermeasures," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 7, pp. 1294–1305, 2013.

[21] Y. Cui, F. Bai, Y. Liu, P. L. Fuhr, and M. E. Morales-Rodríguez, "Spatio-temporal characterization of synchrophasor data against spoofing attacks in smart grids," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5807–5818, 2019.

[22] R. Deng, G. Xiao, and R. Lu, "Defending against false data injection attacks on power system state estimation," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 1, pp. 198–207, Feb. 2017.

[23] J. Tian, R. Tan, X. Guan, Z. Xu, and T. Liu, "Moving target defense approach to detecting stuxnet-like attacks," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 291–300, 2020.

[24] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong, "False data injection attacks targeting DC model-based state estimation," in *Proc. IEEE Power & Energy Society General Meeting*, 2017, pp. 1–5.

[25] J. Zhao, L. Mili, and M. Wang, "A generalized false data injection attacks against power system nonlinear state estimator and countermeasures," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 4868–4877, 2018.

[26] Y. Chakhchoukh, V. Vittal, and G. T. Heydt, "Pmu based state estimation by integrating correlation," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 617–626, 2014.

[27] J. Zhao, G. Zhang, M. La Scala, Z. Y. Dong, C. Chen, and J. Wang, "Short-term state forecasting-aided method for detection of smart grid general false data injection attacks," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1580–1590, 2015.

[28] J. Qi, K. Sun, and W. Kang, "Optimal pmu placement for power system dynamic state estimation by using empirical observability gramian," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 2041–2054, 2015.

[29] Y. Fan, Z. Zhang, M. Trinkle, A. D. Dimitrovski, J. B. Song, and H. Li, "A cross-layer defense mechanism against GPS spoofing attacks on pmus in smart grids," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2659–2668, 2015.

[30] X. Liu and Z. Li, "Local load redistribution attacks in power systems with incomplete network information," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1665–1676, 2014.

[31] G. Chaojun, P. Jirutitijaroen, and M. Motani, "Detecting false data injection attacks in AC state estimation," *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2476–2483, 2015.

[32] A. Ashok, M. Govindarasu, and V. Ajjarapu, "Online detection of stealthy false data injection attacks in power system state estimation," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 1636–1646, 2018.

[33] E. Al-Shaer, Q. Duan, and J. H. Jafarian, "Random host mutation for moving target defense," in *Security and Privacy in Communication Networks*, A. D. Keromytis and R. Di Pietro, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 310–327.

[34] K. R. Davis, K. L. Morrow, R. Bobba, and E. Heine, "Power flow cyber attacks and perturbation-based defense," in *Proc. IEEE Third Int. Conf. Smart Grid Communications (SmartGridComm)*, 2012, pp. 342–347.

[35] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 93–109, 2015.

[36] B. Satchidanandan and P. R. Kumar, "Dynamic watermarking: Active defense of networked cyber–physical systems," *Proceedings of the IEEE*, vol. 105, no. 2, pp. 219–240, 2017.

[37] G. S. Ledva, S. Peterson, and J. L. Mathieu, "Benchmarking of aggregate residential load models used for demand response," in *Proc. IEEE Power Energy Society General Meeting (PESGM)*, Aug. 2018, pp. 1–5.

[38] A. Molina-Garcia, F. Bouffard, and D. S. Kirschen, "Decentralized demand-side contribution to primary frequency control," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 411–419, Feb. 2011.

[39] W. Zeng, Y. Zhang, and M. Chow, "Resilient distributed energy management subject to unexpected misbehaving generation units," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 1, pp. 208–216, Feb. 2017.

[40] H. Mortaji, S. H. Ow, M. Moghavvemi, and H. A. F. Almurib, "Load shedding and smart-direct load control using internet of things in smart grid demand response management," *IEEE Transactions on Industry Applications*, vol. 53, no. 6, pp. 5155–5163, Nov. 2017.

[41] T. W. Haring, J. L. Mathieu, and G. Andersson, "Comparing centralized and decentralized contract design enabling direct load control for reserves," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 2044–2054, May 2016.

[42] A. Mohsenian-Rad and A. Leon-Garcia, "Distributed internet-based load altering attacks against smart power grids," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 667–674, Dec. 2011.

[43] A. K. Marnerides, P. Smith, A. Schaeffer-Filho, and A. Mauthe, "Power consumption profiling using energy time-frequency distributions in smart grids," *IEEE Communications Letters*, vol. 19, no. 1, pp. 46–49, Jan. 2015.

[44] C. Mellucci, P. P. Menon, C. Edwards, and A. Ferrara, "Load alteration fault detection and reconstruction in power networks modelled in semi-explicit differential algebraic equation form," in *Proc. American Control Conf. (ACC)*, Jul. 2015, pp. 5836–5841.

[45] T. Pan, S. Mishra, L. N. Nguyen, G. Lee, J. Kang, J. Seo, and M. T. Thai, "Threat from being social: Vulnerability analysis of social network coupled smart grid," *IEEE Access*, vol. 5, pp. 16 774–16 783, 2017.

[46] S. Amini, F. Pasqualetti, and H. Mohsenian-Rad, "Dynamic load altering attacks against power system stability: Attack models and protection schemes," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 2862–2872, Jul. 2018.

[47] A. Di Giorgio, A. Giuseppi, F. Liberati, A. Ornatelli, A. Rabezzano, and L. R. Celsi, "On the optimization of energy storage system placement for protecting power transmission grids against dynamic load altering attacks," in *Proc. 25th Mediterranean Conf. Control and Automation (MED)*, Jul. 2017, pp. 986–992.

[48] M. J. Osborne, *An Introduction to Game Theory*. Oxford University Press, 2004.

[49] T. Alpcan and T. Basar, "A game theoretic approach to decision and analysis in network intrusion detection," in *Proc. 42nd IEEE Int. Conf. Decision and Control (IEEE Cat. No.03CH37475)*, vol. 3, Dec. 2003, pp. 2595–2600 Vol.3.

[50] Y. W. Law, T. Alpcan, and M. Palaniswami, "Security games for risk minimization in automatic generation control," *IEEE Transactions on Power Systems*, vol. 30, no. 1, pp. 223–232, Jan. 2015.

[51] G. Chen, Z. Y. Dong, D. J. Hill, and Y. S. Xue, "Exploring reliable strategies for defending power systems against targeted attacks," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1000–1009, Aug. 2011.

[52] Q. Zhu and T. Basar, "Game-theoretic methods for robustness, security, and resilience of cyberphysical control systems: Games-in-games principle for optimal cross-layer resilient control systems," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 46–65, Feb. 2015.

[53] A. Farraj, E. Hammad, A. A. Daoud, and D. Kundur, "A game-theoretic analysis of cyber switching attacks and mitigation in smart grid systems," *IEEE Transactions on Smart Grid*, vol. 7, no. 4, pp. 1846–1855, Jul. 2016.

[54] P. Chen, S. Cheng, and K. Chen, "Smart attacks in smart grid communication networks," *IEEE Communications Magazine*, vol. 50, no. 8, pp. 24–29, Aug. 2012.

[55] M. Esmalifalak, G. Shi, Z. Han, and L. Song, "Bad data injection attack and defense in electricity market using game theory study," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 160–169, Mar. 2013.

[56] Y. Li, L. Shi, P. Cheng, J. Chen, and D. E. Quevedo, "Jamming attacks on remote state estimation in cyber-physical systems: A game-theoretic approach," *IEEE Transactions on Automatic Control*, vol. 60, no. 10, pp. 2831–2836, Oct. 2015.

[57] Z. Ni, S. Paul, X. Zhong, and Q. Wei, "A reinforcement learning approach for sequential decision-making process of attacks in smart grid," in *Proc. IEEE Symp. Series Computational Intelligence (SSCI)*, Nov. 2017, pp. 1–8.

[58] L. Wei, A. I. Sarwat, W. Saad, and S. Biswas, "Stochastic games for power grid protection against coordinated cyber-physical attacks," *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 684–694, Mar. 2018.

[59] J. Ma, Y. Liu, L. Song, and Z. Han, "Multiact dynamic game strategy for jamming attack in electricity market," *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2273–2282, Sep. 2015.

[60] Z. Ni and S. Paul, "A multistage game in smart grid security: A reinforcement learning solution," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–12, 2019.

[61] Y. He, G. J. Mendis, and J. Wei, "Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2505–2516, Sep. 2017.

[62] J. Yan, H. He, X. Zhong, and Y. Tang, "Q-learning-based vulnerability analysis of smart grid against sequential topology attacks," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 1, pp. 200–210, Jan. 2017.

[63] H. Wang, Z. Lei, X. Zhang, J. Peng, and H. Jiang, "Multiobjective reinforcement learning-based intelligent approach for optimization of activation rules in automatic generation control," *IEEE Access*, vol. 7, pp. 17 480–17 492, 2019.

[64] S. Ciavarella, J. Joo, and S. Silvestri, "Managing contingencies in smart grids via the internet of things," *IEEE Transactions on Smart Grid*, vol. 7, no. 4, pp. 2134–2141, Jul. 2016.

[65] J. D. Glover, M. S. Sarma, and T. J. Overbye, *Power System Analysis and Design*, 5th ed., Ed. Cengage Learning, 2009.

[66] J. C. M. Vieira, W. Freitas, Wilsun Xu, and A. Morelato, "Performance of frequency relays for distributed generation protection," *IEEE Transactions on Power Delivery*, vol. 21, no. 3, pp. 1120–1127, Jul. 2006.

[67] S. Kiliccote, S. Lanzisera, A. Liao, O. Schetrit, and M. Piette, "Fast dr: Controlling small loads over the internet," *Proc. ACEEE Sum. Study Energy Efficien. Build.*, pp. 196–208, Jan. 2014.

[68] L. Yao and H. Lu, "A two-way direct control of central air-conditioning load via the internet," *IEEE Transactions on Power Delivery*, vol. 24, no. 1, pp. 240–248, 2009.

[69] S. A. Raziei and H. Mohscnian-Had, "Optimal demand response capacity of automatic lighting control," in *Proc. IEEE PES Innovative Smart Grid Technologies Conf. (ISGT)*, 2013, pp. 1–6.

[70] K. Vanthournout, R. D'hulst, D. Geysen, and G. Jacobs, "A smart domestic hot water buffer," *IEEE Transactions on Smart Grid*, vol. 3, no. 4, pp. 2121–2127, 2012.

[71] T. Masuta and A. Yokoyama, "Supplementary load frequency control by use of a number of both electric vehicles and heat pump water heaters," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1253–1262, 2012.

[72] B. Otomega and T. Van Cutsem, "Undervoltage load shedding using distributed controllers," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1898–1907, Nov. 2007.

[73] Q. Wang, X. Cai, W. Tai, and Y. Tang, "A multi-stage game model for the false data injection attack against power systems," in *Proc. and Intelligent Systems (CYBER) 2018 IEEE 8th Annual Int. Conf. CYBER Technology in Automation, Control*, Jul. 2018, pp. 1450–1455.

[74] R. Ma, H. Chen, Y. Huang, and W. Meng, "Smart grid communication: Its challenges and opportunities," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 36–46, 2013.

[75] M. M. E. A. Mahmoud, J. Mišić, K. Akkaya, and X. Shen, "Investigating public-key certificate revocation in smart grid," *IEEE Internet of Things Journal*, vol. 2, no. 6, pp. 490–503, 2015.

[76] R. Hassan, M. Abdallah, G. Radman, F. Marco, S. Hammer, J. Wigington, J. Givens, D. Hislop, J. Short, and S. Carroll, "Under-frequency load shedding: Towards a smarter smart house with a consumer level controller," in *Proc. IEEE Southeastcon 2011*, 2011, pp. 73–78.

[77] E. Alpaydin, *Introduction to Machine Learning.* Cambridg, MA: MIT Press, Aug. 2012.

[78] L. S. Shapley, "Stochastic games," *Proceedings of the national academy of sciences*, vol. 39, no. 10, pp. 1095–1100, 1953.

[79] M. L. Littman, "Markov games as a framework for multi-agent reinforcement learning," in *Machine learning proceedings 1994.* Elsevier, 1994, pp. 157–163.

[80] M. Tokic, "Adaptive ε-greedy exploration in reinforcement learning based on value differences," in *Annual Conference on Artificial Intelligence.* Springer, 2010, pp. 203–210.

[81] A. Pai, *Energy Function Analysis for Power System Stability.* Springer, 1989.

[82] S. Hasan, A. Dubey, G. Karsai, and X. Koutsoukos, "A game-theoretic approach for power systems defense against dynamic cyber-attacks," *International Journal of Electrical Power & Energy Systems*, vol. 115, p. 105432, 2020.

[83] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proc. 16th ACMConf. Comput. Commun. Secur.*, 2009, pp. 21–32.

[84] A. S. Musleh, G. Chen, and Z. Y. Dong, "A survey on the detection algorithms for false data injection attacks in smart grids," *IEEE Transactions on Smart Grid*, p. 1, 2019.

[85] C. Li, Y. Wu, Y. Sun, H. Zhang, Y. Liu, Y. Liu, and V. Terzija, "Continuous under-frequency load shedding scheme for power system adaptive frequency control," *IEEE Transactions on Power Systems*, vol. 35, no. 2, pp. 950–961, 2019.

[86] B. Li, T. Ding, C. Huang, J. Zhao, Y. Yang, and Y. Chen, "Detecting false data injection attacks against power system state estimation with fast go-decomposition approach," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2892–2904, May 2019.

[87] M. Esmalifalak, L. Liu, N. Nguyen, R. Zheng, and Z. Han, "Detecting stealthy false data injection using machine learning in smart grid," *IEEE Systems Journal*, vol. 11, no. 3, pp. 1644–1652, Sep. 2017.

[88] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, and S. Mishra, "Decision tree and SVM-based data analytics for theft detection in smart grid," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 1005–1016, Jun. 2016.

[89] M. Ozay, I. Esnaola, F. T. Yarman Vural, S. R. Kulkarni, and H. V. Poor, "Machine learning methods for attack detection in the smart grid," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 8, pp. 1773–1786, Aug. 2016.

[90] M. E. Hariri, T. A. Youssef, H. F. Habib, and O. Mohammed, "Online false data detection and lost packet forecasting system using time series neural networks for IEC 61850 sampled measured values," in *Proc. IEEE Power Energy Society Innovative Smart Grid Technologies Conf. (ISGT)*, Apr. 2017, pp. 1–5.

[91] G. Fenza, M. Gallo, and V. Loia, "Drift-aware methodology for anomaly detection in smart grid," *IEEE Access*, vol. 7, pp. 9645–9657, 2019.

[92] A. Pinceti, L. Sankar, and O. Kosut, "Load redistribution attack detection using machine learning: A data-driven approach," in *Proc. IEEE Power Energy Society General Meeting (PESGM)*, Aug. 2018, pp. 1–5.

[93] D. Xue, X. Jing, and H. Liu, "Detection of false data injection attacks in smart grid utilizing elm-based ocon framework," *IEEE Access*, vol. 7, pp. 31 762–31 773, 2019.

[94] M. N. Kurt, O. Ogundijo, C. Li, and X. Wang, "Online cyber-attack detection in smart grid: A reinforcement learning approach," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5174–5185, Sep. 2019.

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

[96] Q. Yang, J. Yang, W. Yu, D. An, N. Zhang, and W. Zhao, "On false data-injection attacks against power system state estimation: Modeling and countermeasures," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 3, pp. 717–729, Mar. 2014.

[97] S. Bi and Y. J. Zhang, "Graphical methods for defense against false-data injection attacks on power system state estimation," *IEEE Transactions on Smart Grid*, vol. 5, no. 3, pp. 1216–1227, May 2014.

[98] Z. Zhang, R. Deng, D. K. Y. Yau, P. Cheng, and J. Chen, "Analysis of moving target defense against false data injection attacks on power grid," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 2320–2335, 2020.

[99] T. R. B. Kushal, K. Lai, and M. S. Illindala, "Risk-based mitigation of load curtailment cyber attack using intelligent agents in a shipboard power system," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 4741–4750, Sep. 2019.

[100] Y. Wang, M. M. Amin, J. Fu, and H. B. Moussa, "A novel data analytical approach for false data injection cyber-physical attack mitigation in smart grids," *IEEE Access*, vol. 5, pp. 26 022–26 033, 2017.

[101] J. Liu, Z. Zhao, J. Ji, and M. Hu, "Research and application of wireless sensor network technology in power transmission and distribution system," *Intelligent and Converged Networks*, vol. 1, no. 2, pp. 199–220, 2020.

[102] I. J. Goodfellow, J. P. Abadie, M. Mirza, B. Xu, D. W. Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proceedings of Neural Information Processing Systems (NIPS)*, 2014, pp. 2672–2680.

[103] Y. Li, Y. Wang, and S. Hu, "Online generative adversary network based measurement recovery in false data injection attacks: A cyber-physical approach," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 2031–2043, 2020.

[104] H. Zhang, B. Liu, and H. Wu, "Smart grid cyber-physical attack and defense: A review," *IEEE Access*, vol. 9, pp. 29 641–29 659, 2021.

[105] Z. Liu, Z. Jia, C. Vong, S. Bu, J. Han, and X. Tang, "Capturing high-discriminative fault features for electronics-rich analog system via deep learning," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 1213–1226, Jun. 2017.

[106] J. T. Springenberg, "Unsupervised and semi-supervised learning with categorical generative adversarial networks," in *ICLR*, 2015.

[107] T. Salimans, I. J. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," *CoRR*, 2016.

[108] C. Li, K. Xu, J. Zhu, and B. Zhang, "Triple generative adversarial nets," in *Neural Information Processing Systems (NIPS 2017)*, 2017.

[109] G. Wijeweera, U. D. Annakkage, W. Zhang, A. D. Rajapakse, and M. Rheault, "Development of an equivalent circuit of a large power system for real-time security assessment," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 3490–3499, Jul. 2018.

[110] J. J. Grainger and J. W. D. Stevenson, *Power System Analysis*. McGraw-Hill, 1994.

[111] N. Mohan, *Electric Power Systems*. New York, NY, USA: Wiley, 2013.

[112] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations (ICLR)*, 2015.

[113] Y. Fu, M. Shahidehpour, and Z. Li, "Long-term security-constrained unit commitment: hybrid dantzig-wolfe decomposition and subgradient approach," *IEEE Transactions on Power Systems*, vol. 20, no. 4, pp. 2093–2106, Nov. 2005.

[114] J. Yan, B. Tang, and H. He, "Detection of false data attacks in smart grid with supervised learning," in *Proc. Int. Joint Conf. Neural Networks (IJCNN)*, Jul. 2016, pp. 1395–1402.

[115] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014.

[116] A. M. Mohan, N. Meskin, and H. Mehrjerdi, "A comprehensive review of the cyber-attacks and cyber-security on load frequency control of power systems," *Energies*, vol. 13, no. 15, p. 3860, 2020.

[117] Y. Guo, L. Wang, Z. Liu, and Y. Shen, "Reinforcement-learning-based dynamic defense strategy of multistage game against dynamic load altering attack," *International Journal of Electrical Power & Energy Systems*, vol. 131, p. 107113, 2021.

[118] L. Zhu, M. Li, Z. Zhang, X. Du, and M. Guizani, "Big data mining of users' energy consumption patterns in the wireless smart grid," *IEEE Wireless Communications*, vol. 25, no. 1, pp. 84–89, 2018.

[119] X. He, X. Liu, and P. Li, "Coordinated false data injection attacks in AGC system and its countermeasure," *IEEE Access*, vol. 8, pp. 194 640–194 651, 2020.

[120] Y. Wu, Z. Wei, J. Weng, X. Li, and R. H. Deng, "Resonance attacks on load frequency control of smart grids," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4490–4502, 2018.

[121] Y. Li, R. Huang, and L. Ma, "False data injection attack and defense method on load frequency control," *IEEE Internet of Things Journal*, vol. 8, no. 4, pp. 2910–2919, 2021.

[122] K. Xiahou, Y. Liu, and Q. H. Wu, "Decentralized detection and mitigation of multiple false data injection attacks in multiarea power systems," *IEEE Journal of Emerging and Selected Topics in Industrial Electronics*, vol. 3, no. 1, pp. 101–112, 2022.

[123] A. G. Pillai, E. R. Samuel, and A. Unnikrishnan, "Optimal load frequency control through combined state and control gain estimation for noisy measurements," *Protection and Control of Modern Power Systems*, vol. 5, no. 1, pp. 1–12, 2020.

[124] G. Hug and J. A. Giampapa, "Vulnerability assessment of AC state estimation with respect to false data injection cyber-attacks," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1362–1370, 2012.

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

[126] Z. Li, M. Shahidehpour, A. Alabdulwahab, and A. Abusorrah, "Bilevel model for analyzing coordinated cyber-physical attacks on power systems," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2260–2272, 2016.

[127] X. Liu, Z. Bao, D. Lu, and Z. Li, "Modeling of local false data injection attacks with reduced network information," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1686–1696, 2015.

[128] Z.-H. Yu and W.-L. Chin, "Blind false data injection attack using PCA approximation method in smart grid," *IEEE Transactions on Smart Grid*, vol. 6, no. 3, pp. 1219–1226, 2015.

[129] J. Kim, L. Tong, and R. J. Thomas, "Subspace methods for data attack on state estimation: A data driven approach," *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1102–1114, 2015.

[130] J. Liang, L. Sankar, and O. Kosut, "Vulnerability analysis and consequences of false data injection attack on power system state estimation," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3864–3872, 2016.

[131] M. Jin, J. Lavaei, and K. H. Johansson, "Power grid AC-based state estimation: Vulnerability analysis against cyber attacks," *IEEE Transactions on Automatic Control*, vol. 64, no. 5, pp. 1784–1799, 2019.

[132] M. Du, G. Pierrou, X. Wang, and M. Kassouf, "Targeted false data injection attacks against AC state estimation without network parameters," *IEEE Transactions on Smart Grid*, vol. 12, no. 6, pp. 5349–5361, 2021.

[133] R. Jiao, G. Xun, X. Liu, and G. Yan, "A new AC false data injection attack method without network information," *IEEE Transactions on Smart Grid*, vol. 12, no. 6, pp. 5280–5289, 2021.

[134] R. Tan, H. H. Nguyen, E. Y. S. Foo, D. K. Y. Yau, Z. Kalbarczyk, R. K. Iyer, and H. B. Gooi, "Modeling and mitigating impact of false data injection attacks on automatic generation control," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 7, pp. 1609–1624, 2017.

[135] X. Chen, S. Hu, Y. Li, D. Yue, C. Dou, and L. Ding, "Co-estimation of state and FDI attacks and attack compensation control for multi-area load frequency control systems

under FDI and dos attacks," *IEEE Transactions on Smart Grid*, vol. 13, no. 3, pp. 2357–2368, 2022.

[136] H. H. Alhelou, M. E. H. Golshan, and N. D. Hatziargyriou, "Deterministic dynamic state estimation-based optimal lfc for interconnected power systems using unknown input observer," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1582–1592, 2019.

[137] R. Lowe, Y. I. Wu, A. Tamar, J. Harb, O. Pieter Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," *Advances in neural information processing systems*, vol. 30, 2017.

[138] A. Moeini, I. Kamwa, P. Brunelle, and G. Sybille, "Open data ieee test systems implemented in simpowersystems for education and research in power grid dynamics and control," in *2015 50th International Universities Power Engineering Conference (UPEC)*. IEEE, 2015, pp. 1–6.

[139] X. Dong, H. Lin, R. Tan, R. K. Iyer, and Z. Kalbarczyk, "Software-defined networking for smart grid resilience: Opportunities and challenges," in *Proceedings of the 1st ACM Workshop on Cyber-Physical System Security*, 2015, pp. 61–68.