Graduate Theses and Dissertations

Iowa State University Capstones, Theses and Dissertations

2019

Data-driven cyber attack detection and mitigation for decentralized wide-area protection and control in smart grids

Pengyuan Wang
Iowa State University

Recommended Citation

Wang, Pengyuan, "Data-driven cyber attack detection and mitigation for decentralized wide-area protection and control in smart grids" (2019). *Graduate Theses and Dissertations*. 17597.
https://lib.dr.iastate.edu/etd/17597

Data-driven cyber attack detection and mitigation for decentralized wide-area protection and control in smart grids

by

Pengyuan (Bruce) Wang

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Co-majors: Electrical Engineering; Computer Engineering

Program of Study Committee:
Manimaran Govindarasu, Major Professor
Ajjarapu Venkataramana
Doug Jacobson
Umesh Vaidya
Kris De Brabanter

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2019

DEDICATION

TO MY BELOVED FAMILY.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

ACKNOWLEDGEMENTS

First and foremost, sincere gratitude and thankfulness goes to my advisor Dr. Manimaran Govindarasu for his priceless guidance and advice on my research and career development during the last six years. Dr. Govindarasu has introduced me to THE field. He is always able to wisely lift the bar to the right height in the right time and then successfully persuades me to embrace the challenges. I have learned a lot from him about both research and life, as a student and also as a friend. “Always try to solve real problems” is what he encourages us to achieve and will be my life-long motto.

I really appreciate the valuable feedback from my committee members. More importantly, in another sense, my committee members are also the ones who help lay the foundation of my future career. Dr. Jacobson teaches me cyber security, the way that Dr. Ajjarapu explains the power system stability is so impressive that it will never be forgotten once you remember, Dr. Vaidya makes the non-linear system no longer obscure for me, and Kris shows me the entrance of a magnificent palace made of data. I’m grateful to what I got a chance to learn about in your classes.

My family has always been an endless source of support during my PhD time. The encouragement from my girlfriend, Elsa, is the best cure when I’m in depression. I would not be able to reach this milestone without the company of you guys. I have realized, only recently, that all kinds of difficulties become trivial when you can picture the thing that you really want to chase in your life, and this is especially true for the life with you all being part of it.

Acknowledgement should also be presented to all my group members, especially Aditya and Srayashi, who actively lend their opinions in some really inspiring discussions and always offer great help with peer reviews.

Last but not least, I’m glad that I have developed the addiction to basketball during my PhD and I’m thankful to getting to know all the mates that I’ve met on the courts.

ABSTRACT

Modern power systems have already evolved into complicated cyber physical systems (CPS), often referred to as smart grids, due to the continuous expansion of the electrical infrastructure, the augmentation of the number of heterogeneous system components and players, and the consequential application of a diversity of information and telecommunication technologies to facilitate the Wide Area Monitoring, Protection and Control (WAMPAC) of the day-to-day power system operation. Because of the reliance on cyber technologies, WAMPAC, among other critical functions, is prone to various malicious cyber attacks. Successful cyber attacks, especially those sabotage the operation of Bulk Electric System (BES), can cause great financial losses and social panics. Application of conventional IT security solutions is indispensable, but it often turns out to be insufficient to mitigate sophisticated attacks that deploy zero-day vulnerabilities or social engineering tactics.

To further improve the resilience of the operation of smart grids when facing cyber attacks, it is desirable to make the WAMPAC functions per se capable of detecting various anomalies automatically, carrying out adaptive activity adjustments in time and thus staying unimpaired even under attack. Most of the existing research efforts attempt to achieve this by adding novel functional modules, such as model-based anomaly detectors, to the legacy centralized WAMPAC functions. In contrast, this dissertation investigates the application of data-driven algorithms in cyber attack detection and mitigation within a decentralized architecture aiming at improving the situational awareness and self-adaptiveness of WAMPAC.

First part of the research focuses on the decentralization of System Integrity Protection Scheme (SIPS) with Multi-Agent System (MAS), within which the data-driven anomaly detection and optimal adaptive load shedding are further explored. An algorithm named as Support Vector Machine embedded Layered Decision Tree (SVMLDT) is proposed for the anomaly detection, which provides satisfactory detection accuracy as well as decision-making interpretability. The adaptive

load shedding is carried out by every agent individually with dynamic programming. The load shedding relies on the load profile propagation among peer agents and the attack adaptiveness is accomplished by maintaining the historical mean of load shedding proportion. Load shedding only takes place after the consensus pertaining to the anomaly detection is achieved among all interconnected agents and it serves the purpose of mitigating certain cyber attacks. The attack resilience of the decentralized SIPS is evaluated using IEEE 39 bus model. It is shown that, unlike the traditional centralized SIPS, the proposed solution is able to carry out the remedial actions under most Denial of Service (DoS) attacks.

The second part investigates the clustering based anomalous behavior detection and peer-assisted mitigation for power system generation control. To reduce the dimensionality of the data, three metrics are designed to interpret the behavior conformity of generator within the same balancing area. Semi-supervised K-means clustering and a density sensitive clustering algorithm based on Hieararchical DBSCAN (HDBSCAN) are both applied in clustering in the 3D feature space. Aiming to mitigate the cyber attacks targeting the generation control commands, a peer-assisted strategy is proposed. When the control commands from control center is detected as anomalous, i.e. either missing or the payload of which have been manipulated, the generating unit utilizes the peer data to infer and estimate a new generation adjustment value as replacement. Linear regression is utilized to obtain the relation of control values received by different generating units, Moving Target Defense (MTD) is adopted during the peer selection and 1-dimensional clustering is performed with the inferred control values, which are followed by the final control value estimation. The mitigation strategy proposed requires that generating units can communicate with each other in a peer-to-peer manner. Evaluation results suggest the efficacy of the proposed solution in counteracting data availability and data integrity attacks targeting the generation controls. However, the strategy stays effective only if less than half of the generating units are compromised and it is not able to mitigate cyber attacks targeting the measurements involved in the generation control.

CHAPTER 1. INTRODUCTION

As one of the largest and most complicated artificial critical infrastructures, power girds have been undergoing rapid development and evolution worldwide since Thomas Edison and his company constructed the first direct current power system in New York City in September 1882. Over years, countless unprecedented challenges continually arise and the sophistication of power system planning, operation and maintenance never stops increasing. In the meanwhile, great number of assorted technical innovations have either been adopted by or directly emerged from the power industry. As of now, the modern power systems have already evolved into what is referred to as smart grids, with salient features such as the replacement of fossil energy resources with renewable resources, the bi-directional flow of both the electricity and information and also grid monitoring and control that is "closer" to real-time. Though the ultimate goal of power industry is to deliver the electricity to the customers, reliable and economic power generation, transmission and consumption heavily rely on heterogeneous information and communication technologies (ICT). Therefore, a smart grid is often perceived as a Cyber Physical System (CPS), which comprises a physical layer transmitting the electricity from power plants to load centers and a cyber layer that oversees and manages the operation of the physical layer.

1.1 Smart Grid Overview

According to NIST (National Institute of Science and Technology) [1], a modern power system is composed of the following domains: generation, transmission, distribution, customers, markets, operations and service providers, etc. In general, the operation of every domain relies on various interconnected primary equipment and secondary devices. Electric energy and data flow continually among these domains as depicted in Figure 1.1. The energy grid is often referred to as primary system or physical layer and the communication network and all end nodes connected to it is

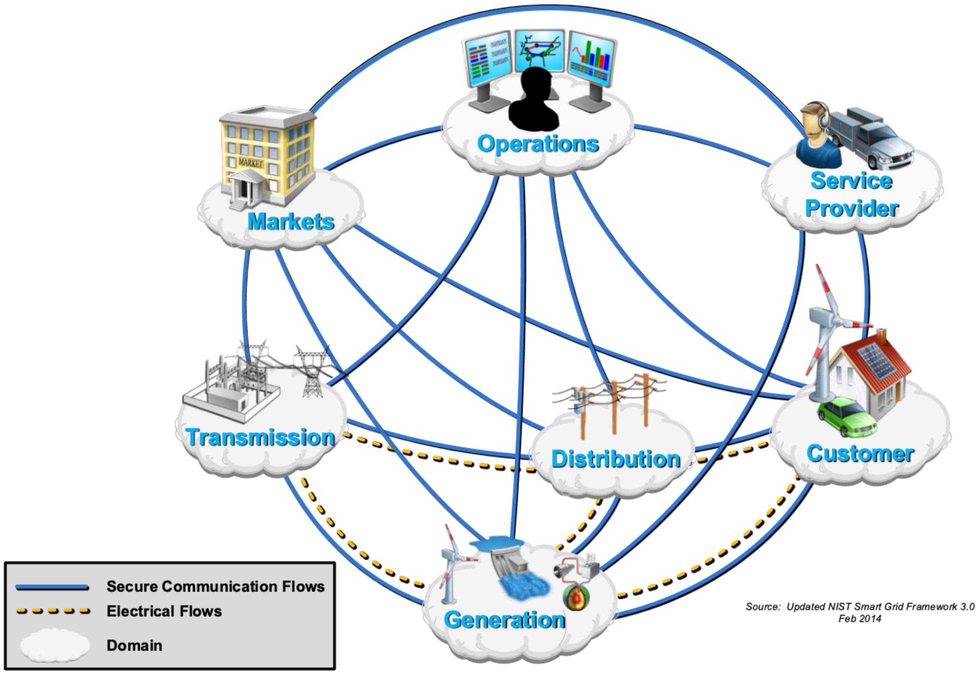


Figure 1.1 Smart grid domains [1]

referred to as secondary system or cyber layer. In this context, all the equipment directly involved in the electricity generation, transmission and consumption process belong to the physical layer and all the intelligent electronic devices (IED), networking devices, computers and servers equipped to monitor, protect and control this end-to-end process constitute the cyber layer.

1.1.1 Physical Layer

The physical layer of a power system contains four aforementioned domains: generation, transmission, distribution and consumption. The expansion and upgrade of the physical layer never really stops since 1890's due to the continuous increase in electricity demand. For the last two decades, the physical layer of power systems has become more heterogeneous. Renewable energy farms, Flexible AC Transmission System(FACTS) equipment, High Voltage Direct Current Transmission (HVDC), micro-grids, electricity storage and electric vehicles, etc. have been gradually integrated. Accommodation of these new elements is inevitable and it also brings new challenges for the operation of power systems.

1.1.1.1 Generation

Electricity generation is the process of converting none-electric energy to electricity, which is clean and easy to transmit over long distance. Common resources utilized in generation include hydro power, nuclear, natural gas, coal, wind, solar, geothermal energy, etc. and the energy conversion is usually achieved via different types of turbines and generators.

Traditionally the electricity is generated by large generators at power plants close to the energy sources in order to improve the generation efficiency. Great number of big power plants were built up during 20th century worldwide. As of now, the Three Gorges Dam in China is the largest power plant in the world, which has 34 hydro generating units in total with the maximal capacity as 225,000 MW. Nowadays, the generation of small scales also become attractive such as micro-grids containing roof PV panels and wind turbines.

1.1.1.2 Transmission

Due to the dispersed distribution of energy resources, which are often far from the load center, people have built High Voltage Alternate Current (HVAC) electricity transmission system to deliver the electricity from power plants to customers over long distance. To reduce the energy loss during transmission (i.e. to reduce the line current), electric voltage level needs to be boosted with transformers before electricity gets sent out of power plants. In US, the commonly used AC high voltage levels include 138kV, 230kV, 345kV, 500kV and 765kV. Recently, High Voltage Direct Current (HVDC) transmission is becoming popular because of the progress achieved in power electronic devices manufacturing. The major benefits that HVDC provides include more economic solution of electricity delivery over very long distance compared to HVAC, connecting two AC systems operating at different frequencies, better system stability control, etc. Conventionally, electricity generation and high voltage transmission together is referred to as Bulk Electric System

(BES) and its reliability is of the most importance for the daily operation of power systems since severe disturbances take place in BPS can potentially result in catastrophic blackouts influencing a multitude of customers across a vast region.

1.1.1.3 Distribution

When the electric energy has been sent to load centers, the electric voltage needs to be stepped down in distribution substations. Common voltage levels at the distribution system vary from 4kV to 35kV, some industry customers can be directly connected to this voltage level and for residential and other commercial users, the voltage needs to be further lowered to 120V/220V. Besides, distribution systems allow the direct connection of renewable power plants such as wind and solar farms.

1.1.1.4 Consumption

In the old days, the domain of consumption means nothing but passive loads for power industry. But today, electricity consumers has become interactive with the utilities. For example, the Distributed Energy Resource (DER) technology becomes mature so customers can choose to install their own DER generation, which needs to be accommodated properly if they need to be connected back to the distribution system. On the other hand, utilities find that demand response can be adopted as an alternative to improve the reliability and security of system operation other than solely putting more capital investment to strengthen the power grid.

1.1.2 Cyber Layer

Electricity is something that cannot be stored in large amount, and therefore, the balance of generation and consumption needs to be attained in real time. Due to this and also the complexity of the interconnection and interaction within the physical layer, almost every piece of equipment of the physical layer has to be monitored, protected or controlled whenever it is energized. This is where the cyber layer functions come into play.

The cyber layer of a smart grid often comprises Information technology (IT) network and Operation Technology (OT) network. In Figure 1.1, for instance, the market and service provider are both part of the IT network while the "operations" is part of the OT network. In general, the IT network hosts functions such as corporate web server, market functions, cloud functions etc. that are not directly related to the operation of the power grid and it allows public access to a certain extent. In contrast, the OT network functions are more critical in a sense that they directly monitors and controls the day-to-day operation of the power system. IT, OT and public network are typically isolated with firewalls and authentication mechanisms like VPN (virtual private network).

Like any others communication networks, the cyber layer of power system contains large number of end nodes, routing devices and communication channels. End nodes in this layer include IEDs (such as phasor measurement units, smart meters, relays), data pre-processing modules (like remote terminal units, phasor data concentrators), decision making scheme modules, etc. End nodes have to gather, report and exchange data/information with each other to achieve certain pre-designed functions.

The geographic span of the communication in cyber layer is determined by the needs of various functions. But in general, system wide communication is getting more and more popular so as to facilitate the function realization and improve the speed of operations. The two most critical wide area communication systems in smart grids are Supervisory Control and Data Acquisition (SCADA) system and Wide Area Measurement System (WAMS).

1.1.2.1 SCADA

SCADA is widely used for industrial automation system's monitoring and control. In power system, each station has a Remote Terminal Unit (RTU) that gathers data from the local secondary devices such as transducers, sensors and meters. SCADA collects data from RTUs located at different stations and transmits them to the control center via wide area communication every a few seconds. Applications including State Estimation (SE), Automatic Generation Control (AGC), economic dispatch, contingency analysis, wide area protection and control schemes, etc. utilize those data to monitor the operation status of power system, decide the necessary controls to launch and then send control commands back to RTUs and close the control loop. The communication protocol commonly used in SCADA system by North American utilities is DNP3, and the version that is widely used still lacks security features such as encryption or authentication.

1.1.2.2 WAMS

With the advent of Phasor Measurement Unit (PMU) and synchronized measurement technology, Wide Area Monitoring System (WAMS) emerges as an alternative to SCADA, which collects data at a much higher sampling rate and so that many applications can potentially be realized close to real time [2]. The typical sample rate in WAMS is 30, 60 or 120 samples/second. With all PMUs being synchronized to the same time source, GPS for instance, WAMS has the access not only to the voltage or current magnitudes, it can also measure the relative phasor angles directly. WAMS is often constructed in a hierarchical architecture, where Phasor Data Concentrators (PDC) at different levels will first concentrate the data from PMUs and then forward the processed data to the control center. Nowadays, applications developed based on WAMS are mainly for monitoring, such as transient instability detection and linear state estimation. Wide area protection and controls based on WAMS data are also under development. More and more applications will be put into use when the data accuracy of WAMS gets further improved. North American Synchrophasor Initiative (NASPI) has showed a road map about the application of synchronized measurement technology in 2006 [2]. IEEE C37.118 is often adopted as the communication protocol in WAMS.

1.1.2.3 Wide Area Monitoring, Protection and Control

A group of critical functions, referred to as Wide Area Monitoring, Protection and Control (WAMPAC), exists in the cyber layer of smart grid, which help maintain the secure, reliable and economic operation of the smart grid. WAMPAC functions collect wide area measurements cyclically to attain a system-wide profile of the grid operation state, based on what coordinated and economic controls can be determined and performed appropriately. Common WAMPAC functions

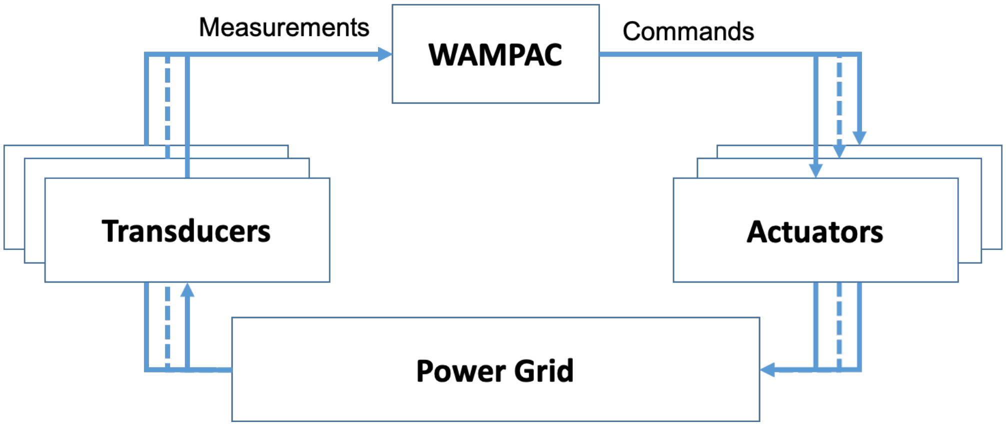


Figure 1.2   WAMPAC in a closed control loop

include System Integrity Protection Schemes (SIPS, also known as remedial action schemes, system protection schemes, etc.), functions encompassed by the Energy Management System (EMS) such as State Estimation (SE), Automatic Generation Control (AGC), Optimal Power Flow (OPF), real time contingency analysis, etc. and those rely on WAMS, which include phase angle monitoring (PAM), power oscillation monitoring (POM), power damping monitoring (PDM), voltage stability monitoring (VSM), and dynamic line rating, etc. [3]

Wide area monitoring is responsible for data collection from geographically dispersed stations and it is capable of identifying disturbances and contingencies such as inter-area oscillation. Beyond operation monitoring, wide area protection and control checks the grid operating state against predefined conditions and then determines the most appropriate remedial actions or control activities. The protection or control actions will be triggered so as to alleviate undesirable operation conditions such as system oscillation and continuous frequency deviation, etc.

Figure 1.2 depicts the WAMPAC as a closed control loop. The three major actions included in the operation of WAMPAC, i.e. measurement collection, decision making and command launching, heavily rely on ICT. High speed networks and modern communication protocols ensure that the measurements and control signals get delivered in time to meet the operation requirements of critical infrastructures. The SCADA and WAMS systems introduced previously both rely on the telecommunication to interconnect substations and control center involved in the WAMPAC.

Similarly, high speed processors and efficient data processing algorithms adopted in the EMS and other standalone WAMPAC modules guarantee the wide area protection and control decisions can be achieved accurately and efficiently. After data analysis, the required control commands will be sent from control center to distributed substations for actuation through the same telecommunication networks. Every link in the closed-loop of WAMPAC needs to stay flawless to guarantee the satisfactory efficacy of the entire function. Since WAMPAC directly interacts the operation of the power grids, its malfunction can be catastrophic.

The wide deployment of ICT has greatly facilitated the management of the smart grids by enabling the implementation of WAMPAC. However, the cyber layer risks also become a great concern recently. Cyber security didn't get as much attention as the reliability in most critical infrastructure such as the power sector in the past. The main purpose of deploying ICT was to speed up the data transmission so that the timing critical WAMPAC functions don't get deteriorated because of latency. Therefore, we can find that the protocols used for smart grid communication often lack encryption or authentication mechanisms and many IEDs and station gateways are even directly connected to the internet. Such breach of cyber security results in the exposure of critical infrastructure to all types of cyber threats as shown by Figure.1.3. The cyber security of WAMPAC in a smart grid will be elaborated in chapter II.

1.2 Research Objective and Contribution

Considering the reliance of the operation of smart grids on ICT and the importance of WAMPAC, this dissertation selects wide area protection and wide area control functions as study objects and investigates the cyber attack detection and mitigation through architecture and function redesign so that the situational awareness, self-adaptivenes and hence the cyber attack resiliency of these critical functions get enhanced. The major contributions are listed as following.

- **Wide area protection (Chapter III)** - A novel Multi-Agent System (MAS) based protection architecture is adopted in the implementation of a load shedding protection scheme. A data-driven attack detection algorithm with good interpretability named as Support Vector

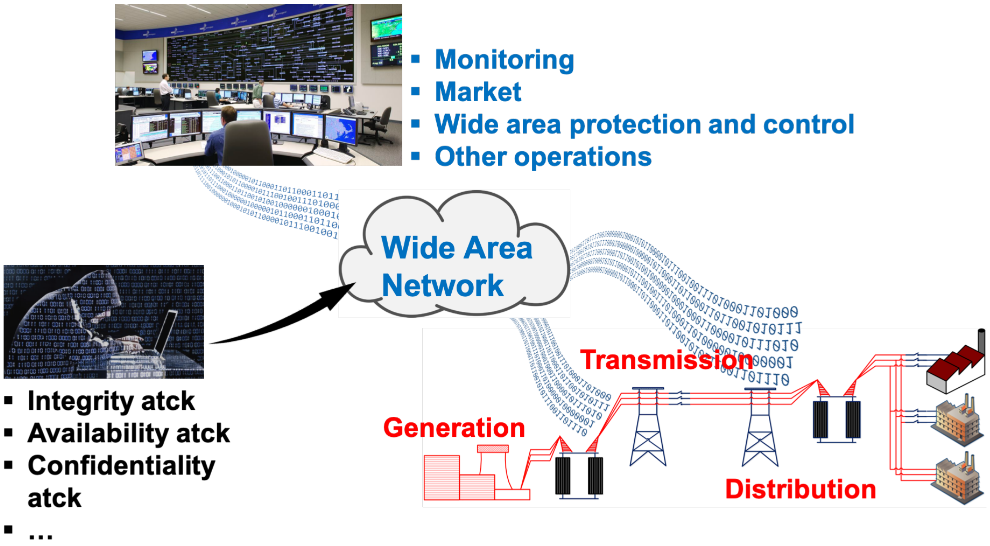


Figure 1.3 Cybersecurity concerns of smart grid

Machine embedded Layered Decision Tree (SVMLDT) is proposed. Besides, an adaptive optimal load shedding strategy is designed to mitigate the impacts of Denial of Service (DoS) attacks.

- **Wide area control (Chapter IV)** - A station-level anomaly detection module based on clustering is proposed for the generation control, which mainly depends on the behavior correlation among peer generators. To counteract the malicious attacks targeting supplementary generation control, a peer-assisted control signal estimation algorithm is presented which addresses the compromised control commands by replacing them with estimated control values based on the information received from the peers with "good reputation". The proposed solution can serve as a countermeasure to data availability and integrity attacks targeting the generation control commands.

CHAPTER 2. CYBERSECURITY OF WAMPAC IN SMART GRIDS

Although the advancement of ICT and its application have facilitated the WAMPAC functions' realization and improved the overall efficiency of operation and maintenance of power grids, the deployment of off-the-shelf commercial ICT solutions also results in the cybersecurity issue. This chapter overviews the cybersecurity vulnerabilities and threats existing in smart grids, briefs the available countermeasures, and then discusses the opportunities to enhance the cybersecurity of critical grid functions - WAMPAC.

2.1 Cyber Vulnerabilities of WAMPAC

Cyber security was not a major concern during the WAMPAC design twenty years ago or even later, when non-digital end-to-end copper cables were still widely used in substation communication and "security through obscurity" worked well so that not many cyber attack trials were observed or reported. When people started to resort to digital WAMPAC solutions based on ICT, vendors still used to focus on improving the availability and reliability of these functions, without fully considering the potential impacts resulted from cyber disturbances. Therefore, as a legacy, most WAMPAC functions in today's power industry lack the necessary security mechanisms and therefore are vulnerable to cyber attacks.

- **Public network connectivity:** Physical segregation of the IT and OT networks in power sector is inconvenient and also expensive. To reduce the implementation cost, it's not rare that utilities choose to leverage public accessible networks (Internet) to realize the data/information transmission involved in WAMPAC [4]. Most of the cases, the corporate networks of utilities (IT networks) are connected to the Internet such that the third party service providers and consumers can access the data and information that are intentionally shared with them. In the meanwhile, the OT network is connected to the IT network such that the operators can have the remote access to the grid control system. Although the IT and OT networks are often separated with DMZ (Demilitarized Zone) as a way of isolation, the adversary still possess various means to infiltrate into the OT network by bypassing the authentication or ingress filtering mechanisms provided by the DMZ. For instance, the connection can be established with the socket request initiated from the OT network or the leakage of the remote access credentials to the OT network from an IT domain workstation. The integration and overlap between the IT and OT networks provide the attackers with potential chances to gain the access to the critical OT functions [5]. When the attacker finds his foothold in the OT network, he is able to access the sensitive data, alter critical information and even manipulate the settings of digital devices, etc.

- **Communication protocol:** Most communication protocols used today in power industry like Modbus and DNP3 do not provide security measures. The lack of data encryption and authentication makes it possible for the threat actors to alter, intercept and spoof the data in transmission. Data theft leads to privacy concerns, and even worse, the adversary can exploit the vulnerabilities of the communication protocol to launch data integrity attack so as to cause malfunctioning by sending erroneous data or command signals to the WAMPAC functions or the actuators involved.

- **IT and OT devices:** The components that constitute the IT and OT networks in power system include software, computer operating systems, IEDs, routing devices, etc. Well-known vulnerabilities of these components can be directly exploited by the attackers, and it is also not difficult for the adversary to attain the same devices or software and then find out the zero-day vulnerabilities through comprehensive testing. Trojan and virus can be delicately designed and launched to compromise those components.

- **Supply chain:** The vulnerabilities embedded in the supply chain also leave the adversary opportunities to carry out attacks targeting the down streaming applications. Sometimes it is relatively easier to target the vendors than directly intrude into the power system OT network.

The control access and the anomaly detection tools of the vendors are often less secure since they are not directly running the critical infrastructures. For example, backdoors can be injected into the WAMPAC software before they are brought online in the power system.

- **Function design:** Most WAMPAC functions are designed without fully considering the functional robustness or resilience, such that whenever all the defensive strategies fail, it costs the adversary nothing to disable or manipulate the WAMPAC at his will.It is desirable that WAMPAC functions can stay attack-resilient itself to a certain extent and this will serve as the last line of defense.

- **Human:** Last but not least, human employees often turn out to be the most vulnerable link of the security chain [6]. Careless or disgruntled employees might cause great damage either indirectly or intentionally, since they may possess the privilege to access critical data base or WAMPAC configurations and setting, etc.

2.2 Potential Threats and Attacks

Various cyber threats exist in today's smart grid [5]. Cyber threats could potentially arise from individual hackers, industrial spies, terrorists and hacktivists, nation-state affiliated adversary, and even discontented employees [7].

In general, the goal of the attackers is to interrupt the WAMPAC system by sabotaging three things: confidentiality, availability and integrity. For traditional IT sector, confidentiality is the most recognized concern since people worry about private data/info leakage and financial loss because of data theft. But for the operation of critical infrastructures like a power system, availability and integrity are more critical. Most potential threat actors aim at compromising the availability and integrity of WAMPAC systems to degrade the system reliability.

A vast amount of cyber attacks targeting smart grid are happening every day, but most of them do not actually result in damages and are not reported in public. PJM Interconnections former CEO, Terry Boston, once mentioned that "the utility experiences 3,000-4,000 hacking attempts every month [8]". Out of so many attacking trials, a few indeed succeed. On Dec 23, 2015, a cyber attack took place in the Ukraine power grid after a 6-month-long reconnaissance and the attackers successfully disconnected 30 distribution substations for about 3 hours causing significant outages to 80,000 customers [9]. Dec 17th, 2016, another Ukranian transmission substation was disconnected from the grid by a new malware called CRASHOVERRIDE and this caused 200 MW loss of load. A couple of real cyber attacks were seen by other critical infrastructures, and researchers agree that the similar tools can be utilized against power girds too [4]. Stuxnet is a malware developed in 2005 to target Industrial Control System (ICS) SCADA systems and it reprograms ICS PLCs specifically to compromise the system control capabilities. It shows a high level of sophistication and stealth - 4 zero-day exploits are embedded and has becomes widely known after taking down 1/5 of Irans nuclear centrifuges via SIEMENS PLC devices [10]. Stuxnet was distributed through unauthenticated USB drives, while sophisticated virus and malware can also be delivered to the targets via social engineering tactics in order to infiltrate and take over the OT network [9]. In 2017, a malware named as "Triton" was used to shut down a petrochemical plant in Saudi Arabia and the target of the malware is the security instrumented systems Triconex from Schneider [11].

Except for the limited number of real attacks that have been publicly reported, different research efforts help identify the potential risk when critical power system functionality is under malicious cyber attacks [12]. Liu et al. have investigated the impacts of switching attack from the perspective of system stability [13, 14]. Cyber attacks targeting WAMPAC are identified, implemented and analyzed on hardware-in-the-loop testbeds [15, 16, 17, 18] and it was demonstrated how malicious attacks could potentially cause severe reliability and stability impacts through the experimental studies.

For most of the cyber attacks targeting on WAMPAC of smart grids, the attacks can be categorized into three types according to the initial targets in the context of the closed control loop shown in Figure 1.2.

- **Backward path attack,** i.e. measurements attack. The adversary can cause the WAMPAC to malfunction by injecting false measurement data along the sensing path. This can be achieved either locally by tampering the the sensors or manipulating the data along the telecommunication routes with means such as Man-in-The-Middle attack (MITM). This type of attack can be detected and mitigated by countermeasures available in the control center such as centralized anomaly detection, bad data detection adopted in state estimation and other redundant data based cross-checking methods, etc.

- **Forward path attack,** i.e. control signals attack. If the attacker knows that the manipulated measurements are likely to be detected in control center, he might turn to the control signals seeking for better luck. For example,Area Control Error (ACE) transmitted from control center to generating stations can be changed in the WAN by MITM attack too, and not like control center, when a power plant receives an erroneous ACE value, it is much more difficult for it to detect that due to the lack of a global view.

- **Direct control.** Although rarely happens, potentially the adversary does have a capability to control the grid operation at will if he is able to intrude into a control center or a substation and attains the necessary control privilege. Equipment damages will be easily induced like what has been demonstrated by Idaho National Lab with the Aurora attack [19].

2.3 Existing Countermeasures

Except for cyber vulnerability, threat and risk assessment or the system restoration strategy design that serve as tactics to improve the system preparedness when facing cyber attacks [20, 21], three types of countermeasures that people often resort to as efforts to secure the critical smart grid functions like WAMPAC include adoption of conventional IT security solution, best-practice regulation, and attack resilient function design.

Conventional IT security solutions such as communication network segregation, multi-factor authentication, cryptography and IT domain IDS (intrusion detection system) and IPS (intrusion prevention system) are widely adopted as countermeasures to potential cyber attacks [22, 23, 24, 25, 26]. More secure communication protocols are developed [23, 24] for industrial control systems that bring security features such as authentication, packets encryption, etc. Similar to trying to protect an infant from a wolf with the most solid shield, the IT domain security solutions help cover the sensitive data and components of the power grid and make it more difficult for the adversary to obtain access to those assets. Such defensive actions seek to secure the operation of power grid by pinpointing and defending every vulnerability, but unfortunately no system stays flawless. Besides, even a secure periphery is not enough to stop a disgruntled insider or the adversary who possesses tremendous resources and tactics.

Different institutes, such as North American Electric Reliability Corporation (NERC) and National Institute of Science and Technology (NIST), have made best-practice standards as mandatory compliance or recommendations for power system participants to follow [27, 28]. NERC is designated by FERC and has been working on Critical Infrastructure Protection (CIP) standards posed as enforceable regulations for bulk electric systems (BES) [28]. NERC CIP 6 covers the topics as following:

- **CIP-002-5.1a** BES Cyber System Categorization
- **CIP-003-6** Security management controls
- **CIP-004-6** Personnel and training
- **CIP-005-5** Electronic security perimeters
- **CIP-006-6** Physical security of BES cyber systems
- **CIP-007-7** System security management
- **CIP-008-5** Incident reporting and response planning
- **CIP-009-6** Recovery plans for BES cyber systems
- **CIP-010-2** Configuration change management and vulnerability assessments
- **CIP-011-2** Information protection

- **CIP-014-2** Physical security

The "Framework for Improving Critical Infrastructure Cybersecurity" proposed by NIST results from Cybersecurity Enhancement Act of 2014, and it is supposed to be "voluntarily adopted" by critical infrastructure owners [29]. This framework suggested put emphasis on the business drivers which can help the utilities secure their assets and it also advise to incorporate the cybersecurity risk into the risk management process.

Cyber security of the power system can be highly improved but again the regulation cannot cover all possible attack vectors and such that it is not able to help when facing unprecedented zero-day exploits. Recently, the idea to make WAMPAC functions themselves situation-aware and self-adaptive to achieve better cyber resilience has become attractive. Under the dome of this philosophy, the WAMPAC functions do not have to stay uncompromised all the time, but if they have the knowledge to properly deal with known or unknown situations where cyber attack occurs, its resilience will be significantly improved. Exploration in this work is mainly conducted towards this direction.

2.4 Challenges and Opportunities

History has proved that the advent of any technical innovations can become a two-edged sword. This is also true in terms of applications of ICT and big data technologies in smart grid.

2.4.1 Information and Communication Technology

Before power industry observed the advancement of information and communication technology at the beginning of 21st century, most of power plants and substations in the power system utilized analog devices and communication channels to monitor and control the grid operation. The efficiency of power system operation, control, maintenance and restoration is low back to that time and the capabilities of these functions are very limited too. Later on, with the progress in the IT domain, application of assorted ICT technologies becomes popular in the power industry and power systems start to enjoy many benefits like wide-area data acquisition, highly efficient and accurate data processing, and automatic control of the system at much lower cost and evolve towards today's "smart grid". Market, PMU/WAMS, demand response, highly integrated renewable energy control and many other ancillary services are all representatives of successful application of ICT in power systems, and none of them can be realized without the high speed communication and information processing techniques. However, in the early days of applying ICT, the first priority in function design is the functionality and reliability rather than security, which renders the power systems no longer immune to cyber threats, as one of a few undesirable side effects along with all the benefits.

Pros and cons are obvious for the ICT application in power industry, and apparently it is not possible to go back and remove ICT from power systems in order to rid the cyber issues. ICT has become an integral part of the smart grid. To the opposite, we should explore how to better apply the ICT to reduce the cyber issues it brought in the first place.

Though digital intelligent devices and optic fiber have replaced the analog transducers and copper cables, the communication pattern in today's "smart grid" does not differ much from its predecessor. The control center still plays as a master among slaves (substations) in most interconnected power grids. The intelligence and capabilities of the substations are limited, not by the technology but rather by the conservative attitudes to improve. The cost of ICT application keeps decreasing nowadays and the utilities do have the ability to construct their own communication infrastructure. Some utilities are even helping the rural areas with their broadband Internet access [30]. Nothing really stops a substation from being constructed in such a way that it possess high intelligence and autonomy and also the capability of peer-to-peer communication. This will be especially beneficial to improve the attack resilience of smart grid. But this time, cyber security has to be surely encompassed when applying ICT.

2.4.2 Smart Grid and Big Data

Daily operation of smart grids creates a vast amount of data that include equipment operating status, bus voltages, transmission line power flows, and load levels, cyber events logs, etc. These data, at least part of them, is required to be stored for future usage such as post event analysis because of the valuable information is underlying. Recently, only around 2 percent of the data get utilized [31]. How to incorporate these data poses great challenges to the power system designer. A power system needs to engineered in such a way that it is able to deal with big data efficiently including mining the valuable information out of the raw data with appropriate data processing methods. Such capabilities could be leveraged to assist the development of various applications such as visualization and anomaly detection. Big data can be characterized in terms of four "V"s - i.e. volume, variety, velocity and veracity and these four facets should be fully considered when trying to develop a data-driven function.

- **Volume:** Like many other industrial critical infrastructures, power system operates continuously 24-7 and thus the sample measurements easily accumulate. For example, the second-level power consumption data of all homes in New York State within a day could be 127.1 TB [32]. The PMU data generated by US power industry only can be up to 7.5 petabytes per month [31].

- **Variety:** Sophistication of the power system has determined the heterogeneous of the data. Data from sources such as transducers, IEDs, control center and networking devices can be in the form of nominal data, numeric data, log texts, and even video/audio files. Reference [33] has provided a list of the possible data sources in smart grid.

- **Velocity:** The sampled data such as SCADA measurements and phasor measurements are fed into the control center as data streams. Specifically for the phasor measurements, its update speed will be quite fast. The sampling rate could be up to 128 times per second.

- **Veracity:** Due to the flaws of sensing devices and the data transmission paths, the measurements could of bad quality. Therefore, bad data should also be treated as one type of anomaly.

The nature of power grid operation data has made the data storage and processing very challenging, but data also contains valuable information which can significantly help with equipment status monitoring, system overarching state detection and control decision making, given that the information can be efficiently mined out of the raw data. Many efforts have been dedicated to the design and application of data-driven methodologies in power systems. Machine learning techniques have demonstrated great competence in extracting information from large amount of raw data. Some general techniques summarized as below in general can help better leverage the bid data available in smart grid in an efficient way.

- **Classification methods:** When the number of data instances obtained is reasonably small, they could be labelled by experts indicating the class that they belong to. Then, supervised learning based classification can be utilized. After the training, a “model can be derived and it judges what class of a given data instance is. Many different classification methodologies exist for this purpose. For instance, Decision Tree (DT) delivers the detection model as a tree which embeds all decision-making procedures and the model that a Support Vector Machine (SVM) attains is a hyperplane specified by a group of optimized parameters. In the detection stage, a new data instance will be classified according to its location in terms of the hyperplane. Ensemble methods leveraging multiple simple classifiers, such as random forest and boosting, can often get better overall performance.

- **Clustering:** When the amount of data instances is too large to label, semi-supervised or unsupervised machine learning techniques can be applied. K-means clustering and hierarchical clustering are distance based methods that are proper for scenarios with evenly distributed data instances, and Density based methods such as Density Based Spatial Clustering for Application with Noise (DBSCAN) performs better when the clusters possess different densities.

- **Data processing:** Another key technique in application of machine learning is the dimensionality reduction. For data set with many features, dimensionality reduction is necessary in order to improve the computation efficiency and visualize the data. The selection of dimensionality reduction technique depends on the concrete problems. For some problems, linear techniques such as Principal Component Analysis (PCA) performs well but for the others, if the instances are concentrated on the nonlinear manifolds in the original hyperspace, nonlinear dimensionality reduction algorithms such as t-SNE is a better option. In some special scenarios, if the relationship among features are well known with domain knowledge, people should try to reduce the dimension by applying their expertise first which could often obtain better results in feature extraction.

All data mining and machine learning techniques attempt to discover certain underlying rules or patterns from available data. It is well known that model obtained via this type of techniques could only be as good as the data, and this reveals one of the salient shortcomings 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, because in most of these cases, to derive an accurate model-based solution is too difficult, if not impossible at all. Besides, the data-driven methods can automate and speed up the induction process, and they provide much better performance in terms of efficiency and efficacy compared to human experts. In the meanwhile, the application of data mining as assistance to automatic control systems such as WAMPAC often results in data processing overhead. Therefore, the selection of methodologies should be carried out based on the characteristics and needs of each concrete problem.

2.5 Enhance WAMPAC Resilience from New Perspective

Considering the vulnerabilities and threats existing in the cyber layer of a smart grid and also the mature technical progresses observed in the sectors like ICT and data science, this work selects two critical WAMAPC functions, i.e. wide-area protection and automatic generation control, and investigates the potential solutions to enhance the cyber security and resilience of the two functions based on functional decentralization and machine learning.

As detailed in Figure 2.1, this work consists of two parts. Part 1 focuses on the study of System Integrity Protection (SIP) schemes, specifically explores the function decentralization based on

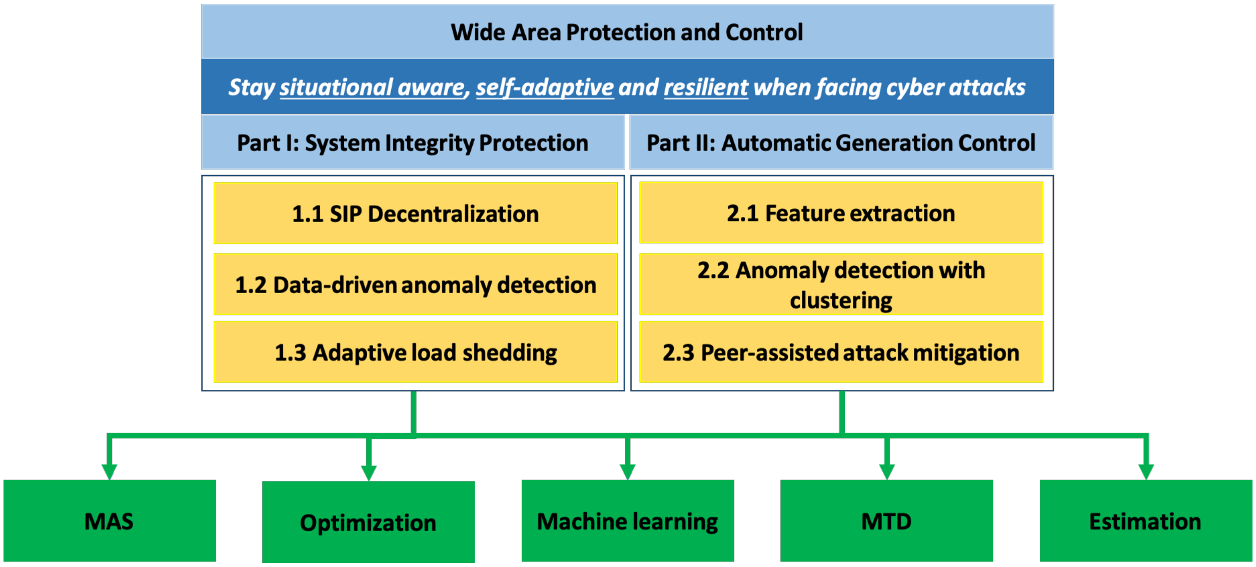

Wide Area Protection and Control
Stay situational aware, self-adaptive and resilient when facing cyber attacks
Part I: System Integrity Protection
Part II: Automatic Generation Control
1.1 SIP Decentralization
1.2 Data-driven anomaly detection
1.3 Adaptive load shedding
2.1 Feature extraction
2.2 Anomaly detection with clustering
2.3 Peer-assisted attack mitigation
MAS
Optimization
Machine learning
MTD
Estimation

Figure 2.1 Holistic dissertation structure

Multi-Agent System (MAS), anomaly detection with data-driven methods and self-adaptive load control as means to make the SIP in smart grid more resilient when facing Denial of Service attack. Data integrity attack has not been considered in this part. Part II examines how peer-to-peer communication among the generating stations can help with detection and mitigation of attacks targeting the power system generation control values. Power industry domain knowledge is utilized in the feature engineering, and clustering methods are adopted to detect attack induced anomalies. Besides, a peer-assisted mitigation is proposed when the forward control path of AGC gets compromised. Both availability and integrity issues along the forward path can be mitigated, and this work does not cover cyber attacks targeting AGC measurements. Although methodologies and algorithms proposed in this work are functionally specific, the philosophy underlying can be applied to other WAMPAC functions. The proposed solutions involve the application of MAS, optimization, machine learning, moving target defense (MTD) and also data estimation, which will be further elaborated in chapter III and IV.

2.6 Summary

This chapter has provided an overview about the potential cyber vulnerabilities and threats of WAMPAC functions in smart grid. The advancement of technologies including ICT and big data provides great opportunity to upgrade the WAMPAC so that the functions per se can become situational-aware and adaptive and thus resilient when under cyber attacks.

CHAPTER 3. MAS BASED ATTACK-RESILIENT SIPS

3.1 Problem Statement

One of the most critical WAMPAC functions in nowadays smart grids is System Integrity Protection Schemes (SIPS) [34]. Unlike the local equipment protection schemes, the responsibility of SIPS is not to protect any single piece of equipment such as a generator, transmission line or a bus bar. Instead, it maintains the stability of the whole system. SIP schemes collect system-wide data and information via telecommunication networks and need to carry out proper protective actions when facing predefined system conditions so that severe system problems such as catastrophic failures can be prevented [35, 36, 37, 38]. The most common controls that SIP schemes could perform include load rejection, generation rejection, etc.[39, 40] and the wide application of Information and Communication Technologies (ICT) in smart grids has significantly augmented the capabilities and improved the efficacy of SIPS. Most of the SIP schemes today are installed as a centralized master-slave system, i.e. the system-wide information is collected by the master where the decision is made about current system condition. If the predefined conditions such as the disconnection of a large generator is detected by the master, it will send control commands to multiple slave stations. The role of slaves is limited to sensing and sending system measurements to the master and then following any control commands received from the master. The cyclic data collection and control actuation of SIPS heavily relies on wide area communication where abundant cyber threats exist. Concerns with conventional centralized SIPS include

1. The centralized master is an ideal target and when it is compromised, a single point failure will render the protection function completely inoperable [41].

2. Static protection settings are commonly utilized and the remedial actions might not be suitable when system operating state changes, either naturally or due to cyber attacks [42].

To enhance the cyber attack-resilience of SIPS, and also to improve the economic performance of SIPS, a novel decentralized SIPS design is proposed in this chapter based on Multi-Agent System (MAS). In the context of a decentralized SIP scheme, cyber anomaly detection and self-adaptive control are investigated as the two main subproblems to ensure that the SIPS solution proposed possesses high level of situational-awareness and adaptiveness against cyber attacks. Assumptions made in this part are listed below.

- Substations involved in the SIPS can be upgraded into intelligent agents.
- MAS is utilized on station-level, i.e., an agent represents a whole substation.
- Peer-to-peer communication is enabled among agents and each agent only needs to exchange data with its directly connected neighbors.
- Only Denial of Service (DoS) attacks are considered. That is, it is assumed that the cyber attack only makes an agent fail to communicate with others, but not send erroneous data. Data integrity attacks can be mitigated by techniques such as cryptography.
- The legacy centralized SIPS is kept as part of the decentralized SIPS.

3.2 Related Work

This section first summarizes the efforts that have been taken to enhance the cybersecurity of SIPS, and then overviews the application of MAS, data-driven algorithms and adaptive control schemes in smart grid respectively.

A rule-based intrusion detection system with Multi-Agent System (MAS) is proposed in [41] to distinguish cyber attacks and normal faults and this study focuses on local protections. Distributed Special Protection Systems (SPS) based on agents and peer-to-peer communication is advocated in [43]. The distributed protection systems presented in that work leverage “reputation-based trust” to identify untrustworthy agents statistically and “data retransmission” mechanism to reduce the impacts of data loss. However, the proposed solution requires large amount of redundant data being simultaneously fed into a few on-line controllers. Although the redundancy of measurements and controllers helps detect and mitigate Byzantine failures, countermeasures such as cryptography could be comparatively more effective and economic to achieve the same goal.

Multi-Agent System (MAS) is an appropriate architecture that supports decentralization. McArthur et al. have conducted a survey about the application of MAS in power system [44, 45], according to which the MAS has been applied to realize back-up protection, system condition monitoring, market and micro-grid control, etc. The authors propose two MAS organizations for dynamic voltage control in [46]. The team organization is statically assigned to one MAS and is not as flexible as the holarchy organization since it allows the MAS to exchange agents dynamically according to the current system operating condition. In [47], the authors propose a protection scheme to prevent cascading events based on MAS such that the status of system can be monitored in real time and the emergency control can be optimized. [48] proposes a MAS that can evolve during attack and X. Tong et al. studies the application of MAS to realize wide area backup protection [49]. Reference [50] has proposed a layered MAS based intrusion/anomaly detection system aligning with IEC 62351. It provides good division in terms of the detection at different layers, i.e. node layer (local) or system layer (system wide). C. Rieger et al. present a hierarchical MAS architecture in [51, 52] which could naturally fit into power system architecture. Paper [53] also proposes to use hierarchical MAS to carry out anomaly detection for wide area protections. The MAS architectures and agents characterization found in all the aforementioned work are inspiring, but not many explorations are conducted on the design of cyber-attack resilient functions using MAS. We can safely conclude that MAS is an attractive and promising solution that power researchers resort to in order to resolve various problems by leveraging its salient features including self-adaptive, autonomous, flexible, etc. But not much efforts are spent trying to answer how the decentralized architecture can in fact help improve the attack resiliency and what cyber concerns exist in a MAS itself.

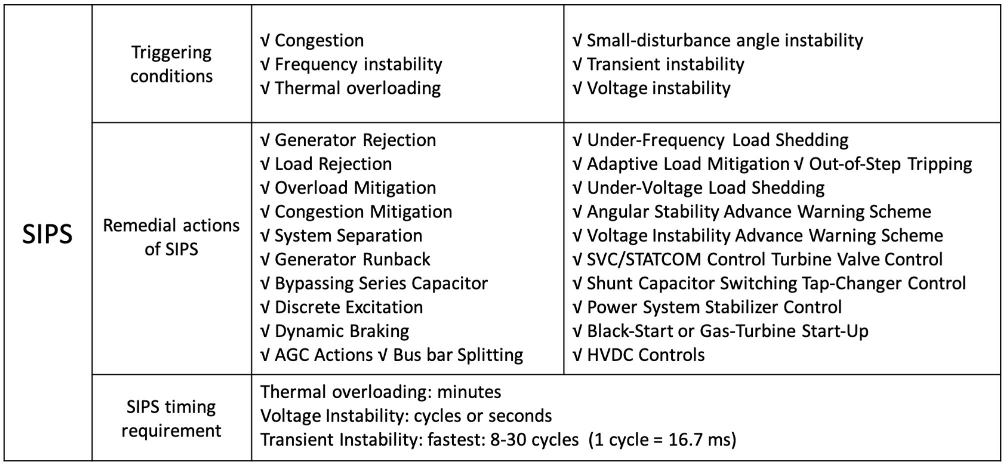
Huge amount of data are generated in the daily operation of power system [33], both from the physical layer and the cyber layer. It is extremely hard for human operators to effectively find the valuable information underlying. Recently, machine learning techniques and artificial intelligence have demonstrated outstanding capabilities in handling big data automatically, and many researchers are trying to leverage it to resolve the problem of data mining in smart grid. Applications in power systems of machine learning include transient instability detection [54, 55, 56, 57], bad data detection and anomaly detection [58, 59, 60], etc. Paper [58] comes up with a state based intrusion detection algorithm called common path mining. The algorithm leverages merged data sources and detect attacks if the common path deviates from the possible normal paths. [55] proposes a PMU data anomaly detection methodology based on unsupervised ensemble machine learning. [59] proposes a multiple event detection based on moving window PCA. A PCA & SVM based false data injection detection method is proposed in [61].

Unlike the common practice that massive off-line events analyses are carried out as reference of on-line applications, a decision support tool, Multi-layer Data-driven Advanced Reasoning Tool (M-DART), is presented in [62]. It handles various data sources in a dynamic way such that the massive data can be processed in time to improve the situational awareness. Most of the data-driven applications proposed are centralized implementation and needs the access to global data. How to apply the data-driven algorithms in a decentralized architecture mainly with locally accessible data is worth investigating. The redesign of critical power system functions could be the key to make the system more attack resilient.

Dynamic schemes used for load shedding are investigated in [63, 64]. Similar idea is adopted in the proposed solution. This will grant the MAS the ability to determine the optimal loads that shall be shed when the triggering condition is observed. Besides, an DoS attack adaptive version is proposed so that the SIPS can deliver the remedial actions to the best effort even under attack.

3.3 Overview of SIPS Application

Figure 3.1 summarizes the common SIPS utilized in smart grids, including the trigger conditions, remedial actions and timing requirements [35]. Among all possible remedial actions, the load rejection and generation rejection are the two mostly adopted to handle the stability issues of



SIPS	Triggering conditions	√ Congestion √ Frequency instability √ Thermal overloading	√ Small-disturbance angle instability √ Transient instability √ Voltage instability
	Remedial actions of SIPS	√ Generator Rejection √ Load Rejection √ Overload Mitigation √ Congestion Mitigation √ System Separation √ Generator Runback √ Bypassing Series Capacitor √ Discrete Excitation √ Dynamic Braking √ AGC Actions √ Bus bar Splitting	√ Under-Frequency Load Shedding √ Adaptive Load Mitigation √ Out-of-Step Tripping √ Under-Voltage Load Shedding √ Angular Stability Advance Warning Scheme √ Voltage Instability Advance Warning Scheme √ SVC/STATCOM Control Turbine Valve Control √ Shunt Capacitor Switching Tap-Changer Control √ Power System Stabilizer Control √ Black-Start or Gas-Turbine Start-Up √ HVDC Controls
	SIPS timing requirement	Thermal overloading: minutes Voltage Instability: cycles or seconds Transient Instability: fastest: 8-30 cycles (1 cycle = 16.7 ms)	

Figure 3.1 Power industry SIPS summary

power grids. The timing performance of SIPS also varies depending on purposes of the protection. Schemes equipped to alleviate transmission line thermal overloading conditions are often allowed up to tens of minutes.

3.4 Decentralized SIPS based on MAS

Most conventional SIP schemes comprise a master and a few geographically dispersed slaves. SIPS master is able to collect data, detect targeting events, and then determine and send the control commands to the slaves. On the other hand, slaves are only required to follow the commands from the master and correctly carry out the remedial actions. In such highly centralized schemes, the master is prone to availability attacks such as Denial of Service (DoS) and intrusion attempts like spear phishing attack. Loss of the master will either lead to hidden protection failures or immediate system impacts when under a coordinated attack. In this section, a peer-to-peer MAS architecture is proposed as an alternative for SIPS implementation.

3.4.1 SIPS Implementation within Decentralized Architecture

An agent is an object possessing a specific set of resources (processors, memories, sensors, actuators, etc.) and behaviors (such as informing, requesting, offering, accepting, rejecting, competing with or assisting one another). Multi-agent System (MAS) is a group of agents that communicate and interact with each other to realize various tasks with certain level of autonomy.

Typical MAS architectures include hierarchical [46, 51], peer-to-peer [49], etc. Considering that hierarchical architectures still discriminate between agent roles, which means that the agents of more significance are more likely get attacked and other agents are not able to replace it if so, the peer-to-peer architecture is adopted in this work. The goal is to ensure that when any part of the MAS gets compromised, the rest agents are still able to provide the expected protective actions.

As depicted in Figure 3.2, the first principle in the proposed MAS design is that an agent represents one substation or generating plant. MAS can also be utilized at a more granular level, within one station for instance [51], but this type of MAS is out of our discussion. Communication among agents are critical. One agent needs to talk to its neighbors to realize global data propagation and to exchange specific situation information to achieve a consensus of the current system state, etc. Thus the communication architecture design can have a significant influence on the performance of MAS. Figure 3.2 shows one possible communication topology, which can be exactly mapped to the physical layer. It's a reasonable choice considering the availability of power line carrier and also that the needs of information propagation is usually constrained to a group of stations close to each other. A peer agent far away from the one needs assistance under stressed condition is not likely to help as much and neither does it need to know about the stressed condition although the relevant information can still be propagated via proper routing methodologies. In the later work, we only care about the de facto communication connection status and do not put much emphasis on the topology design itself.

3.4.2 MAS Characterization

As aforementioned, resources and behaviors are the two major facets of an agent. This section will first discuss the data sources that accessible to agents from two different perspectives and then describe the overall MAS operation.

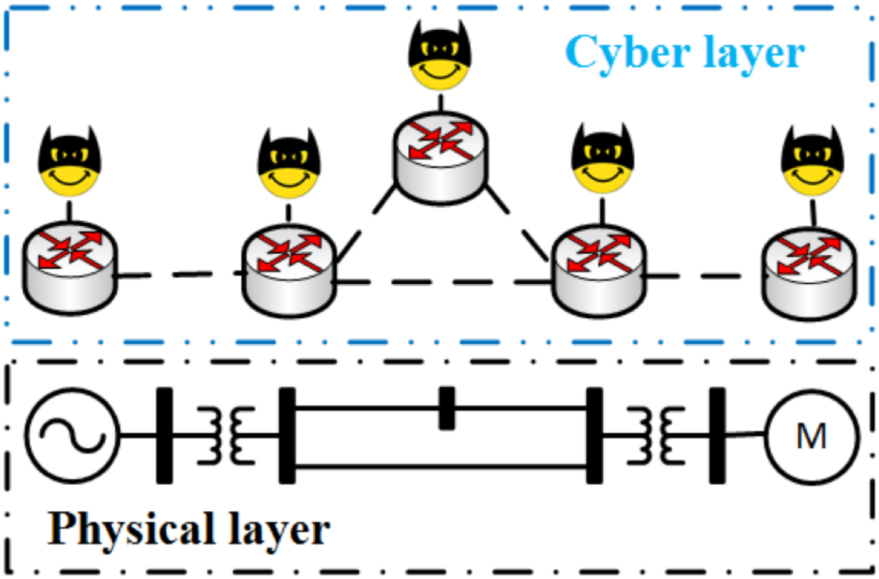


Figure 3.2 Peer-to-peer MAS architecture

3.4.2.1 Data Sources

Locally within one substation, the data accessible to an agent include physical layer measurements, cyber information and processed metrics.

- **Physical measurements** - breaker/switch status, bus/generator terminal voltage, power injection, power flows, and system frequency, etc.

- **Cyber information** - Information from cyber layer such as absence of expected measurements/commands, unexpected packets, incorrect control sequential numbers, anomalous events recorded by log files and indications from security management systems, etc.

- **Preprocessed metrics** - agents can preprocess raw data and compute information enriched metrics to improve the efficacy of online applications. For example, the time interval between two sequential frequency dips may reveal the tendency of cascading events [65].

From the system wide perspective, the data that an agent possesses or accesses could be classified as

- **Public local information** - Information that could be shared with other ally agents when on request.

- **Private local information** - Sensitive data of an agent that could not be shared, such as the load data of a critical feeder that this agent can not lose.

- **Global information** - Other agents’ public local information that have been successfully propagated and collected.

Though barriers still exist to merge all the above data/information at the same module, it is technically achievable. Agents could either obtain such data from sensing devices like what an RTU does or directly get data from RTU.

3.4.2.2 Agent Actions

The coordination among agents highly relies on the information propagation protocol utilized. One possible protocol has been proposed and it only requires one agent to communicate with its directly connected neighbors. If two agents are not directly connected, they have to share the data via other intermediate agents. The main actions of an agent involve

- **Inform** - One agent actively informs its neighbors about the data sets it possesses. An “inform” message contains the data set(s) that just get updated and also a time stamp. For instance, an ”inform” message could read as ”obtain new data from substation i at time 16:30:00”.

- **Confirm** - One agent should reply to the neighbors who have sent it global information.

- **Request** - One agent will send out data request to its neighbors to get the data in need.

- **Send** - One agent sends the requested public local data back to the agent requests for it.

- **Receive** - Upon the arrival of any packet, the agent will read it in and make a state transition.

The overall behaviors of a single agent can be described by a Finite State Machine (FSM) as shown in Figure 3.3. After the “Initiated” state, the agent will first send out “Inform” packets to its neighbors and then automatically enter the “Receive” state. State transition takes place according to the type of packets that an agent has received. Numbers marked on the outbound arrows of a state correspond to the packet type labels inside the same state. For example, when in the state “Receive”, the agent will go to state “Send Confirm” following arrow 1 when it receives an “Inform” packet.

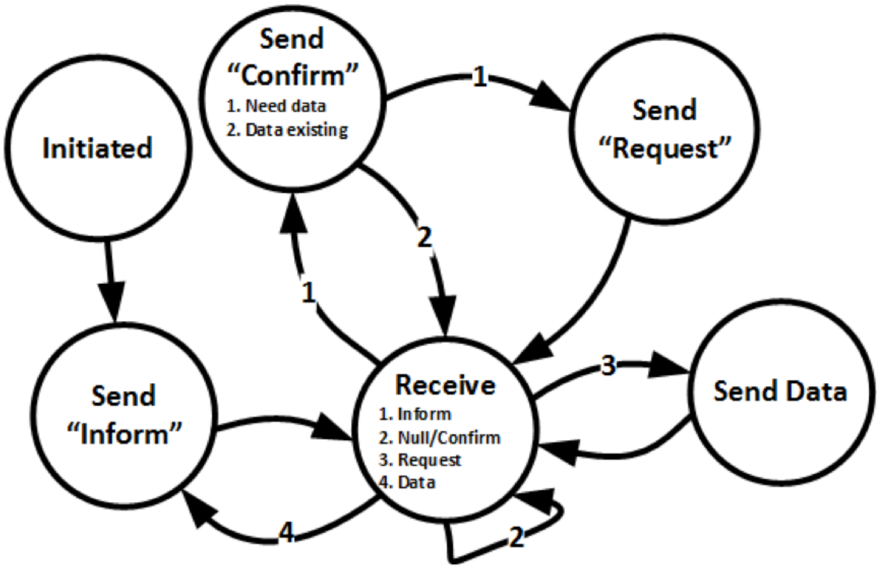


Figure 3.3 Agent actions as FSM

Figure 3.4 shows an example of data propagation among agents. The timing flow starts when Agent 1 has obtained a new dataset. Then Agent1 will send ”inform” message to all its neighbors except the one from whom it gets the data at the first place. In this case, Agent 2 and Agent 3 are the message receivers. As depicted, Agent 3 receives the ”inform” message without an issue, but it finds that it already has the same data set with the same time stamp. Therefore, it only replies with a ”Confirm” message. We assume that the connection between Agent 1 and Agent 2 is temporarily blocked at the beginning and thus the ”inform” message takes several trials to get to Agent 2. Agent 2 checks his database and notices that it does not have the same data set or the one it has is outdated, thus it will send ”request” to Agent 1 following the ”Confirmation”. Agent 1 will send the requested data back to Agent 2.

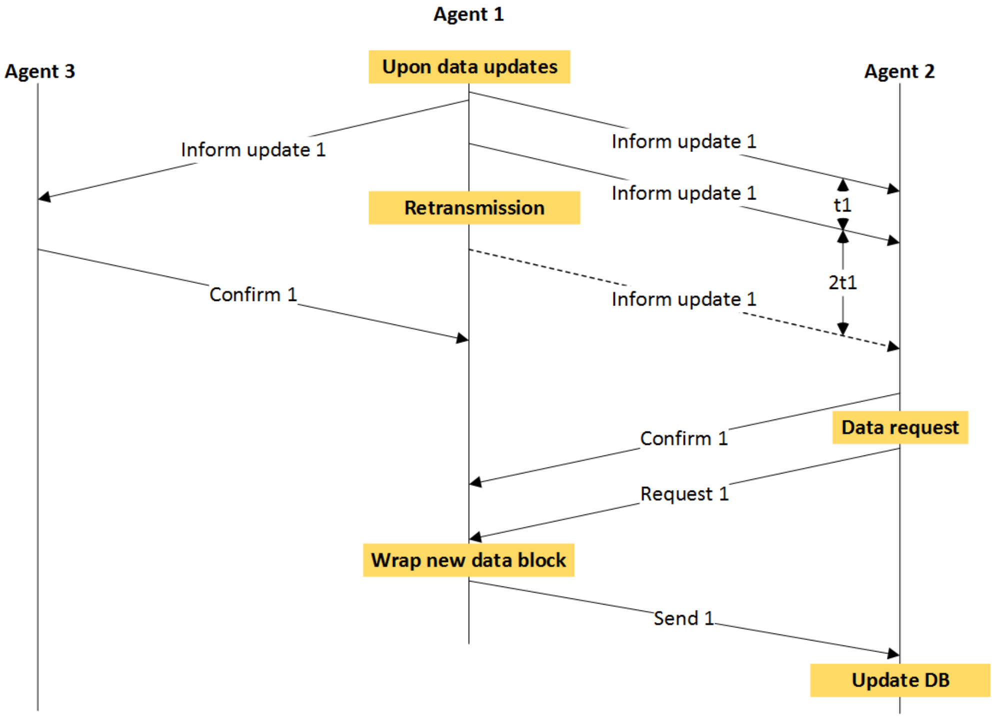


Figure 3.4 Data propagation timing-flow among agents

3.4.3 Overall SIPS Design

The overall MAS operation flowchart is shown as Figure 3.5 with anomaly detection and optimal load shedding as embedded functions. Anomaly detection is carried out locally by each agent but a situation consensus has to be achieved by all the agents that are still interconnected before taking any protective actions. Since the interconnected agents share predetermined data set with each other regularly, after the consensus is obtained, all interconnected agents will form the same optimization problem for load shedding. The remedial actions that one agent should carry out will be attained after it solves the optimal problem. More details about the anomaly detection and adaptive load shedding will be discussed in the following sections.

A most noteworthy issue about any MAS design would be how to make the communication and coordination among agents efficient. Timing requirement differs from application to application, and it needs to get fully considered during the system design phase to make sure that the remedial strategy can fit into the time window that it is allowed.

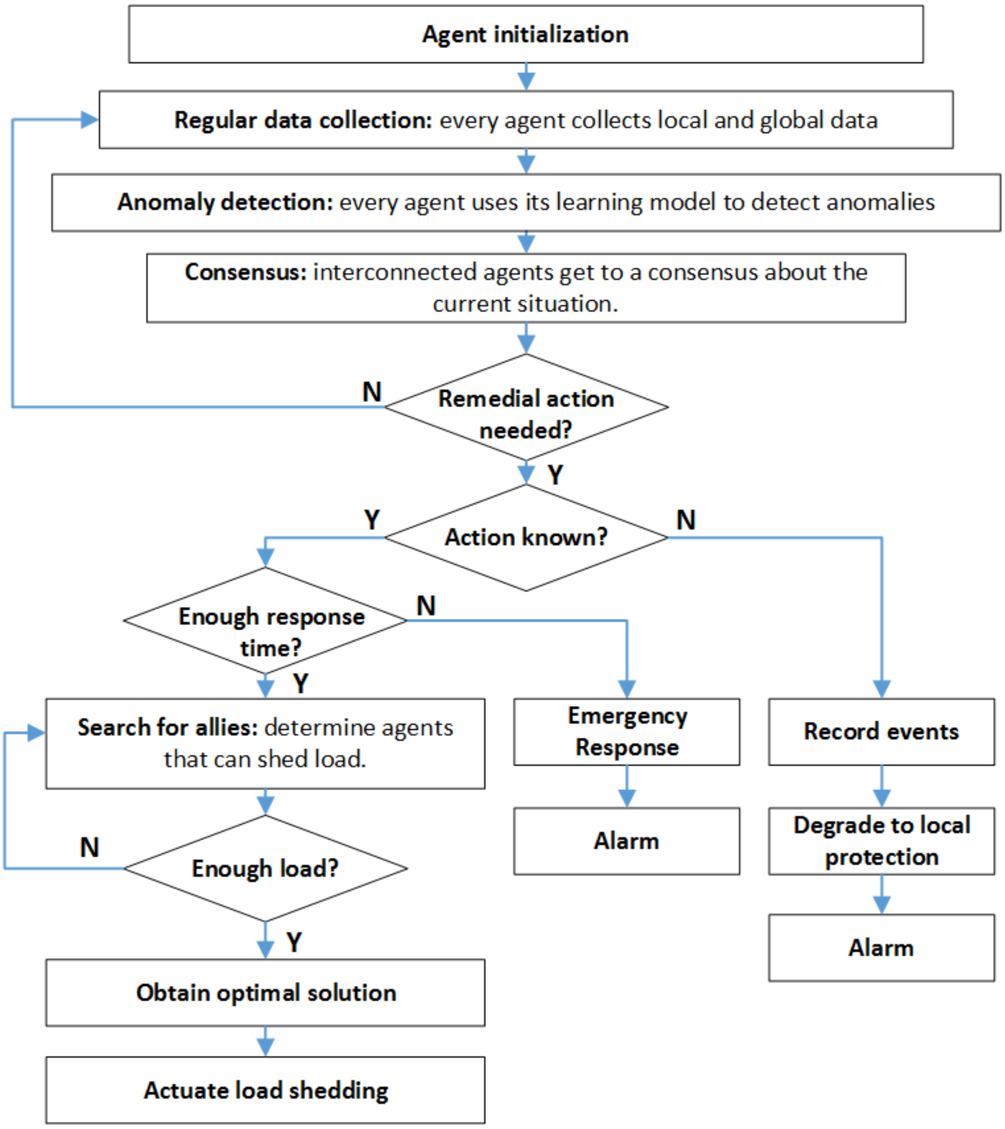

Agent initialization
Regular data collection: every agent collects local and global data
Anomaly detection: every agent uses its learning model to detect anomalies
Consensus: interconnected agents get to a consensus about the current situation.
Remedial action needed?
N
Y
Action known?
Y
N
Enough response time?
N
Y
Search for allies: determine agents that can shed load.
Emergency Response
Record events
Alarm
Degrade to local protection
N
Enough load?
Y
Alarm
Obtain optimal solution
Actuate load shedding

Figure 3.5 Agents overall operation flowchart

3.5 Data-Driven Anomaly Detection of SIPS

A simple cyber security question one can have for centralized SIPS is that when the slave agent does not receive any commands from the master controller, is it because no emergency occurs or the commands are actually blocked by the adversary? There is no good way for centralized SIPS to distinguish the two scenarios. For decentralized SIPS to handle such issues, a data-driven anomaly detection algorithm named as Support Vector Machine embedded Layered Decision Tree (SVMLDT) is suggested in this section and the anomaly detection only requires the local cyber physical measurements. In order to label data instances during training, the CPS operations states are defined qualitatively first.

3.5.1 Cyber Physical System Operating States Categorization

The operation states of a power system are conventionally classified as normal, alert, emergency, in-extremis and restorative, according to whether the equality and inequality operation constraints are satisfied or not [66]. Considering the increasing interaction between physical and cyber systems, we have come up with a more comprehensive state transition diagram for a CPS as shown in Figure 3.6. Blue texts in a state block represents the cyber system state and red represents physical system state.

- **Normal state (N/0)** - Both the physical and the cyber systems are free of anomaly, this state is represent by either letter "N" or number 0. Similar idea applies to other states.
- **Post contingency state (P/1)** - When a physical contingency happens and the protection scheme is free of attack and works as expected.
- **Hidden failure state (H/2)** - Part of the cyber system has been compromised but no physical emergencies are observed yet.
- **Alert state (A/3)** - A protected physical contingency happens while the protection scheme is compromised.

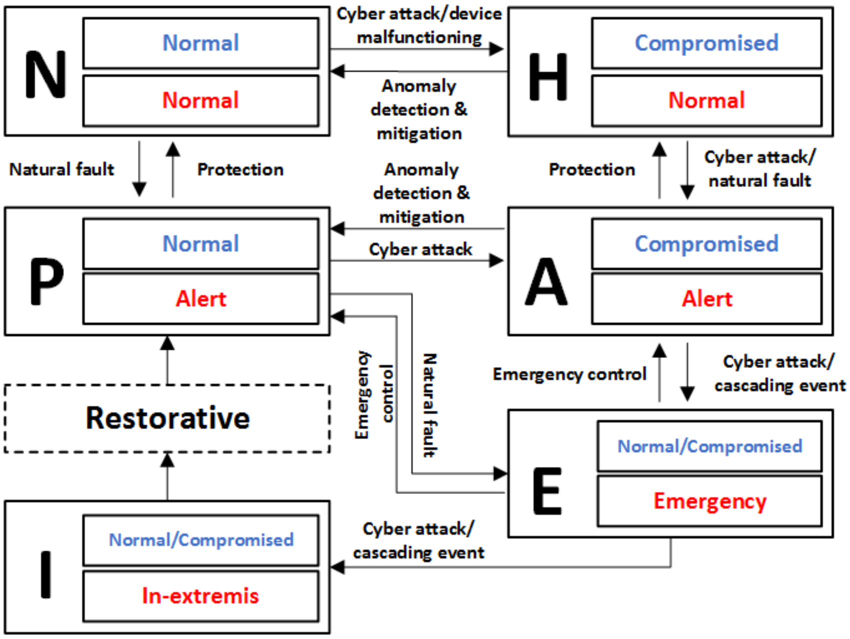


Figure 3.6 Cyber physcial system states

- **Emergency state (E/4)** - The physical system is in a position where cascading events can be triggered due to the malfunctioning of protection scheme.
- **In-extremis state (I/5)** - Power system becomes unstable where isolation is required. This state will not be considered in anomaly detection.

Smart grid operates in the normal state most of the time, but it may enter abnormal states after natural disturbances or malicious cyber attacks. The system enters restorative state to recover from in-extremis state. In the restorative state, we assume that both the physical and cyber issues get resolved. Some of the cyber capability of an agent in fact could have intermediate state between "normal" and "compromised". For instance, the communication bandwidth can be only partly taken by DoS attack which results in certain delays. In this situation, the agent might still be able to carry out the required data exchange with others. But to simplify the problem, these partially compromised states are ignored.

From the perspective of anomaly detection, each CPS state can be perceived as a specific cluster locating in the hyperspace. Noticeable distances may or may not exist for any two clusters depending on the selected feature profile. The anomaly detection can be formed as a multi-class classification problem. Ideally, the anomaly detection needs to distinguish the normal states (N/0) from the abnormal states that are either induced by cyber attack (H/2, A/3 or E/4) or natural contingency (P/1). All the possible states except the in-extremis state will be further utilized for data labeling and the goal of on-line application of anomaly detection is to find the Alert and Emergency states where load shedding is required.

The performance requirements for anomaly detection in power system include

1. In the multi-class classification, the basic requirement is that the anomaly detection module should be able to correctly detect compromised states (H/2, A/3 and E/4).

2. It is preferable to shorten the time needed for the data processing and detection.

3.5.2 SVM Embedded Layered Decision Tree

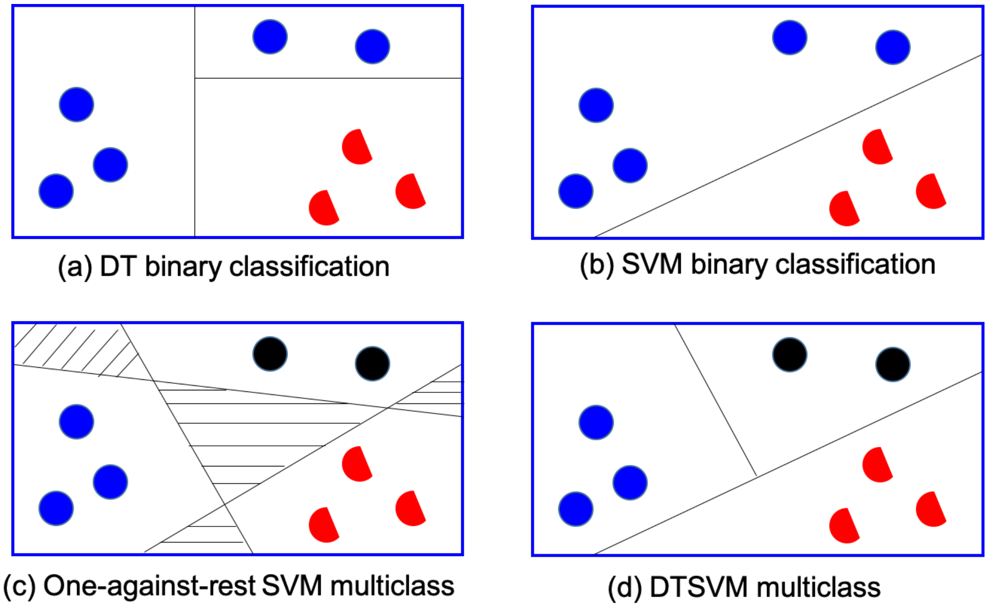
In [67, 68], researchers have proposed a decision tree based support vector machine (DTSVM) for multi-class classification problem. It resolves the problem of unclassifiable region in the traditional application of SVM in multi-class classification. This method focuses on all-continuous-feature problem, but in power system, the feature space is a mixture of many continuous variables and a similar amount of, if not much more, nominal variables. An improved version of DTSVM named Support Vector Machine embedded Layered Decision Tree (SVMLDT) is proposed to solve the multi-class classification in a mixed nominal continuous feature space.

DT stratifies the hyperspace one split a time with a specific feature and a specific feature value. As shown in Figure 3.7(a), a split of a DT forms a linear hyperplane, whose normal is the axis of the feature selected as the the decision condition at this split. Compared to numeric features, this type of splitting works better with nominal features in terms of accuracy, although the most advanced decision tree solution (for example, C5.0) can also classify in mixed feature space by discretizing the numeric features [69]. As its name implies, the stratification process of DT can be visually modeled by a top-down tree, whose internal nodes represent splits of the hyperspace and the leaf nodes represent the separated subspaces. Every leaf node of a DT has a class label. The classification of a new data sample is achieved by walking along the DT from the root node according to the decision conditions of internal nodes until a leaf is hit. This process can simply be described by a sequential of "if else " statements and thus DT is famous for its interpretability and classification speed. The complexity of building a decision tree arises with feature selection, over-grown tree pruning, tree ensemble methodologies, etc., and mature techniques have already been developed [69].

Support Vector Machine (SVM) solves the classification problem by directly looking for an optimal non-linear hyperplane that maximizes the margin between the samples from two different classes in a numeric feature space by solving the optimal problem as shown in (3.1). Application of non-linear kernels allows SVM to find a non-linear hyperplane [70]. Therefore, SVM has more flexibility in terms of hyperspace division compared to DT (see Figure 3.7(b)). However, SVM is ideal for binary classification problems. When it comes to multi-class classification, SVM can be applied recursively with either one-against-one or one-against-rest strategy [70], which may leave undefined regions [67] (shadowed regions in Figure 3.7(c)). In [67], the authors have proposed DTSVM for multi-class classification. It successfully resolves the issue of unclassifiable regions as shown in Figure 3.7 (d). However, DTSVM does not try to distinguish the numeric features and the nominal ones during the learning process.

$$\begin{aligned} \min_{\omega,b} \quad & \frac{1}{2} \parallel \omega \parallel^2 + C\sum_{i=1}^{N} \xi_i \\ \text{s.t.} \quad & \xi_i \geq 0, \\ & y_i(\omega^\top \Phi(X_i) + b) \geq 1 - \xi_i, \; i = 1, \ldots, \text{N}. \end{aligned} \tag{3.1}$$

Considering the nature of CPS anomaly detection and pros and cons of DT and SVM, SVMLDT is proposed so that it leverages the evaluation speed of decision tree and the accuracy of SVM. First all the nominal features will be utilized to divide the hyperspace according to a decision tree, i.e., Layer I tree. Second, for each leaf node in Layer I tree, DTSVM will be applied for further division


(a) DT binary classification

(b) SVM binary classification

(c) One-against-rest SVM multiclass

(d) DTSVM multiclass

Figure 3.7 Algorithms comparison

and render multiple Layer II trees. The final model obtained by SVMLDT looks like the one as shown in Figure 3.8 In Layer I tree formulation, information gain (3.3) based on cluster entropy (3.2) is used to determine the space division sequence while for Lyaer II trees formulation, the type I DTSVM in [67] applied. More details are provide in Algorithm 1.

$$Entropy = -\sum_{i=1}^{N}[p(C_i)logp(C_i) + q(C_i)logq(C_i)] \tag{3.2}$$

$$Gain(S, A) = Entropy(S) - \sum_{v \in value(A)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{3.3}$$

The overall SVMLDT training process is summarized as Algorithm 2. The process to form layer 2 trees mainly include following procedures

- Step 1: Call selectClass() to determine on one specific class C from all the class levels. This algorithm will find a class whose centroid is "most away" from other classes;
- Step 2: Form a binary classification problem by modifying all other classes label;

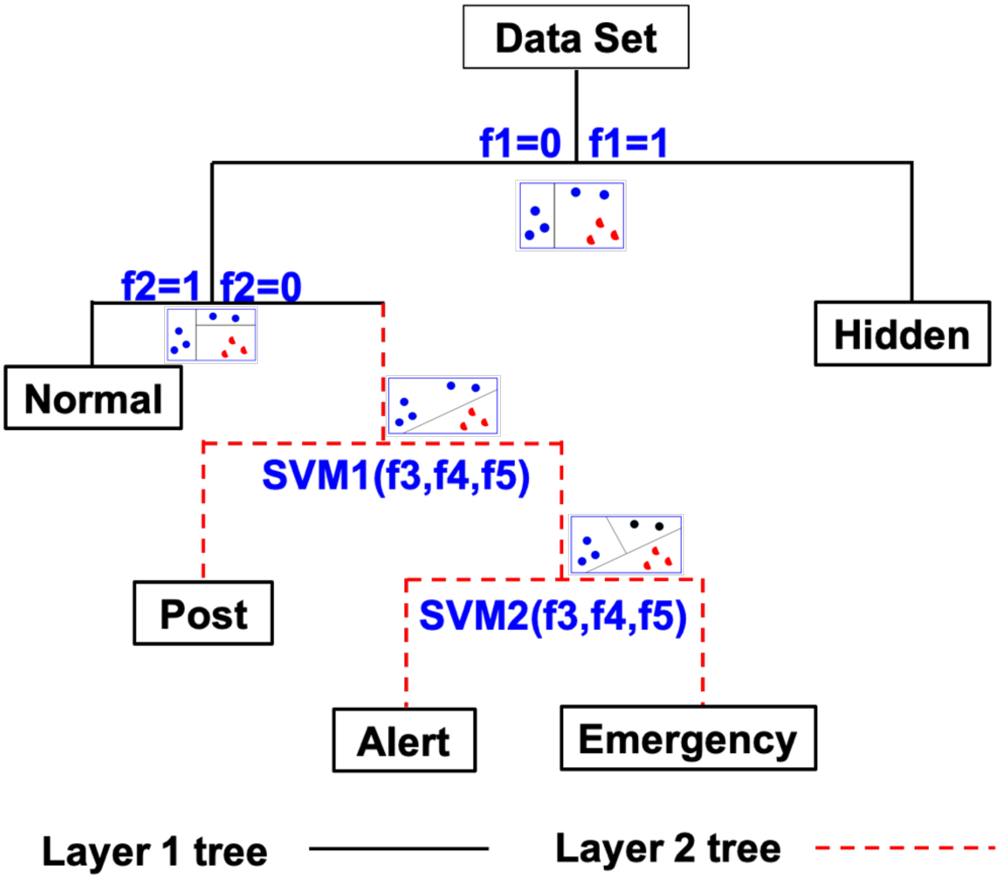


Figure 3.8 SVMLDT model

- Step 3: Remove all samples labeled as C;

- Step 4: Repeat Step 1 to 3 until sample data set becomes empty.

In fact, either SVM, DT or DTSVM can still provide an solution to our problem, but they all have some disadvantages in theory. Table.3.1 shows the pros and cons of the 4 relevant algorithms.

The substation-level anomaly detection serves as the last line of defense by validating the control commands received from the legacy centralized master module. This can be illustrated by the two examples below.

- **Replay attack when load shedding is not needed.** Assuming that the attacker has the capability to inject and send fake triggering commands to the substation agents, the anomaly

Table 3.1 Multi-class classification algorithms comparison

	Operation	Pros	Cons
DT	Discretize continuous features	speed	accuracy in non-linear boundary
SVM	1. Treat nominal features as continuous 2. One-against-one or one-against-rest	non-linear boundary accuracy	1. speed 2. non-classifiable regions
DTSVM	1. Treat nominal features as continuous 2. Utilize SVM as nodes in DT	1. non-linear boundary accuracy 2. no non-classifiable regions	speed
SVMLDT	1. Use nominal features to form layer 1 tree 2. Apply DTSVM for every impure leaf node of layer1 tree	1. leverage the speed of DT 2. non-linear boundary accuracy 3. no non-classifiable regions	speed is not as good as DT

detection is expected to detect it. In fact, when a substation agent cross-checks all the data sources it has, especially the physical measurements like system frequency and line flows, it will notice that the load shedding triggering commands come from nowhere and the current system state should be classified as “Hidden failure” state where no load shedding should be performed.

- **DoS attack when load shedding is needed.** To the opposite of the first example, this case represents a scenario where the physical contingency has taken place, however the load shedding commands from the centralized node get blocked by the adversary. In this case, the agents are also able to detect the anomaly since the physical system symptoms would reveal the needs for remedial actions. As long as most of the interconnected agents reach an agreement about the current situation, i.e. a consensus, the load shedding will be triggered even without commands from the central module.

```
Data: dataset $S$
Result: class $C_{sep}$
for each class $C_i$ do
    find centroid distances $\{d_{ij}, j \neq i\}$ between $C_i$ and other classes
end
$C_{sep} = arg\underset{i}{M}ax\underset{j}{M}in(d_{ij})$;
return $C_{sep}$;
```

Algorithm 1: selectClass(S)

```
    Data: dataset $S$
    Result: SVMLDT model
 1  Tree=NULL;
 2  L1_tree=DT(nominal features);
 3  Tree=L1_tree;
 4  for each leaf node $S_v$ in L1_tree do
 5      if $S_v$ is "pure" then
 6          label the leaf node;
 7      else
 8          take $S_v$ as the root of a L2_tree;
 9          class $C_{sep}$ =selectClass($S_v$);
10          run SVM between $C_{sep}$ and rest data;
11          if both leaves are pure != TRUE then
12              do 9-14 recursively for the impure leaf;
13          end
14          embed L2_tree into Tree;
15      end
16  end
```

Algorithm 2: SVMLDT(S, Tree)

3.6 Adaptive Optimal Load Shedding

SIPS often carry out remedial actions such as generation/load rejection, reactive power adjustment, etc. when certain stressed operation condition is observed. In this section, we mainly investigate the adaptive optimal load rejection scheme within a MAS.

3.6.1 Optimal Load Shedding via Dynamic Programming

Most load rejection schemes utilized in practice only specify the total amount of load to be shed under an emergency and do not take into consideration the current load profile and the significance of load at different locations. In fact, for a centralized case protection where the global load information is accessible, the load shedding problem can be formed as a 0-1 knapsack problem (3.4). By solving the optimization problem, the amount of load to be shed at each location will be determined. Similarly, via the load profile propagation, every agent in MAS is also able to get the global load information. Then each agent can solve (3.4) independently and figure out the remedial actions it should take and it will also be adaptive to the current system operation state.

$$
\begin{aligned}
\max \quad & \sum_{i=1}^{N}\sum_{j=1}^{K_i} x_{ij} v_{ij} P_{ij} \\
\text{s.t.} \quad & \sum_{i=1}^{N}\sum_{j=1}^{K_i} x_{ij} P_{ij} \le P_D - C \\
& x_{ij} \in \{0, 1\}
\end{aligned}
\tag{3.4}
$$

Objective of the optimization problem is to keep as much load value as possible. In (3.4), N is the total number of substations involved in the load shedding scheme and K_i is the number of feeders in substation i. P_{ij} (MW) and v_{ij} (\$/MWh) represent the amount of load on feeder j in substation i and the corresponding per unit load value respectively, P_D (MW) is the total load in the system and C (MW) is the amount of load must get shed when SIP is triggered. For load rejection schemes, C is normally predetermined by contingency analysis and we assume this value is known to every agent as a constant. Decision variables in the optimization is x_{ij}, and it takes value

0 to shed the load of the specific feeder and 1 to maintain the load. We can see that the overall objective of (3.4) is to preserve as much load value as possible after shedding required amount of load. Per unit load value ($/MWh) v_{ij} is defined as (3.5) where LMP is the locational marginal price and λ_{feeder} is a constant indicating the feeder significance. Hence, the value of a feeder takes the unit $/h and is defined as the product of P_{ij} and v_{ij}.

$$v_{feeder} = LMP \times (1 + \lambda_{feeder}) \tag{3.5}$$

3.6.2 Adaptive Load Shedding under Data Availability Attack

For a centralized load rejection scheme, it is not able to get the optimal load shedding solutions some slave substations are not responsive. For MAS, same issue exist but the best effort remedial actions can still be delivered due to the peer-to-peer nature of MAS. We assume that the impact of data availability attacks is that the completely connected MAS gets broken into several interconnected subgroups and nodes in different subgroups cannot exchange data with each other. In this case, the global data propagation become impossible. Data integrity attacks are not considered here. Given this assumption, our solution for adaptive optimal load shedding when under attack is achieved as following

3.6.2.1 Proposed strategy

1. The total amount of load to be shed is determined by contingency analysis and is still denoted as C.

2. Each agent records its load shedding proportion each time the protection is activated. Denote the historical average as p_i^{avg}, see (3.11).

3. Load profile are propagated among all agents within a subgroup regularly.

4. Each subgroup of agents resolves the same dynamic programming problem as in (3.6).

5. Shed load according to the solution.

$$\begin{aligned} \max \quad & \sum_{i=1}^{\widetilde{N}} \sum_{j=1}^{K_i} x_{ij} v_{ij} P_{ij} \\ \text{s.t.} \quad & \sum_{i=1}^{\widetilde{N}} \sum_{j=1}^{K_i} x_{ij} P_{ij} \leq \widetilde{P}_D - C \sum_{i=1}^{\widetilde{N}} p_i^{avg} \\ & x_{ij} \in \{0, 1\} \end{aligned} \tag{3.6}$$

When the communication among agents gets blocked, the originally interconnected MAS will be segregated into several interconnected subgroups. Within one subgroup, the dynamic load profiles can still be shared "globally". Then the optimization is carried out respectively for each subgroup. $\widetilde{N}$ in (3.6) is the number of interconnected agents in one subgroup, $\widetilde{P}_D$ represents the sum of available load in the subgroup and C in (3.4) is replaced by the load amount that this subgroup needs to shed. In this way, the load shedded in all subgroups will still sum up to C MW. A brief proof is provided below.

Use matrix $P_{n \times m}$ to represent the historical load shedding records for all the n agents and m times of protection activation as in (3.7). P_i^j is the amount of load that agent i has shed in the jth event. C still represents the total amount of load the SIP has to shed every time it is triggered. Equations (3.8)-(3.11) prove that when each subgroup of agents shed their load according to (3.6), C MW of load will be shed in total. One assumption we made here is that the historical load shedding records do not have attacks involved. It's reasonable since the emergency events are rare and any records injected by the adversary can be filtered out during post-event analysis.

$$P = \begin{bmatrix} P_1^1 & P_1^2 & ... & P_1^m \\ P_2^1 & P_2^2 & & P_2^m \\ . & . & ... & . \\ P_n^1 & P_n^2 & ... & P_n^m \end{bmatrix} \tag{3.7}$$

$$\sum_{i=1}^{n} P_i^j = C \tag{3.8}$$

$$\sum_{j=1}^{m}\sum_{i=1}^{n} P_i^j = mC \tag{3.9}$$

$$\sum_{i=1}^{n}\sum_{j=1}^{m} P_i^j = mC \tag{3.10}$$

$$\sum_{i=1}^{n} p_i^{avg} = \sum_{i=1}^{n} \frac{\sum_{j=1}^{m} P_i^j}{mC} = 1 \tag{3.11}$$

3.6.2.2 Performance comparison

The benefits of the proposed load shedding strategy include better economic performance and higher cyber attack resilience when under data availability attack. A comprehensive performance comparison between centralized SIPS and the decentralized one is provided in Figure 3.9. This figure highlights that centralized SIPS is vulnerable to cyber attack (comparison on the left), while decentralization helps improve the attack resilience (comparison on top). Since the proposed load shedding brings the load value into consideration, the economy of load shedding can also be improved by the decentralized SIPS (comparison at the bottom). Adaptiveness of the proposed SIPS is demonstrated when the MAS is attacked (comparison on the right).

To better demonstrate this, a toy example is provided for the rest of this section.

Figure 3.10 depicts a protection scheme used for load shedding under a predetermined system operating condition. The protection is supposed to shed 100MW load in total from substations 1, 2, 3 and 4 when the predetermined condition is met. All end nodes are interconnected with a SONET ring. Each substation has its own load profile labeled. In the context of the centralized SIPS implementation, A0 represents the master controller. Traditionally the centralized protection treats all the load indifferently and it sequentially polls about the available load from the substations and will trigger load shedding when enough load is reported voluntarily following a “first come, first serve” philosophy. Besides, we will consider the same cyber attack which results in blocking of

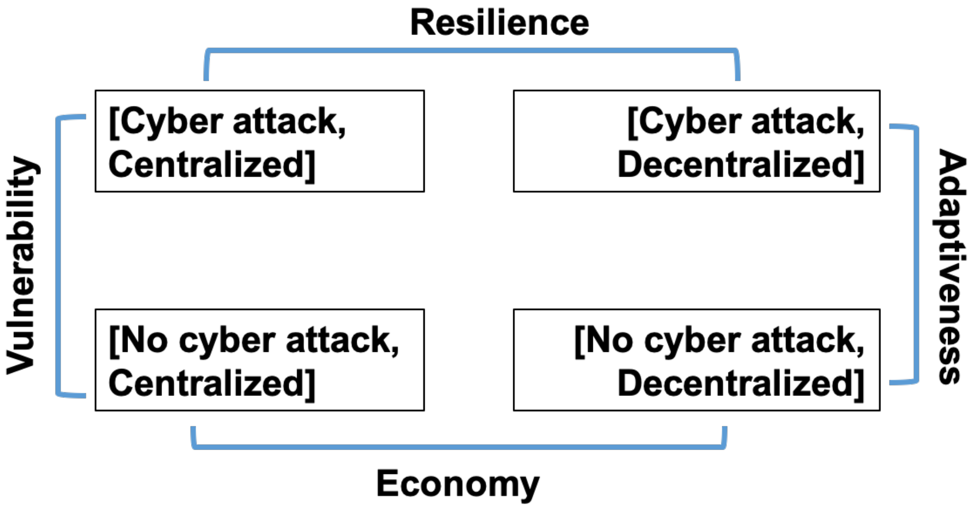

Resilience
[Cyber attack, Centralized]
[Cyber attack, Decentralized]
Vulnerability
Adaptiveness
[No cyber attack, Centralized]
[No cyber attack, Decentralized]
Economy

Figure 3.9 Performance comparison

communication at the two locations as depicted in Figure 3.10. Figure 3.11 compares the centralized and decentralized protection in terms of the performance of load shedding.

In Figure 3.11, the tuple for each scenario contains two elements. The first is the total loss due to load shedding, and second represents if system reliability is at risk. A red cross means the stability of the system is at risk. Four different comparisons can be made corresponding to Figure 3.9.

- **Vulnerability:** The vulnerability of centralized protection is obvious since the controller is able to communicate with only substation 1 and 2 when under cyber attack, which does not provide enough load to be shed. Thus, the system still has stability issues even 90MW of load is shed.

- **Economy:** Since the decentralized protection models the load shedding problem as an optimization problem, the MAS based protection will lead to less loss (92,000 $/h vs. 100,500 $/h).

- **Resilience:** When under the same data availability attack, the decentralized protection is able to shed enough load considering that all the four substations are upgraded into intelligent

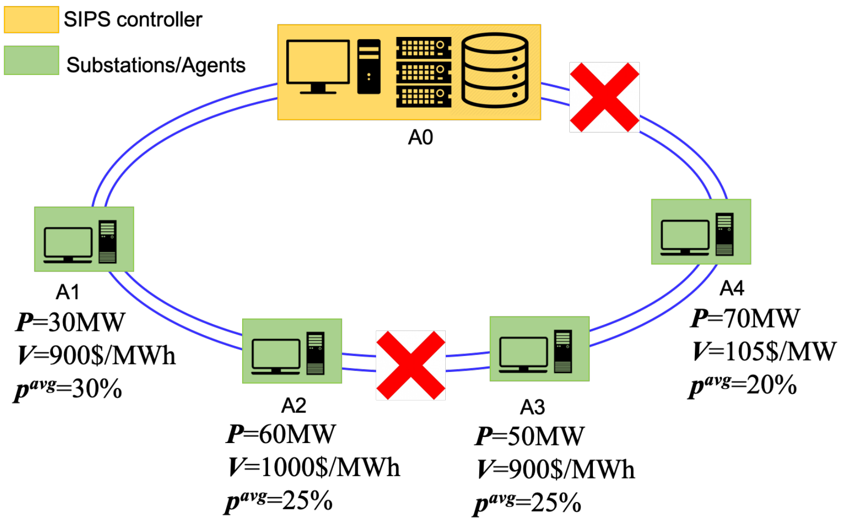


Figure 3.10 Load shedding protection scheme

	DoS Attack	No DoS Attack
Decentralized scheme	[30*900+25*1000+45*900=92,500$/h; -]	[(30+50)*900+20*1000=92,000 $/h; -]
Centralized scheme	[30*900+60*1000=87,000$/h; +]	[30*900+70*1050=100,500$/h; -]

Figure 3.11 Load shedding performances

agents and each has the capability to detect the system operating condition and shed the load as needed. Thus the system stability will not be an issue if the decentralized protection is adopted.

- **Adaptiveness:** The optimization problem as modeled in (3.4) cannot get resolved when under cyber attack for each subgroup. In contrast, the adaptive version of load shedding as represented in (3.6) adds more adaptiveness to the protection with negligible degrade in terms of economic performance.

3.6.3 Agent Consensus Achievement

As depicted in Figure 3.5, before the load shedding can be carried out by all the agents, the consensus about the current system condition has to be achieved among the interconnected agents. For a MAS, "consensus" means that an agreement among all the agents is attained pertaining to a certain quantity of interest. According to Lemma 1 in reference [71], for a connected undirected graph $G(V, E)$, the algorithm in (3.12) asymptotically solves an average consensus problem for all initial states. x_i is the state value of node $i \in V$, and a_{ij} represent the element of the adjacency matrix of G, when node i and j are connected by an edge $e \in E$, $a_{ij} = 1$ otherwise $a_{ij} = 0$. (3.13) provides the discrete-time average consensus model, in which ϵ is a small time step.

In the consensus algorithm, state x is the decision about load shedding of every agent. x takes value 1 when the load shedding is necessary, and 0 otherwise. The average consensus algorithm will asymptotically converge to the average of initial x values. Therefore, as long as the minority of the agents make type 1 error (false positive) or type 2 error (false negative), the MAS as a group is able to get to the same conclusion regarding the anomaly detection outcome with a specified threshold. When the consensus value reaches an equilibrium that is beyond the threshold, the anomaly is confirmed and such that the load shedding will be conducted by all the agents.

$$\dot{x}_i(t) = \sum_{j \in N_i} a_{ij}(x_j(t) - x_i(t)) \tag{3.12}$$

$$x_i(k+1) = x_i(k) + \epsilon \sum_{j \in N_i} a_{ij}(x_j(k) - x_i(k)) \tag{3.13}$$

Ideally to implement a consensus algorithm in MAS, the messages about the state values should have time stamps (such as PMU packets) such that data from all agents can be synchronized in the discrete-time calculation in (3.13).

3.7 Evaluation Results

In this section, one SIPS that is used in Western Electricity Coordinating Council (WECC) [40] is selected as study case to evaluate the proposed attack-resilient SIP design. Palo Verde is a nuclear power plant in Arizona and it has 3 nuclear units whose generation is mainly consumed by the load in the area of Arizona and Nevada. Through contingency analysis people have found that when any two units in Palo Verde plant get disconnected from the system, electricity will be imported from California and Oregon which results in the thermal overload on the COI (California-Oregon Inter-tie). Geographic locations of the nuclear power plant is noted in Figure 3.12. The SIPS detailed in [40] is installed to shed 120MW amount of load close to Phoenix and this helps avoid the COI overloading. The scheme arms load shedding when the sum of any two generators' output is greater than 2550MW, and it initiates load shedding only if two units out of three get disconnected. The protection scheme is equipped at Palo Verde plant and the load shedding is performed by 14 substations near Phoenix. The 15 stations involved in this SIPS is connected via a SONET ring (Synchronous Optical Network), which is often used in power system to improve the communication reliability.

3.7.1 Experiment Set-up

After examining the WECC system behavior and the specific protection needs elaborated in [40], it is found that the same system behavior and needs for load shedding protection can be reproduced in the IEEE 39-bus system, which is depicted in Figure 3.13.

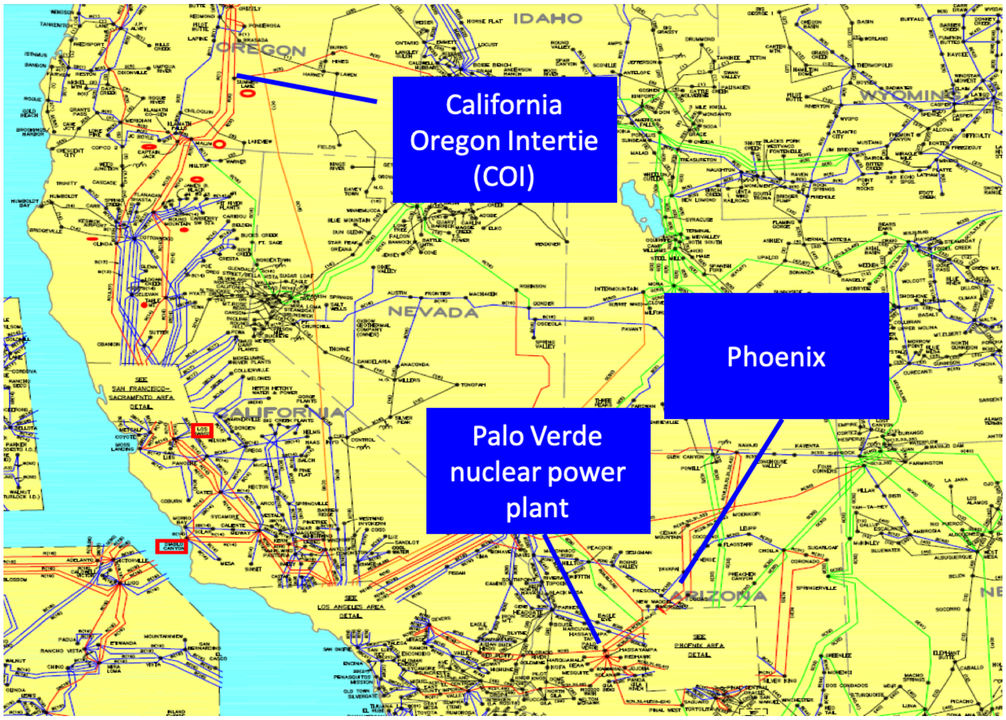

California Oregon Intertie (COI)
Phoenix
Palo Verde nuclear power plant
IDAHO
OREGON
WYOMING
NEVADA
CALIFORNIA
ARIZONA

Figure 3.12   Load rejection SIPS in WECC

The bottom right region in the IEEE 39 bus system as shown in Figure 3.13 has 2350MW generation (37.9%) and 927.5MW load demand (15%), and the extra generation will be sent out through tie lines 16-19, 21-22 and 23-24. Simulation shows that when G8 is out of service, line 16-19 will gets overloaded and at least 100MW load needs to be shedded from load center near G8 so as to clear the overloading as well as the possibility of potential cascading events. In the original SIPS design, the load to be shedded are those close to Palo Verde with low priority, and similarly, the same centralized load rejection scheme is implemented in 39 bus system as following:

1. Load buses including bus 25, 26, 27 and 28 are selected as the load sources. 20% from 224MW on bus 25, 20% from 139MW on bus 26, 20% from 281MW from bus 27 and 10% from 206MW are preselected as the sheddable load.

2. The communication media is a SONET ring which connects the plant and 4 substations as in Figure 3.14. It contains 5 interconnection links.

3. Whenever the generation level of G8 is above 500MW, the sheddable load gets armed in sequence from bus 25, and the arming signal will be turned off when the total load prepared for shedding is more than 100MW.

4. In case when G8 gets tripped, a load shedding command from the power plant will be sent to all substations in the loop and upon receiving this command, load armed previously will get shedded.

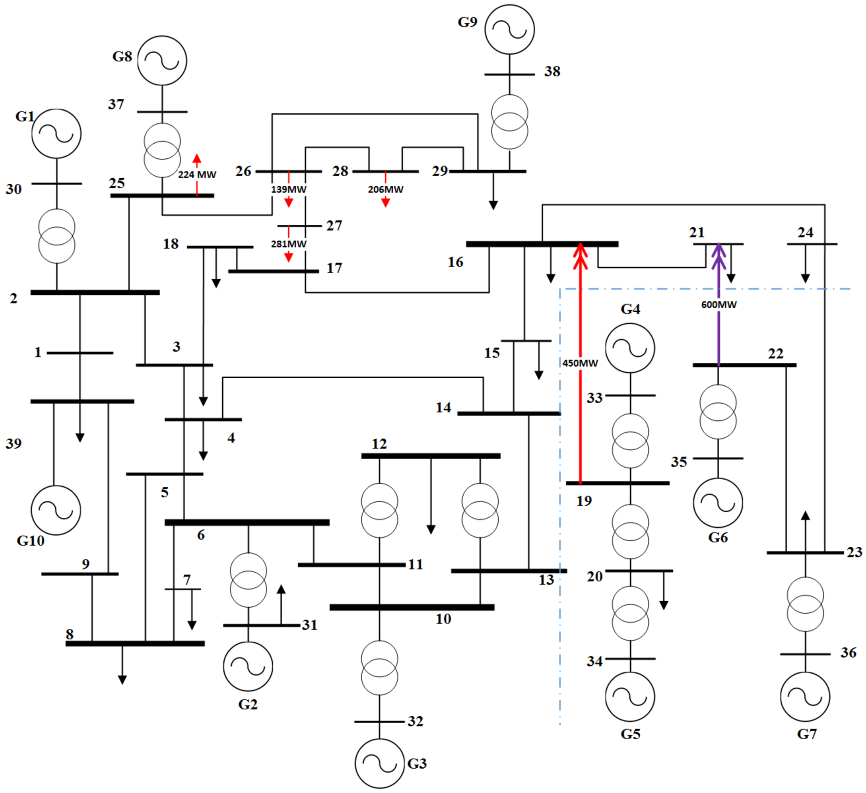


Figure 3.13 IEEE-39bus with the load rejection RAS

In this study, the line overload threshold is set to 1.3 times of the pre contingency power flow, and the overload tolerances are set to 50s for line 16-19 and 60s for line 21-22.

Overall experiment set-up is shown as Figure 3.15. The 39-bus system will be simulated in real-time with Opal-RT simulator and agents that represent substation 25 - 27 are virtual machines whose behaviors are coded in Java. The synthetic data exchange is achieved through OPC server

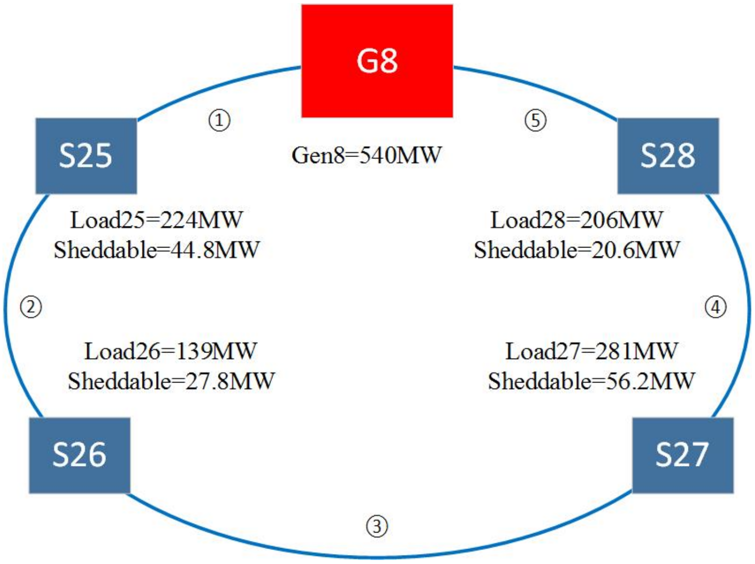


Figure 3.14 Communication topology

and peripheral functions such as the centralized protection scheme, attacks and load changes are all done with python scripts. Before getting to the evaluation of proposed solution, the rest of this subsection will further analyze the nature of the system and the centralized protection scheme.

3.7.1.1 System Dynamics Exploration

System dynamics after loss of G8 while without any protection scheme are investigated to demonstrate the system nature. Table 3.2 shows 3 tie line currents under different scenarios. The sequential critical events include G8 tripping (about 85s), line 16-19 tripping (about 135s) and line 21-22 tripping (about 190s). Figure 3.16 shows the frequencies at different generator buses. After G8 is out of service, line 16-19 first gets overloaded ($6.542 > 1.3 \times 4.781$) and is tripped after 50 seconds. Line 21-22 will immediately become overloaded ($9.379 > 1.3 \times 6.552$) after that. It can be observed that after G8 and line 16-19 are tripped, the system is still stable. However, when line 21-22 is also tripped 60s later, instability occurs. Figure 3.17 shows the frequency response with load rejection protection equipped, which shows that after G8 trips, the rest of system still maintain stable. Frequency of G8 is set to 60Hz after it is disconnected from the system.

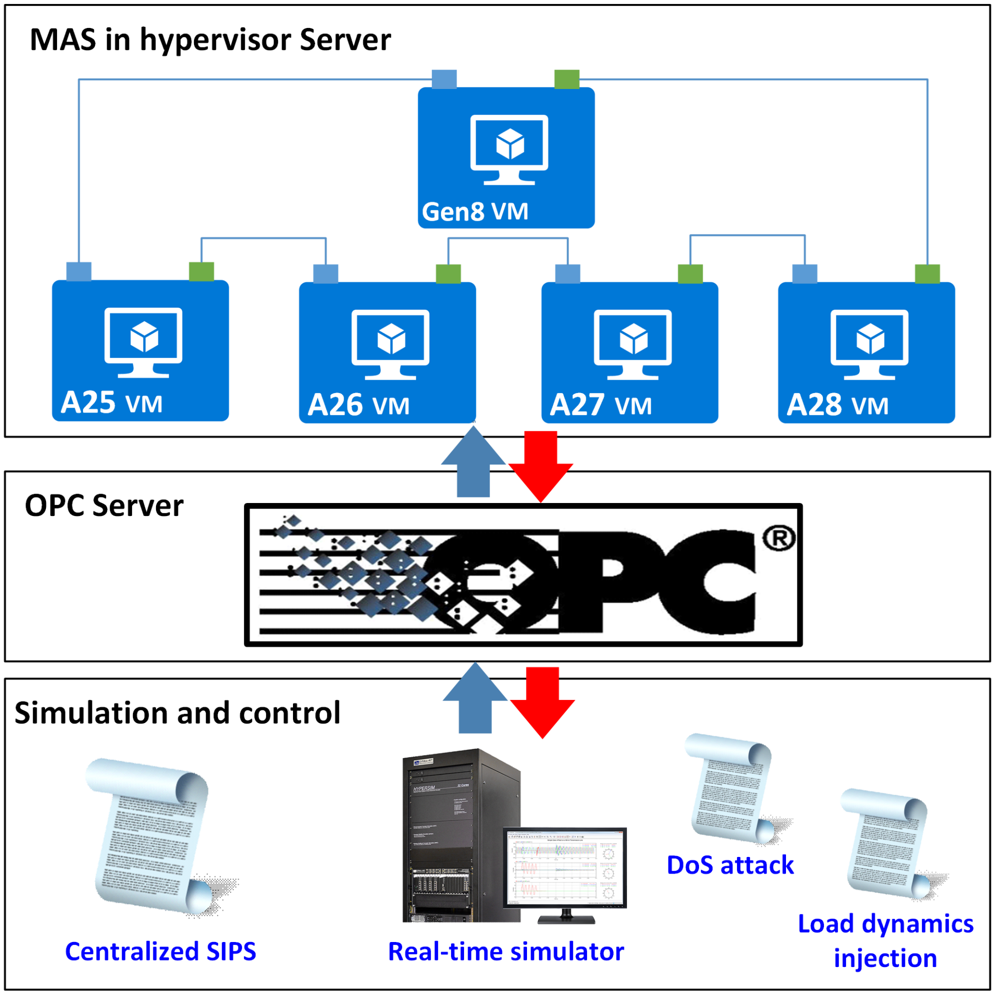


Figure 3.15 Experiment based on PowerCyber testbed

Table 3.2 Power flow of critical tie-lines

	L16-19(pu)	L21-22(pu)	L22-23 (pu)
Pre-contingency	4.781	6.552	0.4466
G8 tripped	6.542	7.560	0.3000
L16-19 tripped	0	9.379	0.3809
L21-22 tripped	0	0	Unstable

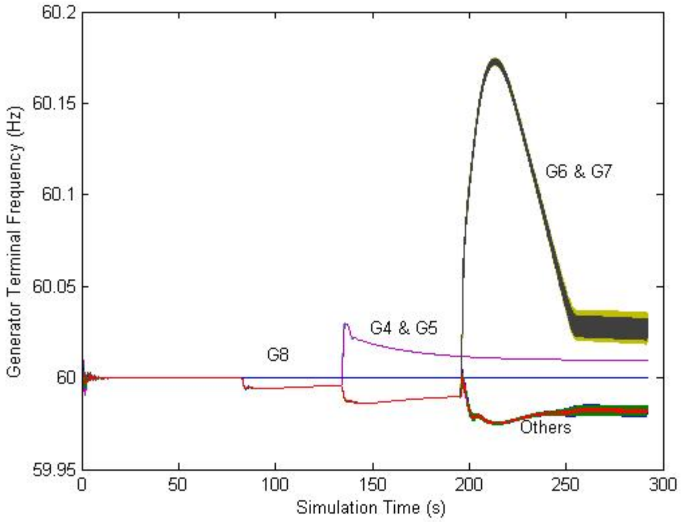


Figure 3.16 Frequency response without load rejection protection

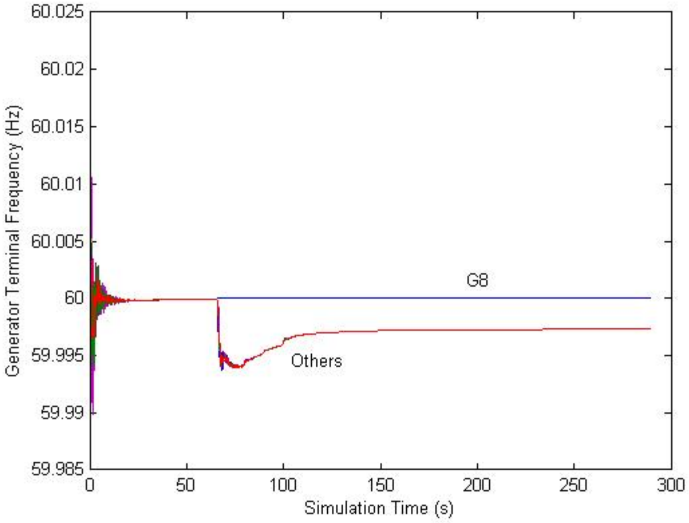


Figure 3.17 Frequency response with centralized load rejection protection

3.7.1.2 Centralized Protection Performance under Attacks

In this section, DoS (Denial of Service) attack and replay attack are chosen to test the performance of the centralized protection scheme. Since the SONET ring can tolerate single point fault on the ring, the DoS attack are designed as a coordinated double points attack where links ① and ⑤ in Figure 3.14 are attacked. Attackers will trip G8 after starting DoS attack. By replay attack, the recorded arming command and shedding command are injected in ① when G8 is operating normally. To save space, only the frequency response when the scheme is under replay attack is given as Figure 3.18 and frequency response under DoS attacks is the same as Figure 3.16. This demonstrates that the centralized protection scheme is quite vulnerable to cyber attacks.

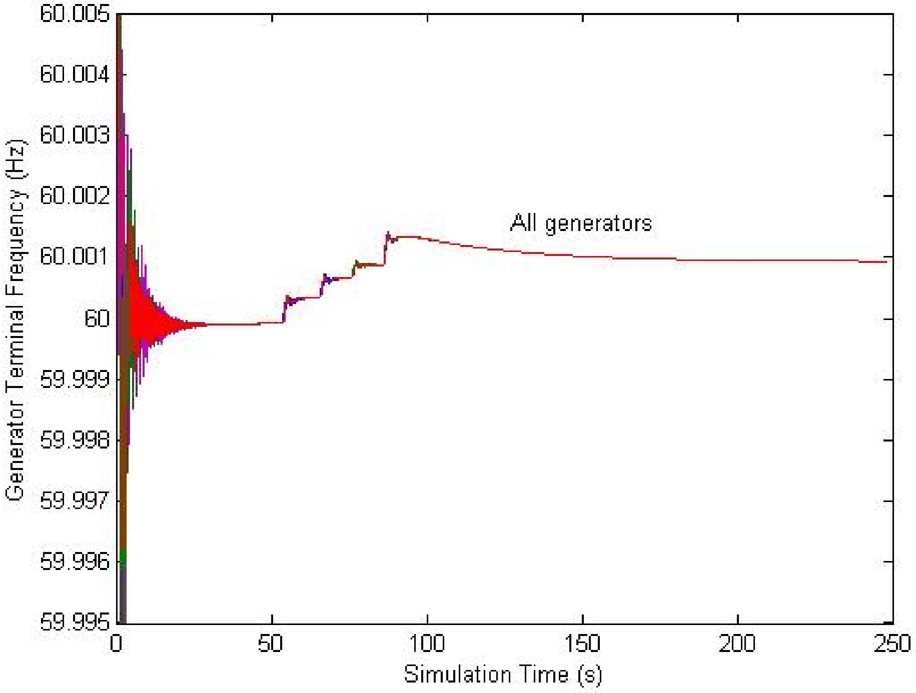


Figure 3.18 Frequency response under replay attack with centralized load rejection protection

3.7.2 Anomaly Detection Evaluation

For each agent, we have listed all the selected features utilized in anomaly detection in Table 3.3. First column **Status** includes the relevant breakers' statuses for every agent. For example, S_{25_26} represents the status of the breaker on line 25-26 near bus 25. Second column contains features relevant to bus voltage. V_i and a_i are the voltage magnitude and relative phase angle respectively of bus i. Active power injections (both generation and load) are involved as shown in column **Inj.**. Features in the fourth column are power flows and that in the fifth is system frequency. According to the data taxonomy in section II, aforementioned features are all physical measurements. In contrast, Δt is a processed metric and it represents the time interval between two sequential frequency dips. Each agent records last two values of this metric, which characterizes the last three frequency dips. C_{arm} and C_{shed} are cyber information which indicate the presence of the "Arm" and "Shed" commands received from the legacy protection master.

Table 3.3 Measurements of agents

Agent ID	**Status**	**V**	**Inj.**	**Flow**	*freq*	**timing**	**Cmds**
A25(Bus 25)	S_{gen_8} S_{25_2} S_{25_26}	V_{25} a_{25}	P_8 L_{25}	P_{25_2} P_{25_26}	f_{25}	$\triangle t_{last1}$ $\triangle t_{last2}$	C_{arm} C_{shed}
A26(Bus 26)	S_{26_25} S_{26_27} S_{26_28} S_{26_29}	V_{26} a_{26}	L_{26}	P_{26_25} P_{26_27} P_{26_28} P_{26_29}	f_{26}	$\triangle t_{last1}$ $\triangle t_{last2}$	C_{arm} C_{shed}
A27(Bus 27)	S_{27_26} S_{27_17}	V_{27} a_{27}	L_{27}	P_{27_26} P_{27_17}	f_{27}	$\triangle t_{last1}$ $\triangle t_{last2}$	C_{arm} C_{shed}
A28(Bus 28)	S_{28_26} S_{28_29}	V_{28} a_{28}	L_{28}	P_{28_26} P_{28_29}	f_{28}	$\triangle t_{last1}$ $\triangle t_{last2}$	C_{arm} C_{shed}

Synthetic training data are collected from following scenarios while system loads are configured as static values.

- the targeting contingency (i.e. G8 trips) occurs without attack and the centralized protection shed the load successfully.

- irrelevant natural contingencies occur with centralized protection and without attack.
- when the legacy centralized protection is under a single-point or double-point DoS attack on the ring network (15 scenarios) and then G8 gets tripped by the adversary.

Training data obtained are labeled according to the CPS operation states discussed in section 3.5. This data set is split into training subset and test subset, and the training subset is used to attain the SVMLDT model while the test subset is used to evaluate the training error.

SVMLDT is realized with R language and packages “C50” and “e1071” are selected for the DT and SVM implementation respectively. As a comparison, four different classification methods: I. C50 (decision tree), II. DTSVM, III. SVMLDT, IV. K Nearest Neighbors (KNN) are trained based on training subset. Table.3.4 - 3.7 shows the testing results of the test subset as confusion matrices for agent 25 - 28. We can notice that C50 and SVMLDT obtain better detection accuracy given that the false classification only occurs between either states **N** and **P** or **A** and **E**. Both cases are acceptable compared to mis-classifying state **N** as **A** or vice versa. Random forest is also applied to the same data set, and the accuracy of the five methods is summaries as in Figure 3.19. Two different measures of accuracy are utilized for each classification method. Accuracy 1 (A1) is the normal accuracy that considers all the classes separately while accuracy 2 (A2) is calculated after merging the states N and P and states A and E respectively. The rationale behind the state merging is that for state N and P, no load shedding is required, and in contrast, load shedding is required for both state A and E. Therefore, the mis-classification between the two states that are merged will not result in differences in terms of protection activities. A2 for every method is higher than A1 because of the state merging. Figure 3.20 illustrates the detection accuracy of A3 before states merging with different parameter selections of SVMLDT. (a) depicts the detection accuracy where the radial kernel is selected for SVM, the x axis represents the γ selected and the box plots vary in the cost value in the quadratic optimization. Similarly, (b) provides box plots in terms of varying constraint violation cost, but the kernel selected in 3 degree polynomial. It's shown that the radial kernel can result in better accuracy and when the value of *gamma* takes 0.1 and the cost value takes 10000, the accuracy is close to 0.995.

Table 3.4 Algorithms comparison for Agent 25

I	N	P	H	A	E	II	N	P	H	A	E
N	633	15	0	0	0	N	629	14	0	0	0
P	17	355	0	0	0	P	8	331	0	13	0
H	0	0	2333	0	0	H	0	0	2333	0	0
A	0	0	0	8398	0	A	0	10	0	8732	0
E	0	0	0	0	1992	E	13	15	0	13	1992
III	N	P	H	A	E	IV	N	P	H	A	E
N	627	11	0	0	0	N	629	27	0	0	0
P	22	359	0	0	0	P	16	268	0	1	0
H	1	0	2333	0	0	H	0	0	2333	0	0
A	0	0	0	8387	3	A	5	75	0	8397	0
E	0	0	0	11	1989	E	0	0	0	0	1992

In order to evaluate the anomaly detection model with data not been observed during training, we collect 2 more synthetic data sets. The first is collected from a scenario where legacy centralized protection is running while the DoS attack is undertaken on both sides of VM G8 and then G8 gets tripped. The second contains data collected during normal system operation involving a natural line fault and dynamic load changes. The online detection results from Agent 28 are plotted in Figure 3.21. Left vertical axis of all the 4 subplots represents the system frequency and the right axis represents the CPS operation state. For example, a pulse ends at value 2 represent the hidden failure state **H**. From the comparison of subplot (c) and (d), we can tell that both DT and SVMLDT are able to distinguish the targeting events correctly. But the comparison between (a) and (b) shows that DT has committed more misclassification between **N** and **P** than the SVMLDT on the unobserved test data.

Figure 3.22 shows the timing performance of all the methods. It turns out that the SVMLDT has least satisfactory timing performance. But since load shedding allows much longer time (mins) compared to the detection time required (milliseconds), it will not be a significant shortcoming. However, the good interpretability and accuracy of SVMLDT can still be leveraged.

Table 3.5 Algorithms comparison for Agent 26

I	N	P	H	A	E	II	N	P	H	A	E
N	638	31	0	0	0	N	630	40	0	0	0
P	12	339	0	0	0	P	6	292	0	12	0
H	0	0	2333	0	0	H	0	0	2333	0	0
A	0	0	0	8398	0	A	0	19	0	8340	0
E	0	0	0	0	1992	E	14	19	0	46	1992
III	N	P	H	A	E	IV	N	P	H	A	E
N	630	34	0	0	0	N	632	52	0	0	0
P	20	336	0	0	0	P	11	237	0	0	0
H	0	0	2333	0	0	H	0	0	2333	0	0
A	0	0	0	8328	0	A	7	81	0	8398	0
E	0	0	0	70	1992	E	0	0	0	0	1992

3.7.3 Adaptive Optimal Load Shedding Evaluation

To evaluate the optimal adaptive load shedding algorithm, first we need to convert the lumped loads on bus 25 $\sim$ 28 in the IEEE 39-bus system into feeder specific loads. Therefore, we assume that substations 25 $\sim$ 28 each has 6, 4, 6 and 8 feeders respectively and each feeder transmits a fixed proportion of the total load on this bus. Load profile of a feeder includes two facets - amount and value, and the load fluctuation should be considered and added to the load values provided in the base case.

3.7.3.1 Dynamic load profile

In order to make the load shedding scenarios more realistic, load dynamics should be involved in the simulation. We utilize Area Control Error (ACE) values observed by Mid-continent Independent System Operator (MISO) to mimic load changes in the IEEE 39-bus model. Two hundred ACE values collected from MISO's website [72] are used to generate the ACE Probability Density Function (PDF) via kernel density estimation. Then we can draw ACE values from the PDF to emulate the system load changes. These ACE values are scaled up to fit IEEE 39-bus model and then they are proportionally split and assigned to each feeder. As for the "value" of the load on each feeder, typical LMP values from the real time market of MISO are leveraged. The LMP for

Table 3.6 Algorithms comparison for Agent 27

I	N	P	H	A	E	II	N	P	H	A	E
N	637	18	0	0	0	N	636	41	0	0	0
P	13	352	0	0	0	P	2	309	0	4	0
H	0	0	2333	0	0	H	0	0	2333	0	0
A	0	0	0	8398	0	A	0	8	0	8363	0
E	0	0	0	0	1992	E	12	12	0	31	1992
III	N	P	H	A	E	IV	N	P	H	A	E
N	635	34	0	0	0	N	629	53	0	0	0
P	15	336	0	0	0	P	13	263	0	3	0
H	0	0	2333	0	0	H	0	0	2333	0	0
A	0	0	0	8376	0	A	8	54	0	8395	0
E	0	0	0	22	1992	E	0	0	0	0	1992

a bus is randomly sampled and then the load value is calculated as in (3.5) to represent the true value of load. With this, we are able to inject the load dynamics in IEEE 39 bus model.

3.7.3.2 Adaptive control results

We first simulate the targeting contingencies (i.e. G8 is out of service) for ten times and run adaptive load shedding algorithm such that the historical load shedding records are obtained as shown in (3.7). Then the algorithm is tested against ten new contingencies that occur under different system operation states. The objective value of the 0-1 knapsack problem is shown as Figure 3.23. The “2-2 subgroups” represent a scenario where substation 25&26 and 27&28 are separated into two subgroups due to DoS attack. Similarly, “1-3 subgroups” is the scenario that substation 25 is isolated from other substations. It can be seen that when the MAS is separated into subgroups, the load value preserved after load shedding is not as much as that when all the agents are interconnected. However, the same amount of load are still being shed to maintain the system stability. That implies that the decentralized agents is capable of delivering the remedial actions to their best effort when under coordinated DoS attack, which is a desired improvement compared to the centralized SIP which will not shed load at all when under the same attack.

Table 3.7 Algorithms comparison for Agent 28

I	N	P	H	A	E	II	N	P	H	A	E
N	641	12	0	0	0	N	638	78	0	0	0
P	9	358	0	0	0	P	2	279	0	8	0
H	0	0	2333	11	0	H	0	0	2333	3	0
A	0	0	0	8387	0	A	2	5	0	8376	0
E	0	0	0	0	1992	E	8	8	0	11	1992
III	N	P	H	A	E	IV	N	P	H	A	E
N	637	72	0	0	0	N	629	34	0	1	0
P	13	298	0	0	0	P	8	318	0	4	0
H	0	0	2333	6	0	H	0	0	2333	0	0
A	0	0	0	8381	8	A	13	18	0	8393	0
E	0	0	0	11	1984	E	0	0	0	0	1992

3.8 Summary

This chapter presents a decentralized load shedding SIPS based on MAS. An data-driven algorithm named as SVMLDT is proposed for anomaly detection and the adaptiveness of the protection is enhanced by the optimal adaptive load shedding strategy proposed.

Conversion from a centralized function architecture to a decentralized one eliminates the possibility of function loss due to the compromise of the centralized end node, since the intact agents of a MAS could still provide the protection function even the other agents are rendered inoperable by cyber attacks. This conversion only requires that certain intelligence is added to the substation agents involved in the legacy centralized protection, instead of replacing the centralized protection with the MAS.

The data-driven anomaly detection capability is added to the MAS to ensure that each agent can stay situational-aware, especially facing cyber attacks. Diversity of the data sources and algorithms adopted by different agents used in anomaly detection further improves the cyber attack resilience. Besides, as the last line of defense, the anomaly detection module serves the purposed of validating the protection commands from the legacy centralized module.

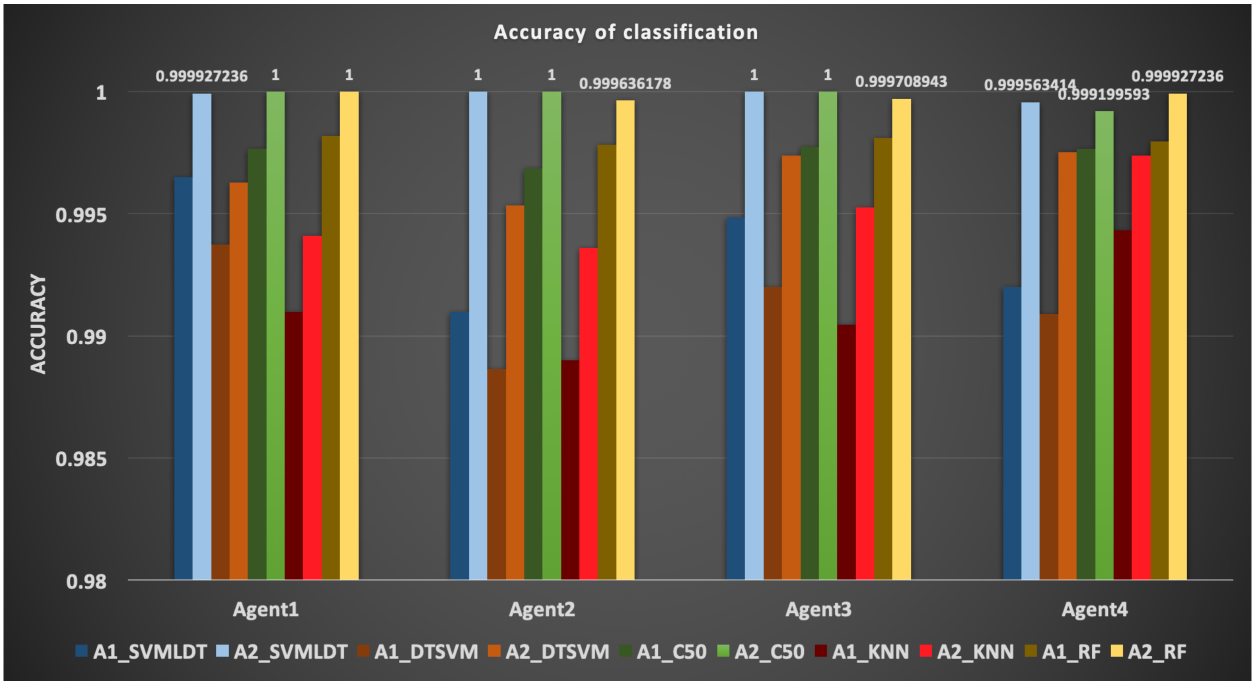
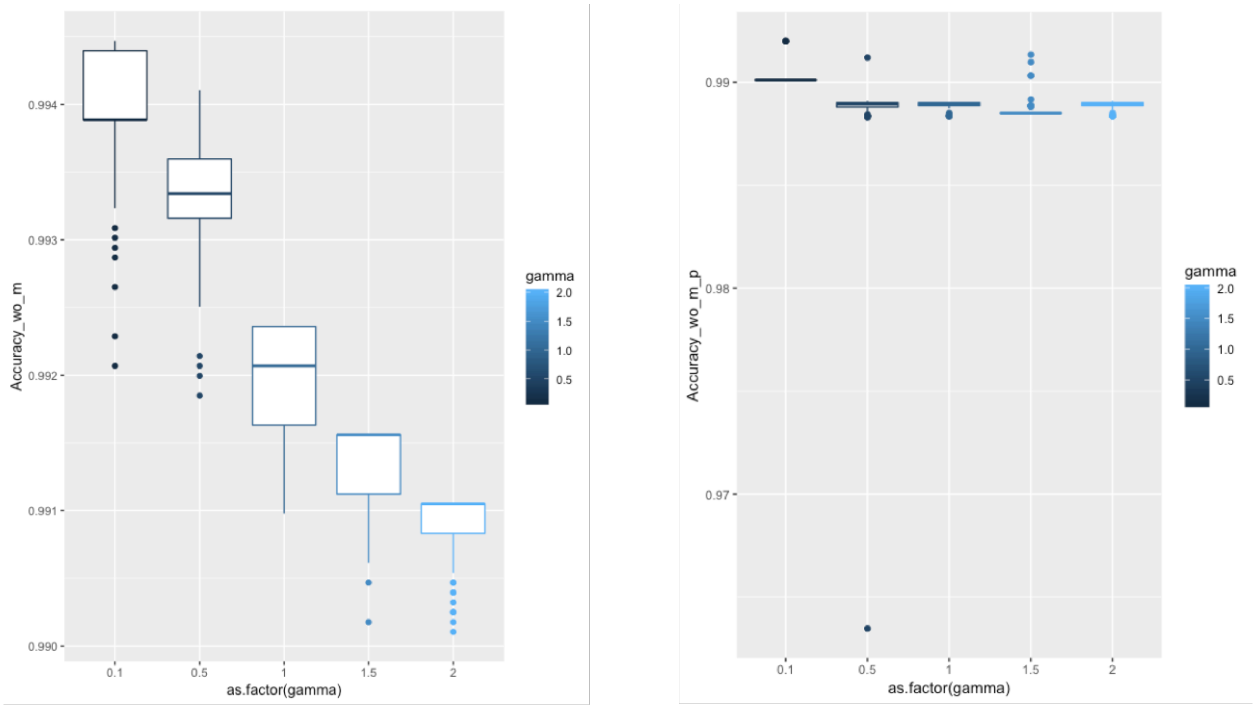


Figure 3.19 Accuracy of classification

The adaptive load shedding strategy proposed not only improves the way how load should be shed when it becomes necessary, it also stays effective when the MAS is under data availability attack. The load shedding scheme proposed is not able to tolerate data integrity attacks, what may require the adoption of cryptography techniques.



(a) Radial kernel

(b) Polynomial kernel

Figure 3.20 Parameter tuning for SVMLDT

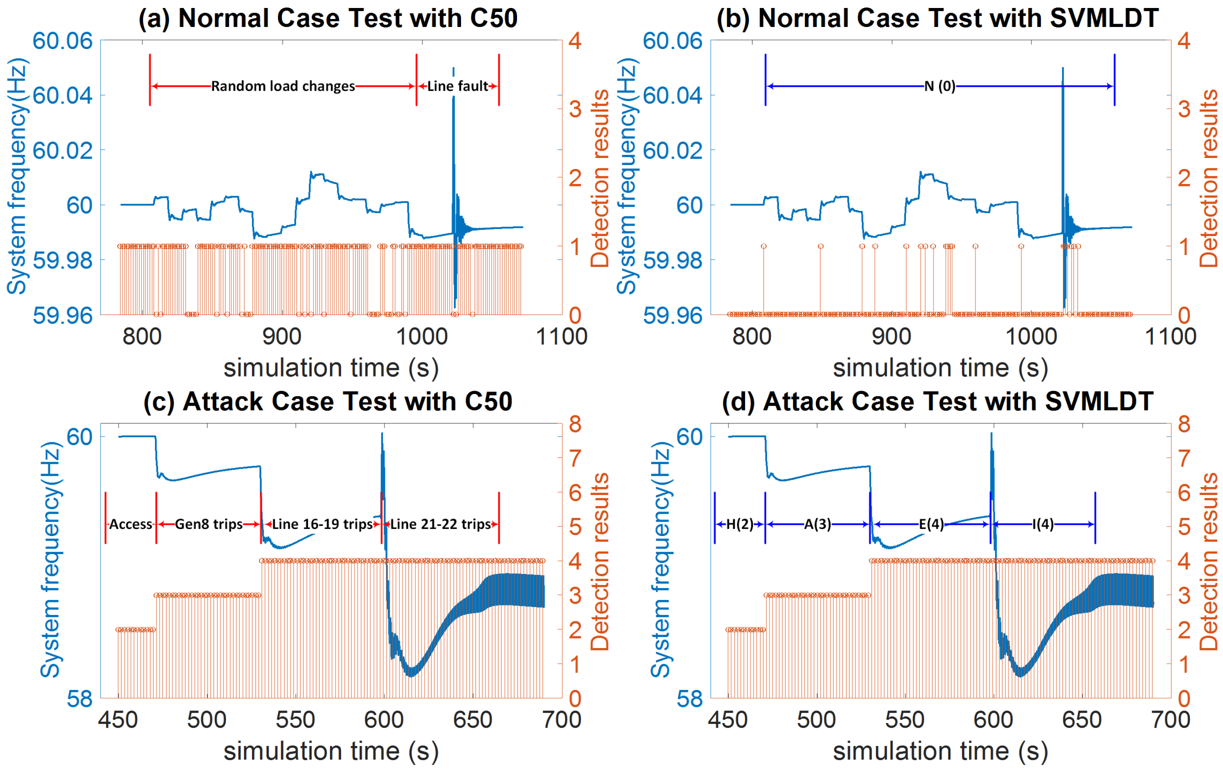


Figure 3.21 Agents overall operation flowchart

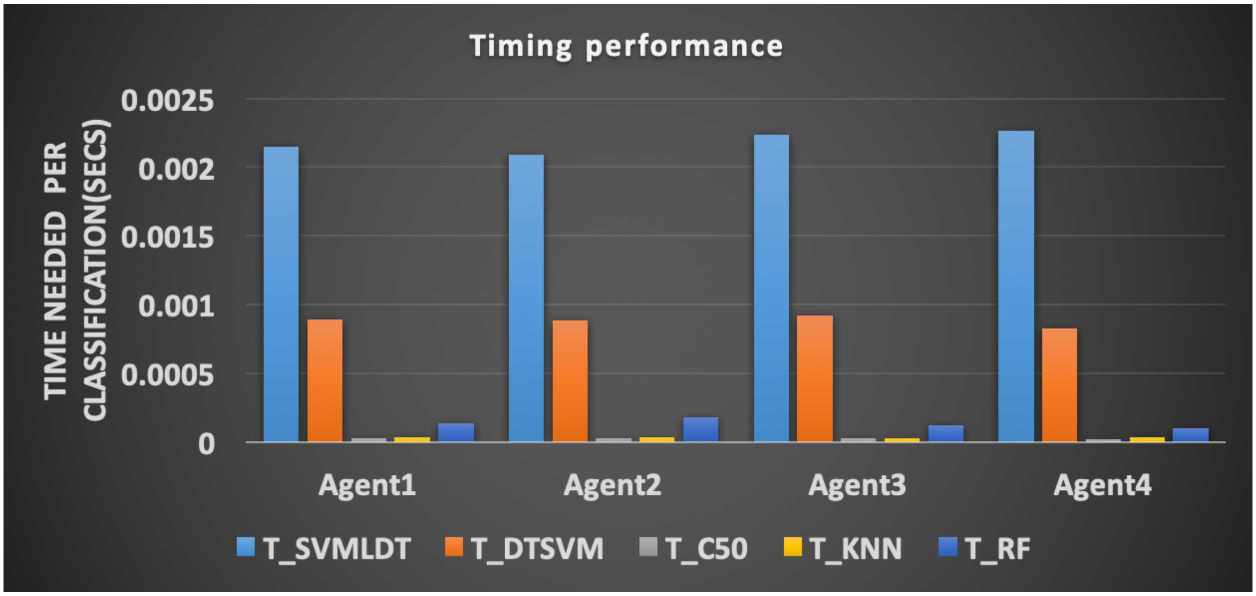


Figure 3.22 Algorithm timing performance

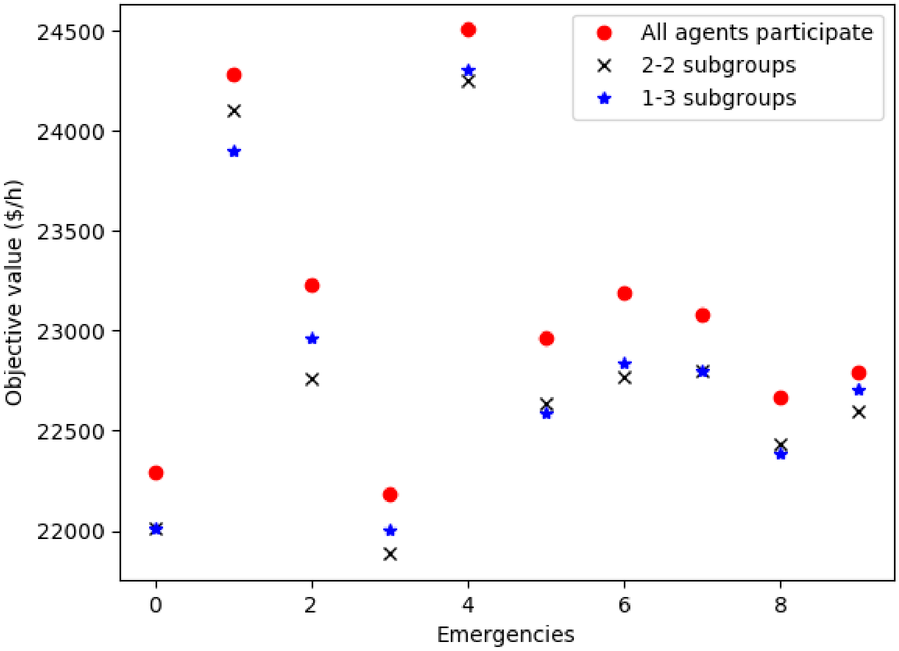


Figure 3.23 Load shedding results

CHAPTER 4. CYBER ATTACK DETECTION AND MITIGATION FOR GENERATION CONTROL

4.1 Problem Statement

Maintaining the balance of electricity generation and consumption is the most critical task of power system daily operation such that system frequency can be constrained around its nominal value (60 Hz in North America). Various control strategies including generation governors and Automatic Generation Control (AGC) have been deployed to achieve this goal, and they are commonly known as ”balancing and frequency control” [73]. System frequency is controlled by these strategies through regulation of the power output of different generators to match the system load. This functionality helps guarantee the adequacy of power supply when the power system undergoes disturbances such as random load changes, loss of generation units, etc.

Balancing and frequency control highly relies on information and communication technologies. For instance, the AGC algorithm running at a control center relies on the Supervisory Communication And Data Acquisition (SCADA) network to collect measurements from different substations and then send the appropriate control commands back to control the output of different generating units. As the SCADA increasingly leverages public communication network infrastructures for the data exchange, the attack surface of balancing and frequency control is exposed to the adversary. Not only the measurements of AGC such as system frequency, tie-line flows among different balancing areas, etc. can be manipulated to induce inaccurate control errors, the generation control commands are also vulnerable to cyber attacks during transmission from control center to generating units. Although rare, some utilities do allow remote operators to control the generation via Virtual Private Networking (VPN) connection during an emergency [74]. When the adversary succeeds in cracking the VPN connection, very likely by first tampering with the remote operators' home area network, he will be able to directly infiltrate into the power plants and carry out malicious controls of the generation. Some utilities may deploy a dedicated communication network as a "Security as obscurity" defense, but careless or disgruntled employees can still cause severe problems to generation control [6, 25]. Under/over frequency load shedding protection, stability issues and even cascading failures might be triggered [15] due to this type of attack.

This chapter proposes methodologies that leverage the real-time generator power output and ACE values received by different generating units to detect and mitigate malicious generation controls at station level. The anomaly detection proposed is an online function that would be implemented for each generator participating in AGC. It depends on the conformity of different generating units by checking the data collected from multiple power plants within a sliding window. The mitigation strategy proposed also needs local data as well as the data from neighbors via a peer-to-peer communication network, the main idea is to estimate the ACE for one generator based on its peers' ACEs. The estimated ACE will be used to replace the suspicious ACE. Assumptions made in this chapter are listed below.

- Stations involved in the AGC can be upgraded into intelligent agents.
- MAS is utilized on station-level, i.e., an agent represents a whole substation.
- Peer-to-peer communication is enabled among agents.
- Transition between two economic dispatches is not considered in the anomaly detection. In other words, the detection will be disabled during every re-dispatch
- Not all the peer data sent to one generator can be compromised at the same time
- Dead band of the governor is ignored.

4.2 Related Work

False data injection attacks (FDIA) is the most widely investigated type of cyber attack that targets AGC. FDIA results in miscalculated ACE at the control center by manipulating the measurements such as system frequency, tie-line flows, etc. When the wrong ACE values are received by the generating units and used to raise or lower the unit output, the generation will mismatch the load so that impacts can be induced. The attacking on AGC is formed as an optimization problem in both [75, 76]. It's been demonstrated that the attack is able to minimize the time needed to drive the system frequency to a desired off-nominal value stealthily assuming that the AGC measurements can be modified. Hardware-in-the-loop test-bed based evaluation of the impact of cyber attacks on AGC has been carried out in [15]. This work demonstrates how the MITM attack could influence the automatic generation control by manipulating the measurements or the control commands. In reference [77], the authors model the load and frequency control with a switched system and show that the instability of the system can be resulted from DoS attack that is launched with proper timing against the measurement sensing path.

Countermeasures against malicious generation controls have also long been explored by researchers, and most countermeasures are proposed to be set up at the control center level. In [78], a model-based attack detection method for AGC is presented, which is aided by short-term load forecasts. The detection rules are proposed in terms of the absolute and cumulative value of Area Control Error (ACE) observed. If either of the two is out of the statistical bounds, the measurements are treated as anomalous. [16] has implemented these rules on the hardware-in-the-loop testbed as further validation. A Unknown Input Observer (UIO) based anomaly detection and identification strategy is proposed in [75], and again the attack under discussion is FDIA on the measurements. A mitigation strategy based on redundant measurements is proposed in [76] such that state estimation can be utilized to determine the measurements needed in AGC by removing the bad measurements identified from the over-observed system.

Figure 4.1 provides a comprehensive schematic of the attack-defense competition in smart grid related to generation control. The left part of Figure 4.1 depicts the structure of today's power system. The SCADA network ① is utilized in the AGC and it connects the stations and the control center. Peer-to-peer networks linking different stations also exist (see ②), but their capability is limited and those networks often are only involved in certain local functions such as transmission line pilot protections. To realize the detection and mitigation strategy proposed latter, such peer networks might need to be augmented in terms of capabilities like bandwidth but they are not required to be as powerful as the SCADA network. With this cyber layer structure, the attacks targeting AGC can potentially occur at four different locations: ❶, the attacker can choose to intrude into the control center directly; ❷, the attacker can modify the measurements before they arrive the control center; ❸, the attacker does have an option to directly manipulate the control signal while it is being sent back to the stations; ❹, the attacker can also intrude into a station to cause impacts by tampering actuators. Most of the existing literature focus on the detection of attacks targeting the AGC measurements, namely type ❷ depicted in Figure 4.1 and the detection modules are placed at the control center level. Anomaly detection in control center is facilitated by the convenient redundant data access, therefore, the requirement for detection accuracy could be satisfied. However, detection functions located in the control center are not able to handle scenarios where false ACE injection or ACE manipulation happens along the actuating path. Besides, only detection is insufficient to eliminate the impacts of the malicious attack.

4.3 Power System Generation Control

Power system balancing and frequency control contains control actions of different timing granularities. Among the existing control actions summarized in [73], primary and secondary frequency controls are activated most often and also more likely to be targeted by different threat actors.

4.3.1 Primary Control - Governor Action and Load Feature

Most of the generators are equipped with speed governors to stabilize power system frequency after a generation-load imbalance occurs. To put it simply, what the governor does is that it controls the generation output linearly according to the frequency deviation as shown in (4.1). β_i is a parameter called "droop value" of generator i and it takes a negative value. The governor will increase the generation when system frequency is below its nominal value (lack of generation) and vice versa till a new generation-load balance is attained again. pu in (4.1) stands for "per unit", which is a normalized value based on a base power and has no unit [79].

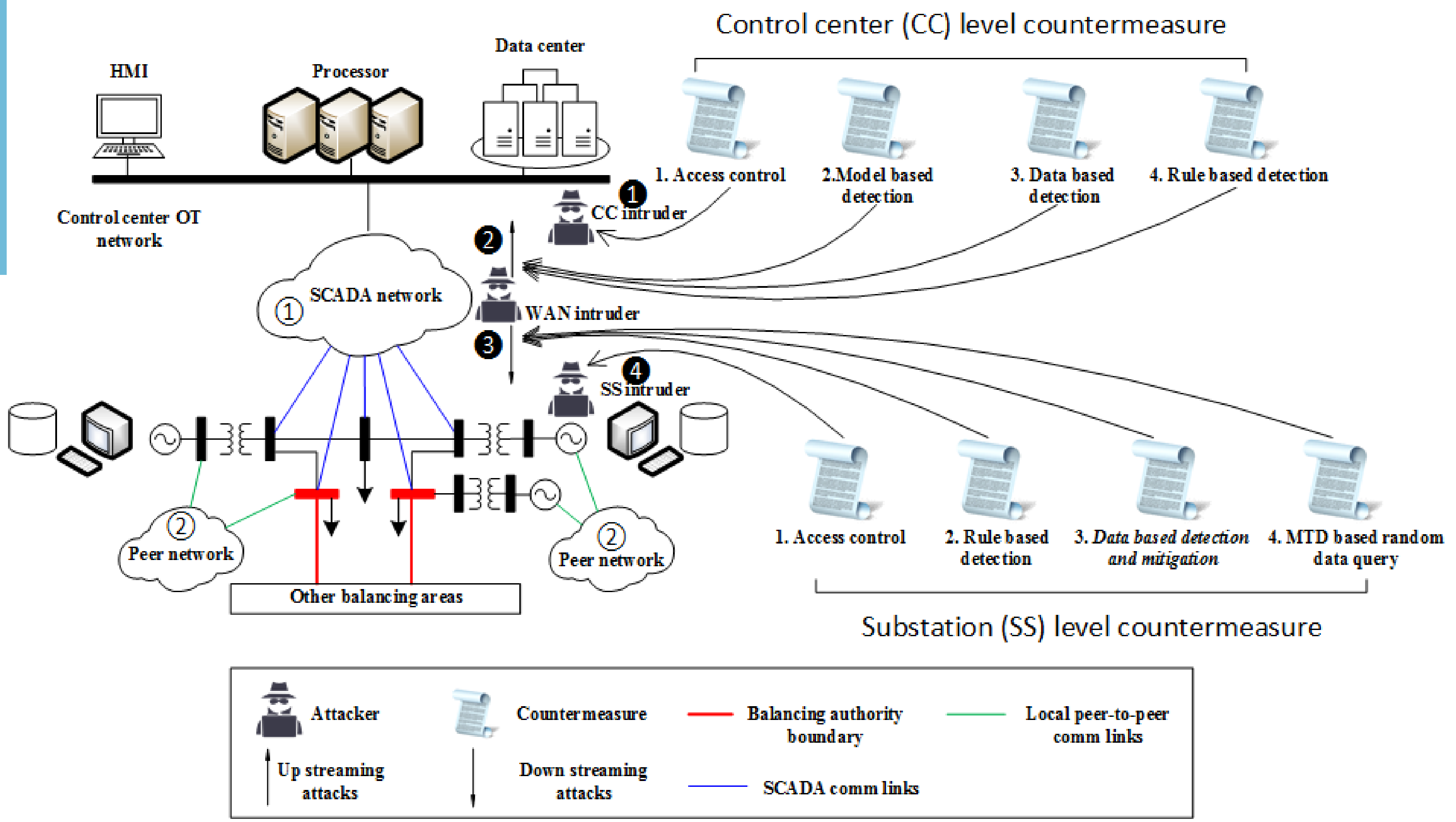


Figure 4.1 Systematic deployment of countermeasures for generation control

$$\Delta P_{i,pu} = \beta_i.\Delta f_{pu} \tag{4.1}$$

Whenever there is a load increase, the system frequency dips as generators provide the excess energy by converting their kinetic energy before the governor action kicks in. Besides, most of the motor loads in the power system will absorb less energy to help the restore the frequency back creating a damping effect. This feature of load is ignored and constant load values are adopted to simplify the problem.

4.3.2 Secondary Control - AGC

The North American power grid is functionally divided into several Balancing Areas (BA) overseen by different control centers. Among other functions, each BA employs the AGC algorithm to maintain load-generation balance within the BA such that the overall grid frequency is maintained close to the nominal value of 60 Hz. Figure 4.2 presents a detailed view of an example AGC implementation at a BA control center. The execution of AGC happens in the following three steps.

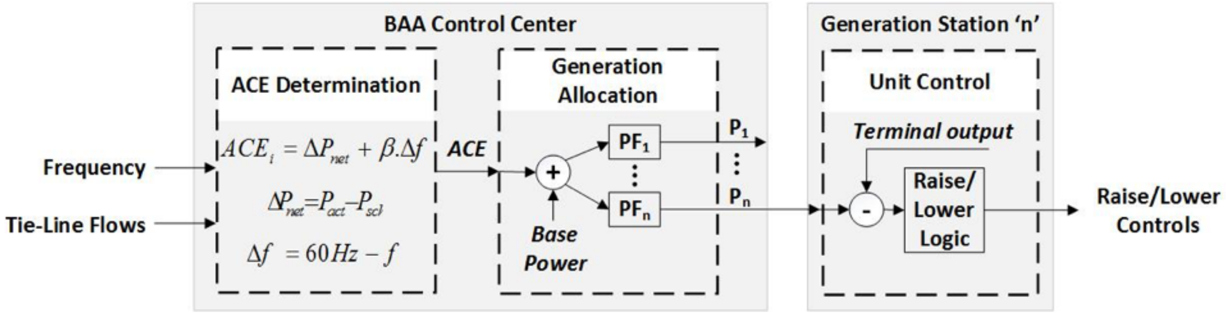


Figure 4.2 AGC overview

Step 1 - ACE determination: The first step involves the calculation of the Area Control Error (ACE) which represents the imbalance between generation and load in the BA. The ACE is calculated using the equations provided in Figure 4.2, where, ΔP_{net} and Δf represent the difference between actual and scheduled, tie-line power flows and frequency, respectively. The actual tie-line power flow and frequency measurements are relayed to the control center from field sensors through SCADA network. The constant β is the frequency bias. The ACE that is calculated using this equation represents the generation correction for the entire BA.

Step 2 - Generation allocation: The generation allocation logic uses participation factors determined through the economic dispatch process to distribute the overall BA generation between the participating generators. The output of this block is a desired power output P_i, for each generator i participating in AGC. The P_i for each generator is directly transmitted via the SCADA to the generation units controller. In other implementation architectures, raise/lower signal for each unit may be calculated at the control center itself and transmitted to the units.

Step 3 - Unit control: The unit control logic compares P_i to the actual unit output to determine corrective controls for the unit. This control is then issued to the governor prime mover as a raise/lower command to directly alter the position of the speed changer, which proportionally alters the power output from the generator.

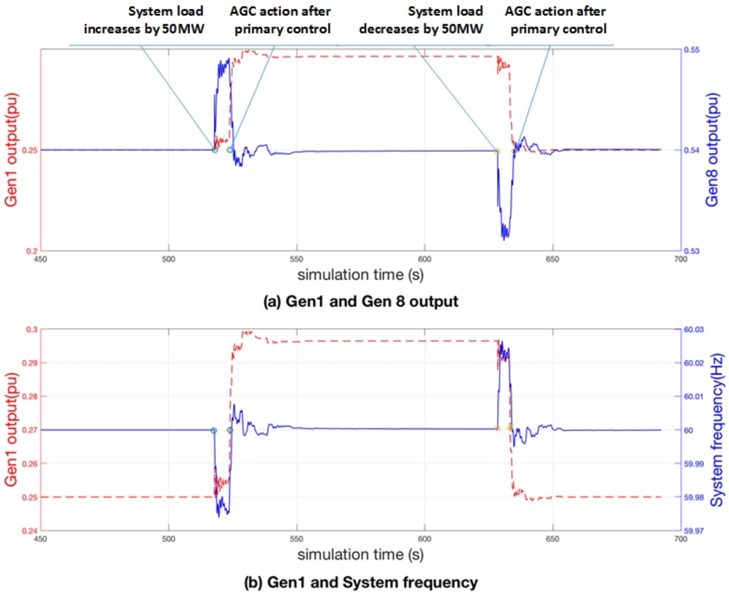
It is to be noted that currently the AGC operates without any protection mechanisms that validate the ACE at the generators. The ACE value that is calculated and sent from the control center to the generators are typically sent over unencrypted DNP3 packets in the SCADA networks. This makes it extremely vulnerable to cyber attacks such as Man-In-the-Middle (MITM) attacks that introduce malicious generation set points to affect the frequency of the power system.

4.4 Generation Attack Detection based on Semi-Supervised Clustering

In this section, we first investigate the power system frequency control characteristics and the specific behavior patterns observed, which is followed by the proposed data processing method and the attack detection strategy based on semi-supervised clustering.

4.4.1 Behavior Patterns Observed in Frequency Control

As implied by their names, primary control responds to frequency disturbance faster than the secondary control. When a generation-load imbalance occurs, primary control will first attempt to



(a) Gen1 and Gen 8 output

(b) Gen1 and System frequency

Figure 4.3 Balancing and frequency control

restore system frequency to 60 Hz but with a small deviation. Then, as a part of secondary control, AGC reduces this deviation in several cycles to bring the frequency to 60 Hz. Figure 4.3 shows how frequency control works after load changes occur in Area 1 of the IEEE 39 bus model (as shown in Figure 4.10), where G1 is participates in AGC while G8 only has the governor function. All generators that have a governor module enabled will help to stabilize the system frequency, and only generators with secondary control enabled will contribute in AGC. From Figure 4.3, we might have already observed the underlying nature of frequency control, such as a non-trivial AGC should follow both the disturbance and the governor's action.

In terms of the measurements' characteristics, the frequency control of a BA could have four general patterns as shown in Figure 4.4. The horizontal axis is the absolute value of the ACE received by a generator and the vertical axis represents the system disturbance level within the sliding window right before the arrival of this ACE. System disturbance can be quantitatively measured by the variance of system frequency and generator outputs, and it reflects whether the current operation state has drifted away from the normal state. Most of the cases, system operation state will fall into region I, where the system is operating in the normal stable state and only a trivial generation adjustment is necessary. Region II, in contrast, contains scenarios where the system does not experience much disturbance, but a large ACE value is received trying to change generation setting points. One ACE observed in this region is either miscalculated or maliciously injected by an adversary. Region III and IV each contains several possible scenarios covering both no attacks and attacks and cannot be distinguished from each other by simply checking the raw data. We use the terms "Inside" and "Outside" to indicate whether an event (a load change) happens inside or outside of the BA in consideration. The problem that anomaly detection needs to resolve is to distinguish between the scenarios under attack and normal scenarios.

In general, frequency control patterns differ for every balancing area based on their interconnections with other systems, and will be specific for different disturbances too. Several special cases are given as follows.

- *Case* 1: Independent BA without power exchange with other interconnections, such as ERCOT [73].
- *Case* 2: Interconnected BA where a single event needs to be adjusted.
- *Case* 3: Interconnected BA and multiple events occur sequentially.

Based on the above discussion, it is obvious that the implementation of the anomaly detection varies according to which type of BA is being considered and also what specific types of attacks it should detect. To avoid over-complicating the problem, the detection methodology proposed provides a general way to detect adversary activities by utilizing generator behavioral conformity metrics and clustering techniques.

4.4.2 Generator Behavior and Conformity Metrics

Generators that are synchronized in the same BA normally respond to the same disturbance such as load change and generation rejection in a similar way, either all ramp up or all ramp down. Figure 4.5 shows this intuitively by providing a zoomed-in view of the load increase event that happens around 518s provided in Figure 4.3. It shows the frequency control process within the

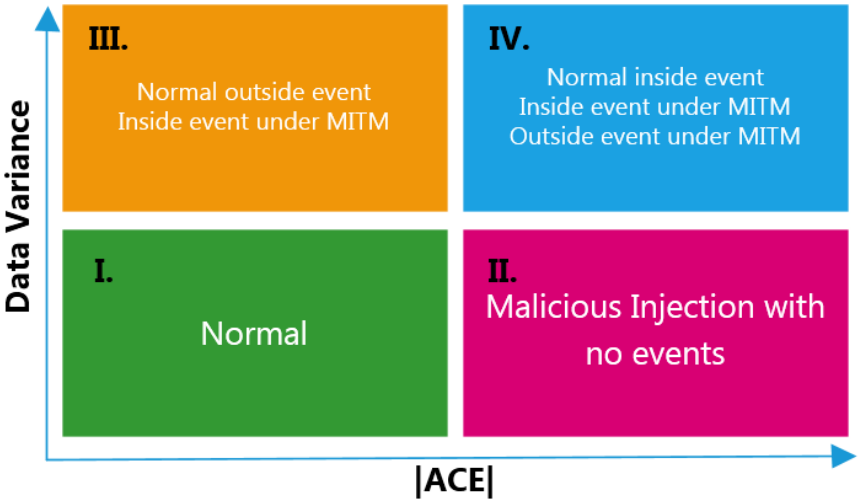
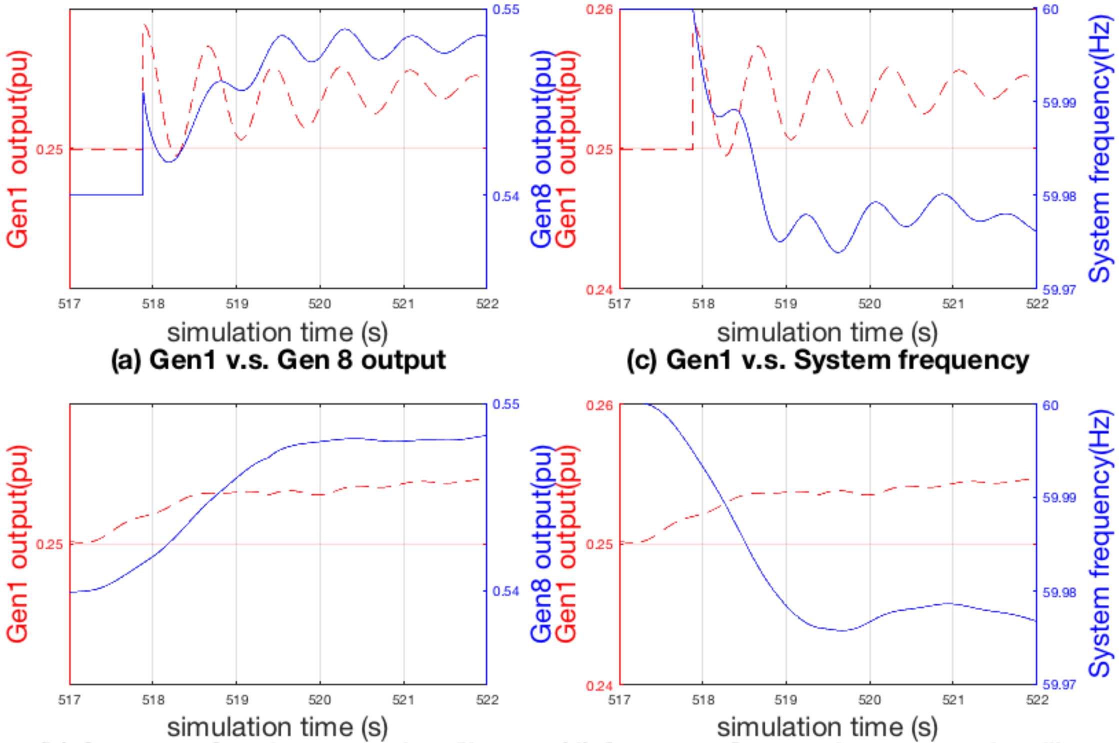


Figure 4.4 Grouping data based on ACE variance

sliding window which contains 1800 simulation steps (step size is 0.003 s) right before the arrival of first post-event ACE. Figure 4.5 (a) and (c) are plotted with the original sample values. Since we do not care much about the excursions in the signal when we check the behavior conformity of generators, we can smooth them out by applying mature filtering techniques. Figure 4.5 (b) and (d) show the same curves after filtering with 3 order Savitzky-Golay filter [80]. We can see that the outputs of two generators and system frequency are highly correlated. This behavior correlation or conformity will be utilized to detect malicious generation controls, and the main idea is to find anomaly where the conformity of generators is lost. To quantitatively measure ”conformity”, 3 metrics are defined. The main rationale behind the 3 metrics is that synchronized generators should adjust their output towards the same direction when facing random load changes. This is true for both primary and secondary generation control phases.

Metric 1: Correlation between two active power outputs of generator i and j, denoted as C_{ij} or that between the power output of a generator i and system frequency, denoted as C_{if}. We denote the discrete active power output of n generators (Gen) during a window with a pre-selected length as matrix $P_{n\times winSize}$, and thus P_{im} is the output of generator i at time point m. Sample mean of generation and system frequency are given by (4.2) and (4.3), and the correlation metric for two generator outputs is denoted as C_{ij} and the correlation metric for a generator output and



(a) Gen1 v.s. Gen 8 output

(c) Gen1 v.s. System frequency

(b) Gen1 v.s. Gen 8 output after filter

(d) Gen1 v.s. System frequency after filter

Figure 4.5 Generation control just before AGC operation

the system frequency C_{if} is denoted as (4.4) and (4.5) respectively. Although the generator power output and system frequency will always contain excursions, there might be constant readings when sensors are malfunctioning or under cyber attack. C_{ij} and C_{if} will be computed as 1.

For the scenario in Figure 4.5, C_{18} is 0.6464 and 0.8642 respectively before and after being filtered, and C_{1f} is -0.4721 and -0.8760. It should be noted that the reverse correlation between a generator output and system frequency will not always hold for Case 3 mentioned in Section 4.4.1.

$$\overline{P_i} = \frac{\sum_{m=1}^{winSize} P_{im}}{winSize} \tag{4.2}$$

$$\overline{f} = \frac{\sum_{m=1}^{winSize} f_m}{winSize} \tag{4.3}$$

$$C_{ij} = \begin{cases} 1, & \text{const } P_i,\, P_j \\ \frac{\sum_{m=1}^{winSize}(P_{im}-\overline{P_i})(P_{jm}-\overline{P_j})}{\sqrt{\sum_{m=1}^{winSize}(P_{im}-\overline{P_i})^2}\sqrt{\sum_{m=1}^{winSize}(P_{jm}-\overline{P_j})^2}}, & otherwise \end{cases} \tag{4.4}$$

$$C_{if} = \begin{cases} 1, & \text{const } P_i,\, f \\ \frac{\sum_{m=1}^{winSize}(P_{im}-\overline{P_i})(f_m-\overline{f})}{\sqrt{\sum_{m=1}^{winSize}(P_{im}-\overline{P_i})^2}\sqrt{\sum_{m=1}^{winSize}(f_m-\overline{f})^2}}, & otherwise \end{cases} \tag{4.5}$$

Metric 2: Stability contribution ratio D_{ij} of generator i and j. After a generation-load imbalance occurs, all generators with the governor function will adjust their output a little to compensate for the discrepancy between generation and load until the frequency becomes stable. Droop value of a governor (β) will determine how much a unit will contribute with given frequency deviation as in (4.1). Equation (4.6) defines the ratio of stability contributions of Gen_i and Gen_j with the droop values β_i and β_j respectively over the length of a sliding window. D_{ij} ideally should be constant, close to 1, under normal operation. This metric is useful when an attacker changes one generation set point after intrusion in a fast and extreme way.

$$D_{ij} = \frac{(G_i[winSize] - G_i[1])\beta_j}{(G_j[winSize] - G_j[1])\beta_i} \tag{4.6}$$

Metric3: Group conformity R_i of a secondary control generator i. This metric measures the relevance of ACE values received by a few secondary generators in a BA. Normally in each AGC operation cycle, the total ACE value calculated for a BA equals the current lack or surplus of generation compared to the scheduled value and would be split among all secondary control generators based on their available capacities. Therefore, ACEs sent out in the same AGC cycle should be approximately proportional to each other. Metric (4.7) is defined for Gen_i based on this feature. Gen_k are neighbors of Gen_i, and ACE values will first be converted to real value from p.u. value by multiplying with the base power P_{base}. We use a very small constant ϵ to avoid Not-A-Number (NaN) being obtained in certain cases.

$$R_i = exp(\sum_k \frac{P_{base}ACE_i + \epsilon}{P_{base}ACE_k + \epsilon}) \tag{4.7}$$

When under malicious generation control, the behavior conformity of a generator will start to show discrepancy, which can be identified by the values of metrics 1-3 defined above.

4.4.3 Semi-supervised Clustering Aided Multi-Class Classification

The detection of malicious generation control is formulated as a semi-supervised multi-class classification problem. Since raw data is collected continually with a temporal sliding window in a real-world scenario, the cost of labeling all instances will be expensive. On the other hand, the conformity of generators' different behaviors is expected to be well separated in the hyperspace. Therefore, instead of trying other complex algorithms, K-means clustering is selected first to cluster the historical data as the detection model for the online classification. A portion of all the historical data instances will be labeled and they will provide the initial cluster centroid locations. Due to the rareness of attack instances, in real applications, real attacks recorded should all get labeled and synthetic data from simulation can be generated for a given system to improve the accuracy of model.

The data are collected with sample windows determined by the arrival of ACEs as shown in Figure 4.6, every pulse in the figure represents one AGC operation signal. The maximum length of a window is the full AGC operation cycle. Data are collected from both local generation unit and other neighbor units, and Savitzky-Golay filter is then applied to these data such that the trend of signals is maintained while the excursions get smoothed out. After the raw data gets preprocessed, it is used to calculate the conformity metrics as discussed earlier. The metric tuple obtained from each window will become one instance in the data mining.

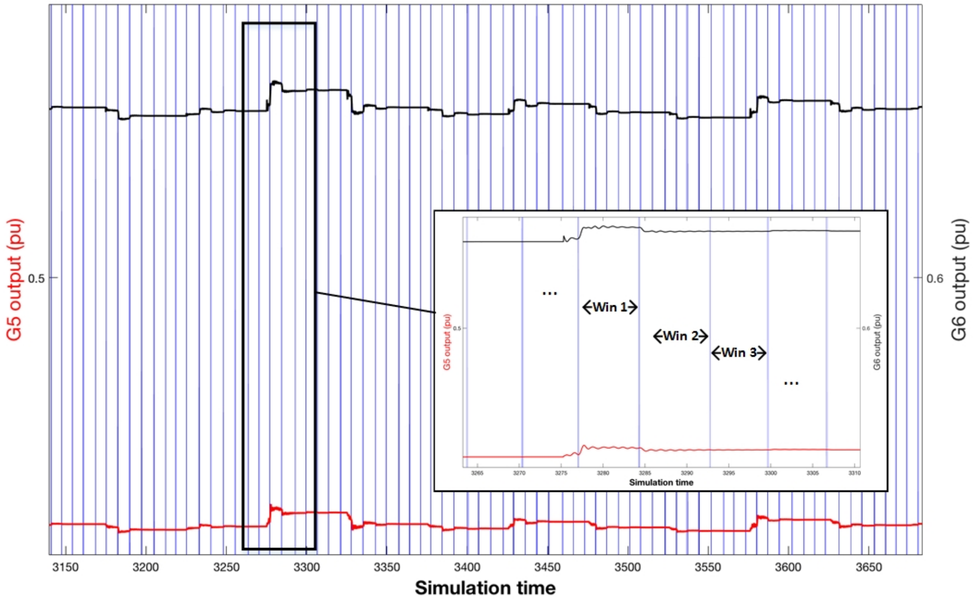


Figure 4.6 Data collection by sample windows

4.4.4 Overall Detection Process

The overall process of the proposed anomaly detection is as depicted in Figure 4.7, in which the solid line links two procedures and the dotted line represents a path of data flow. The main steps of online detection on the left are listed below.

- **Step 1:** When the generator receives a new ACE from AGC controller, retrieve data from the previous sliding window for comparison.

- **Step 2:** Check the variance of the data and if a sample instance falls in region I or II as shown in Figure 4.4, do not send data forward to detection module.

- **Step 3:** Filter out data excursions.

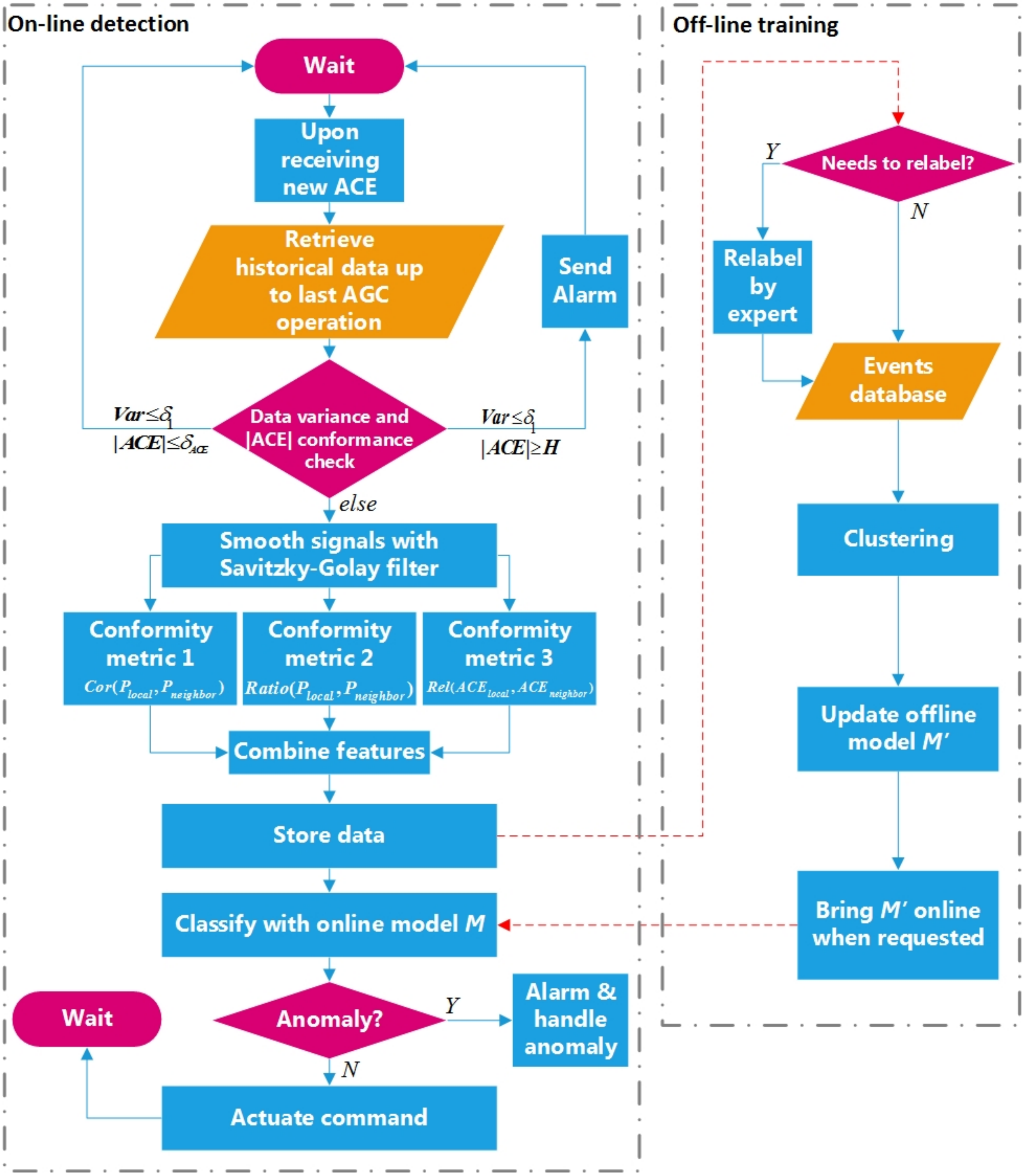


Figure 4.7 Overall flowchart of anomaly detection

- **Step 4:** Perform a dimension reduction on the data and calculate the three conformity metrics based on the processed data.
- **Step 5:** Send conformity metrics to the online model that was built from offline clustering and determine if this instance represents a normal AGC operation.

The offline clustering proceeds as following.

- **Step 1:** Online data gets archived and if it is a novel pattern, it will get relabeled.
- **Step 2:** New offline cluster model M' is obtained through offline training with updated instance set.
- **Step 3:** Online model is updated on request.

4.5 Generation Attack Detection based on Density Sensitive Clustering

In last section, an anomaly detection methodology based on the behavior conformity of a group of generation units is proposed. But through the experimental evaluation [81], a drawback of K-means clustering is observed. That is, K-means clustering does not consider the density information embedded in the data set. Therefore, a density sensitive clustering technologies is applied to solve the same problem [82] as an improvement.

Density based Spatial Clustering of Application with Noise (DBSCAN) [83] was proposed in 1996, which demonstrated excellent capability of extracting the density information out of datasets. Later, Hierarchical DBSCAN (HDBSACAN) [84] was further developed in order to handle the parameter selection issue of DBSCAN. Density based clustering is widely applied in unsupervised learning, and researchers have already tried to use it to process the great amount of PMU data in power system [85].

Clustering is originally proposed as an unsupervised learning technique. It does not require the data instances to be labeled. But in many cases, people would like to provide the clustering process with hints. It is reasonable to apply empirical knowledge to bias the clustering intentionally. Thus, we assume again that certain amount of labeled data instances via post events analysis are available and semi-supervised clustering is adopted in the application of HDBSCAN.

4.5.1 Preliminary of Hierarchical DBSCAN

The objective of density based spatial clustering is to group all instances in a data set $\mathbf{X} = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$ distributed in the hyperspace into clusters based on the regional density. Distance $d(\mathbf{x}_i, \mathbf{x}_j)$ is defined for any two points $\mathbf{x}_i$ and $\mathbf{x}_j$. Every data point either becomes a core point belonging to a cluster or it becomes a noise point (outlier) after clustering. We adopt the following definitions from [84] which are originally designed in DBSCAN:

Core points - A point $\mathbf{x}_i$ is a core point w.r.t. ϵ and m_{pts} if $|\mathbf{N}_\epsilon(\mathbf{x}_i)| \geq m_{pts}$, where $\mathbf{N}_\epsilon(\mathbf{x}_i) = \{\mathbf{x} \in \mathbf{X} | d(\mathbf{x}, \mathbf{x}_i) \leq \epsilon\}$ and $|.|$ denotes cardinality. If a point is not a core point, it is labeled as noise.

ϵ **reachable** - Two core points $\mathbf{x}_i$ and $\mathbf{x}_j$ are ϵ reachable w.r.t. ϵ and m_{pts} if $\mathbf{x}_i \in \mathbf{N}_\epsilon(\mathbf{x}_j)$ and if $\mathbf{x}_j \in \mathbf{N}_\epsilon(\mathbf{x}_i)$.

Density connected - Two core points $\mathbf{x}_i$ and $\mathbf{x}_j$ are density connected w.r.t. ϵ and m_{pts} if they are directly or transitively ϵ reachable.

Cluster - A cluster $\mathbf{C}$ w.r.t. ϵ and m_{pts} is a non-empty maximal subset of $\mathbf{X}$ such that every pair of objects in $\mathbf{C}$ is density-connected.

In [84], core distance and reachability distance are introduced, which are presented as following.

Core distance - The core distance of an object $\mathbf{x}_p$ in $\mathbf{X}$ w.r.t. m_{pts}, $d_{core}(\mathbf{x}_p)$, is the distance from $\mathbf{x}_p$ to its m_{pts}-nearest neighbor (including $\mathbf{x}_p$).

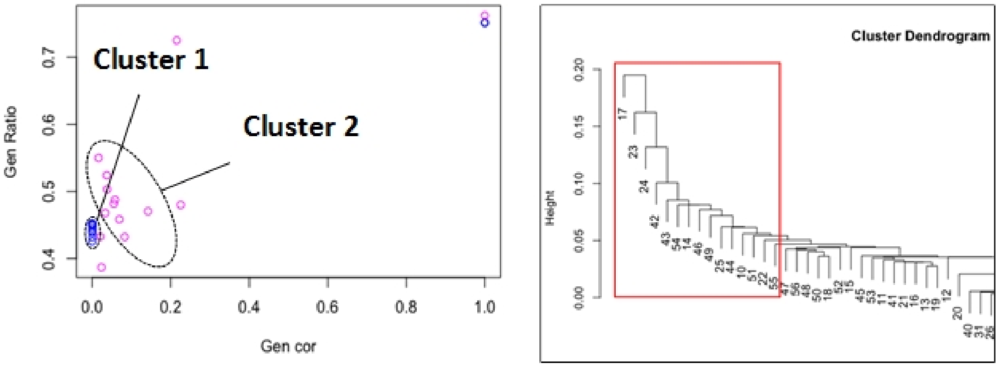
Mutual reachability distance - The mutual reachability distance between two objects $\mathbf{x}_p$ and $\mathbf{x}_q$ in $\mathbf{X}$ w.r.t. m_{pts} is defined as $d_{mreach}(\mathbf{x}_p, \mathbf{x}_q) = max\{d_{core}(\mathbf{x}_p), d_{core}(\mathbf{x}_q), d(\mathbf{x}_p, \mathbf{x}_q)\}$.

HDBSCAN is developed based on DBSCAN and it works better for a data set composed of clusters of different densities and only requires one user input parameter m_{pts}. The way how it works is that it first forms a Minimum Spanning Tree (MST) connecting all data points in the hyperspace. Thus all data points belong to the same cluster at the beginning. An edge between 2 vertex in the MST uses the mutual reachability distance [84] as its weight. Instead of utilizing a specific ϵ value as input, HDBSCAN treats ϵ as a variable. Starting from the complete dataset as a root node, it continuously decreases ϵ value and along the way removes those edges whose weights are greater than ϵ from MST. Finally a hierarchical cluster tree (a divisive dendrogram) will be formed until all the data points are labeled as outliers. Cut the cluster tree at any ϵ level, points that are still connected belong to the same cluster. It is not necessary to cut the tree with a global ϵ value, and therefore, HDBSCAN is able to recognize clusters of different densities. During the construction of the MST, the mutual reachability distance between two data points incorporates the data density information by specifying the minimum number of data points that a core point should possess in its vicinity and therefore, the clustering achieved is sensitive to the density of data points.

4.5.2 Application of HDBSCAN in Divide & Conquer Algorithm

HDBSCAN is by no means perfect. First of all, since the edge weights of the MST is determined by m_{pts}, for certain cases, the global m_{pts} cannot form a MST with satisfactory structure; second, HDBSCAN is not able to distinguish two clusters that are close to each other and have different densities. Figure 4.8 shows one example. From Figure 4.8 (a) we can intuitively see that two clusters exist in the hyperspace, of which cluster 1 is much denser than cluster 2, and the two clusters are closely placed. If we run HDBSCAN on the dataset, Figure 4.8 (b) depicts the dendrogram resulted. It is clear that the splits at the beginning from the root node are not ”true splits” which are supposed to divide a cluster into another two. In fact, those splits throw out a data point of cluster 2 at a time as an outlier. During the whole process, HDBSCAN believes there is only one cluster exist. The key reason is that the two clusters stay too close.

Reference [84] has discussed using HDBSCAN for semi-supervised clustering. It suggests utilize labeled data instances as constraints at the instance level to find the optimal clusters. That is, list all the pairs of points should stay in the same cluster and all the pairs should not, then select clusters with least constraints violation. Considering the number of labeled data can be large too, our algorithm uses entropy, which is strictly defined in information theory, to select the optimal



(a) Illustration of a special case **(b) HDBSCAN dendrogram**

Figure 4.8 Special case when HDBSCAN is not suitable

clusters in terms that the labeled data points with different labels can be separated as much as possible. Entropy of a cluster is calculated as (4.8), in which $p(C_i)$ represents the proportion of class C_i in the labeled sub dataset and $q(C_i) = 1 - p(C_i)$.

$$Entropy = -\sum_{i=1}^{N}[p(C_i)logp(C_i) + q(C_i)logq(C_i)] \tag{4.8}$$

Based on the analysis above, a modified algorithm based on HDBSCAN is suggested as stated in Algorithm 3, which adopts the divide and conquer philosophy. $\mathbf{X}_{N\times p}$ in the pseudo code represents the dataset with N data points and each point has p features, $C_{N\times 1}$ is a vector stores the cluster label for all points, $\{m_{pts}\}$ is a set that contains all candidate m_{pts} used in HDBSCAN and T_{depth}

indicates how deep the algorithm can move to and cut on the dendrogram. This algorithm is able to recursively find the clusters.

```
   Data: $X_{N\times p}$, $C_{N\times 1}$, $\{m_{pts}\}$ and $T_{depth}$
   Result: $C_{N\times 1}$
1  Pick up one element $m$ from $\{m_{pts}\}$;
2  Run HDBSCAN($\mathbf{X}$,$m$) and attain cluster dendrogram $cl$;
3  if Bad cluster tree pattern observed in cl then
4      Divide $\mathbf{X}$ into $X_1$ and $X_2$;
5      call Div_Conq_HDBSCAN() on $X_1$;
6      call Div_Conq_HDBSCAN() on $X_2$;
7  else
8      Find all true splits and cut at depth above $T_{depth}$;
9      Select the cut with min sum of cluster entropy;
10     go to step 1 until $\{m_{pts}\}$ becomes empty;
11 end
12 Modify $C_{N\times 1}$ according to the optimal $cl$ return $C_{N\times 1}$
```

Algorithm 3: Divide and Conquer HDBSCAN

4.6 Peer-Assisted Mitigation of ACE Attacks

A few control center level anomaly detection methodologies have been proposed in the existing literature and they provide solutions to detect false data injection attacks targeting AGC measurements. However, the attacks directly target the ACE values along the forward control path of AGC can not be detected by the anomaly detection algorithm located in the control center. Besides, not any preventive or corrective schemes are proposed to handle the cyber attacks to the best knowledge. Considering such shortcomings, a mitigation solution is proposed in this section which is designed in such a way that it is able to not only detect anomalous ACE commands from the control center, but also mitigate the impacts of attacks by replacing the missing/bad ACE with an estimated value, and hence the AGC control continues without being interrupted. In the context of Figure 4.1, the attacks under investigation is ❸ and the mitigation is placed at the station level.

4.6.1 Overall Mitigation Process

The idea behind the proposed mitigation is that the generating units in one balancing authority are supposed to move together towards the same "direction" every time under the control of AGC and the contribution of every unit in the AGC operation are highly relevant since the AGC is deployed with the economic performance under consideration. Therefore, when the ACE value sent to a generating unit from the control center gets compromised or blocked, if this unit has the capability to retrieve data from its peers, very likely it can still attain the current situational awareness and carry out the generation control actions in need.

Figure 4.9 provides the schematic of the proposed mitigation. The fundamental assumption underlying is that the generating units that participate AGC are able to regularly exchange the ACE information among themselves. One generating unit is supposed to be able to communicate to at least a few peers.

The mitigation relies on the information exchanged among peer generating units. Peer generating units that stay close geographically can be interconnected to aid the ACE info exchange. Every unit will randomly select some of its interconnected peers and request these peers' ACEs in every AGC cycle besides its own from the control center (Step 0). This random selection of peers is achieved based on a "reputation" metric of every peer, which serves as a moving target and makes it harder for the adversary to scan through or compromise the entire community. The peer reputation keeps updating based on the performance of each individual peer in the clustering and the algorithm will be provided latter. It's worth noting that peer ACEs will be shared via either the network ① or ② as shown in Figure 4.1. Anomaly detection tactics such as what's proposed in the previous subsections can be applied first on the ACE values. In case that no anomaly of the ACE received is detected, all the ACE values can be stored in the database, which will be utilized to obtain the relationship between the ACEs from two peers and this is done via linear regression

(Step1). When anomaly does exist, the generating units will resort to the main mitigation strategy. Step 2 to 6 are the major actions involved in the mitigation. The fundamental idea of step 2 is to use the linear regression models obtained in step 1 to calculate inferred values for the suspicious ACE. Then a one dimensional density based clustering of all the inferred ACEs is conducted in step 3. With the majority cluster, ACE estimation will be performed (step 4) and the reputation of peers is updated (step 5). Usually the majority cluster of inferred ACEs will be treated as intact values considering that the attacker is not able to compromise most of the peers. After the ACE estimation, the estimated ACE value will replace the suspicious value received from the control center. Besides, each generating unit has the capability to report its low reputation peers to control center so that the most commonly reported units can be temporarily excluded from the AGC (step 6). Subsections following will elaborate the main mitigation steps.

4.6.2 Random Peer Data Query as Moving Target Defense

The telemetry network configuration in SCADA and WAMS is often static. The statically configured communication set-up is vulnerable when facing malicious reconnaissance, especially when not many redundant end nodes are included in the design. This makes it easy for the adversary to find out what is the data sources and sinks for a certain critical application such as WAMPAC. For instance, the control center gets the system frequency from a transducer remotely and this transducer is the only available data source. The adversary could try to block the communication channel between the transducer and the control center to disable the function or he can manipulate the transducer readings to confuse functions at control center. The attacker can achieve his goal without much efforts since his targets are well scoped. One natural way to counteract this would be to increase the data redundancy and also add randomness to the implementation of function.

Adding to data redundancy is similar to the adoption of back-up workstations against natural failures. If the control center reads the same system frequency from 10 transducers, it becomes much harder for the attacker to cause the same impacts as in the previous example. He might be able to compromise one or two readings, but it is useless as a trial either to block the reading

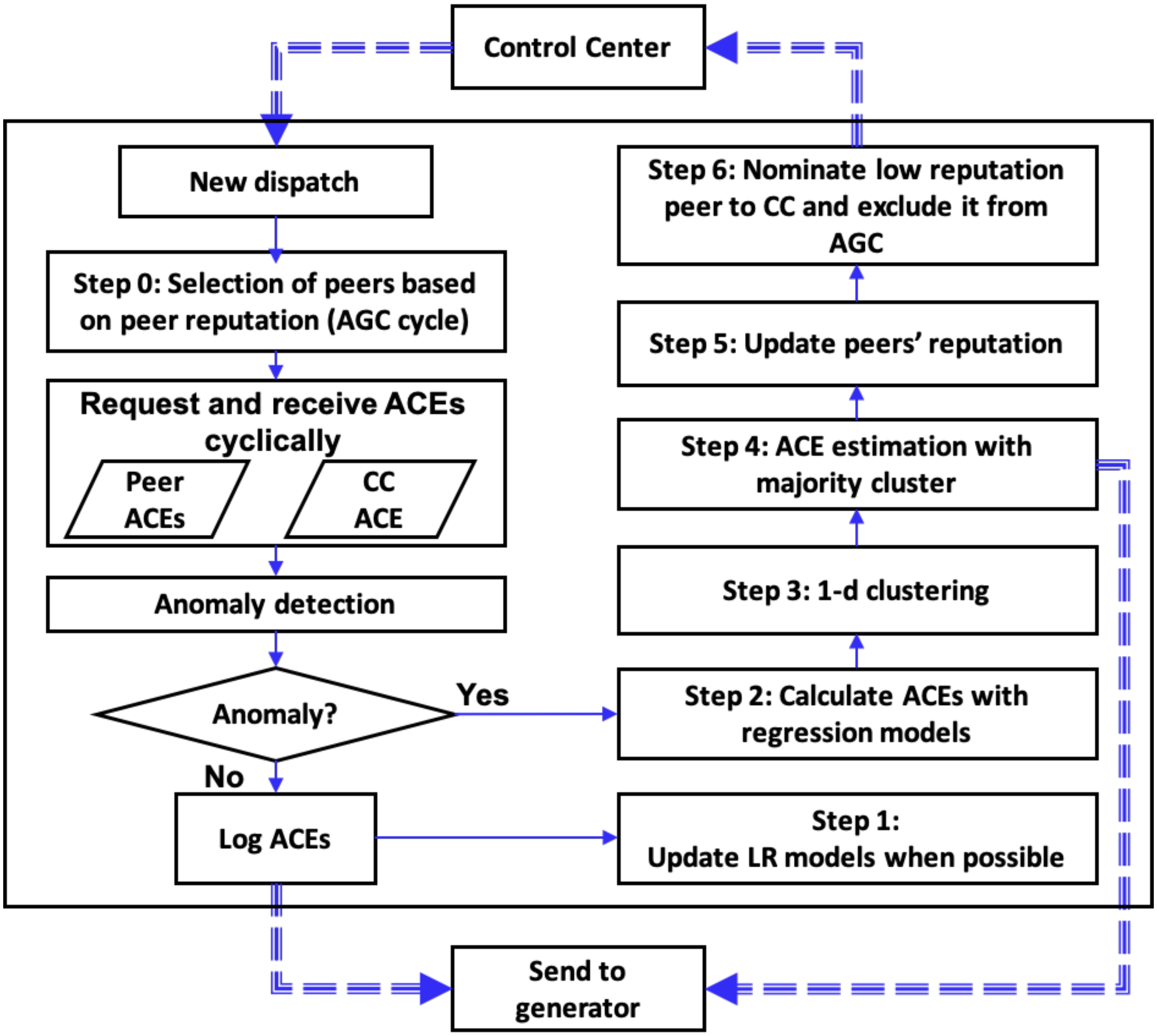

Control Center
New dispatch
Step 0: Selection of peers based on peer reputation (AGC cycle)
Request and receive ACEs cyclically
Peer ACEs
CC ACE
Anomaly detection
Anomaly?
Yes
No
Log ACEs
Step 6: Nominate low reputation peer to CC and exclude it from AGC
Step 5: Update peers' reputation
Step 4: ACE estimation with majority cluster
Step 3: 1-d clustering
Step 2: Calculate ACEs with regression models
Step 1: Update LR models when possible
Send to generator

Figure 4.9 Overall AGC attack mitigation

from control center or, with a majority vote scheme in place, to manipulate some of the readings. Though lifting the level of data redundancy enhances the function reliability, it will also highly increase the communication overhead for the attack free operation state. If every measurement requires data from all accessible sources, the amount of data needs to be exchanged will explode, the complexity of data processing will soar and the efficiency of function will decrease.

Thus, a random data query mechanism is proposed in this section based on the philosophy of moving target defense. In this way, the data and information exchange among agents will not increase very much and in the meanwhile it confuses the attackers by rendering the communication among end nodes “moving”. The foundation of this mechanism is that out of all the redundant and accessible data sources, data query is only sent out to a small part of the sources randomly. In this way, the efficiency and efficacy of the function can be guaranteed, and the cyber security will be increased as well. The random selection of data sources is implemented by a reputation based scheme. That is, every generating unit maintains a reputation vector R that records the reputation of every peer that this unit can communicate with. Each element in R represents the probability that one corresponding peer gets selected during data query and the the elements in R always sum up to 1.

```
   **Data:** Reputation vector $R_{m\times 1}$, clustering label vector $C_{p\times 1}$, indicator vector $Ind_{cl\times 1}$
   **Result:** Updated reputation vector $R$
1  $S_1 = 0$ $S_2 = 0$;
2  for $i \leftarrow 1$ to $m$ do
3      if peer $i$ is not selected during ACE request then
4          Keep $R[i]$ as is;
5      else
6          $S_1 = S_1 + R[i]$;
7          $R[i] = R[i] * exp(Ind(C[i])\frac{cnt(C[i])}{p})$;
8          $S_2 = S_2 + R[i]$;
9      end
10 end
11 foreach $i$ in selected peer do
12     $R[i] = R[i] * \frac{S_1}{S_2}$
13 end
```

Algorithm 4: Peer reputation update

Element values of R should be updated cyclically according to the performance of each peer during clustering. Algorithm 4 is proposed to achieve this goal. Line 7 in this algorithm shows that the reputation of a peer is updated according to the size of the cluster to which the ACE value inferred from this peer's information belongs to. p is the number of all the selected peers. The $Ind(C[i])$ is an indicator function takes values either 1 or -1 and the vector Ind takes the length as the number of clusters cl. When it takes value 1, the underlying principle of the reputation update is that the reputation of the majority should be boosted and that of outliers should be decreased. But it is also well-known that majority is not necessarily correct. If any peer out of a cluster, which could be the majority, has a concrete sign of being compromised, the $Ind(C[i])$ is turned from 1 to -1 to lower the reputation of all peers belong to cluster $C[i]$. If the data from a generating unit is manipulated by the adversary, after a few cycles of AGC, it will have very low reputation from others' perspective and in such a way it is very unlikely to be selected in the data query any more.

4.6.3 ACE Computation and Estimation based on Peer Data

Every Economic Dispatch (ED) cycle (roughly 5 mins), the load reference set point of every generator is calculated in the control center to minimize the production cost and is assigned to corresponding machine. A simple ED model is given by (4.9), assuming N generators are involved in the dispatch. P_i and $F_i(P_i)$ are the active power output and cost function of generator i respectively. C in (4.9) is the total load of the system. The necessary condition that a set of active power output $P_1, ..., P_i, ..., P_N$ to be the most economic dispatch is that the incremental cost of each machine equals the Lagrangian multiplier of the balance constraint as shown in (4.10). This specifies the base generation points within the economic dispatch cycle.

With a dispatch cycle, AGC operates every 4 seconds. Generating units in the same balancing authority will share the overall area control error based on their participation factor (PF) when generation-demand imbalances are observed. The process is coordinated by the AGC function running at the control center. (4.10) should still be satisfied after every AGC control and therefore, the area control error will be split among the generators proportionally according to (4.11) [86].

$$\begin{aligned} \min_{P_i}. \quad & \sum_{i=1}^{N} F_i(P_i) \\ \text{s.t} \quad & \sum_{i=1}^{N} P_i = C \end{aligned} \tag{4.9}$$

$$F_i(P_i)^{'} = \lambda, i = 1, 2, ..., N \tag{4.10}$$

$$\frac{\Delta P_i}{\Delta P_{total}} = \frac{\frac{1}{F_i^{''}(P_i)}}{\sum_i^N \frac{1}{F_i^{''}(P_i)}} \tag{4.11}$$

The above analysis indicates that the ACE control sent to different generators should be linearly correlated and this can be leveraged during mitigation when AGC control signals sent to a generating unit is compromised. In the mitigation strategy proposed, linear regression is adopted to find the relation model between the ACE values of any two generating units and the model obtained is able to infer one unit's ACE value according to the ACE of the other unit. Denote the ACE value that generating unit i receives during AGC cycle k as $a_i[k]$, and the ACE value that its peer j receives as $a_j[k]$. Then given two ACE value sequences, there are two ways to perform the linear regression. Either to take a_i as independent variable and a_j as dependent variable or vice versa. With the first option, unit i can have a sequence of linear models that can be denoted as

$$a_i = \beta_j a_j + \epsilon_j \quad j \in \{Peer(i)\} \tag{4.12}$$

ϵ_j in (4.12) represents the ACE data transmission error following Gaussian distribution. (4.12) can also be rewritten in matrix form as following

$$\begin{bmatrix} a_i \\ a_i \\ \cdot \\ \cdot \\ \cdot \\ a_i \end{bmatrix} = \begin{bmatrix} \beta_1 & & & & & \\ & \beta_2 & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & \beta_p \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_p \end{bmatrix} \tag{4.13}$$

In (4.13), it's assumed that the generating unit i has selected p peers in the mitigation. Similarly, if the second way to run the linear regression is selected, the models that unit i obtain after the LR are as given in (4.14) or (4.15). By utilizing the least square estimation, a_i can be estimated based on a set of a_j. The estimated ACE value $\hat{a}_i$ derived from (4.13) is $\hat{a}_i = \frac{\sum_j \beta_j a_j}{p}$ and that derived from (4.15) is $\hat{a}_i = \frac{\sum_j \beta_j a_j}{\sum_j \beta_j^2}$

$$a_j = \beta_j a_i + \epsilon_j \quad j \in \{Peer(i)\} \tag{4.14}$$

$$\begin{bmatrix} a_1 \\ a_2 \\ . \\ . \\ . \\ a_p \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_p \end{bmatrix} .a_i + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ . \\ . \\ . \\ \epsilon_p \end{bmatrix} \tag{4.15}$$

4.6.4 1-d DBSCAN based Corrupted ACE Replacement

Upon the receiving of the new ACEs from selected peers, the linear regression models obtained can be used to compute a group of inferred ACE values. Out of those peer ACE values, some might be compromised during transmission too. The 1-d clustering is deployed to exclude such values and the rationale behind is that most of the peers should stay intact. Therefore, the ACE estimation will be carried out based on the majority cluster.

4.7 Evaluation Results

4.7.1 Experiment Set-up

IEEE 39-bus system is selected as the study case in this section too. For the evaluation of the detection algorithm, the overall system is divided into 3 BAs as shown in Figure 4.10 and each BA has its own AGC function operating once every 2000 simulation steps (around 6s). Table 4.1 provides the information about the ten generators in this system. Real-time simulations are carried out with the ePHASORSim solver in RT-Lab in order to generate all the synthetic data for various scenarios. The performance of proposed methodology has been tested by focusing on Area 2 in the 39-bus system. This area contains 4 generation units, among which G4 and G7 are the primary control generators and G5 and G6 are both secondary control generators being involved in AGC. From Table 4.1, we can notice that G5 and G6 have the identical governors and same droop values, therefore, their behavior is expected to be highly correlated. For the evaluation of the mitigation strategy, the whole 39-bus system is perceived as a single BA and all the ten generators are configured to participate in the AGC.

Table 4.1 Generation information

	Gen type	Gov type	β	Generation(p.u.)
G1	GENROU	GAST	-0.086	0.25
G2	GENROU	none	none	0.5729
G3	GENSAL	HYGOV	-0.0395	0.65
G4	GENSAL	HYGOV	-0.0395	0.632
G5	GENROU	GAST	-0.046	0.508
G6	GENROU	GAST	-0.046	0.65
G7	GENSAL	HYGOV	-0.0395	0.56
G8	GENROU	GAST	0.046	0.54
G9	GENROU	none	none	0.83
G10	GENROU	none	none	1.012

4.7.2 Anomaly Detection Evaluation

To include common daily load variations in the synthetic data for detection evaluation, a typical day's ACE records from MISO [72] which contains BA ACEs updated every 30 seconds are incorporated in the simulation. Besides, the window size is selected as 1600 simulation steps (around 4.8 seconds) to determine data instances. More detailed generator characterization is necessary to determine the window size and will be further investigated in the future.

Six scenarios have been considered to collect the synthetic data: (1) normal inside load changes (nor in), (2) normal outside load changes (nor out), (3) ACE flip attack together with inside events (flip), (4) ACE constant attack together with inside events (const), (5) ramp attack targeting a compromised generator (ramp), (6) switching attack targeting a compromised generator (switching). Data collection includes 3 stages. First, one simulation is executed for each scenario separately to collect labeled training instances; second, another single simulation run containing mixed types of scenarios is carried out to collect more unlabeled training data; third, repeating stage one with

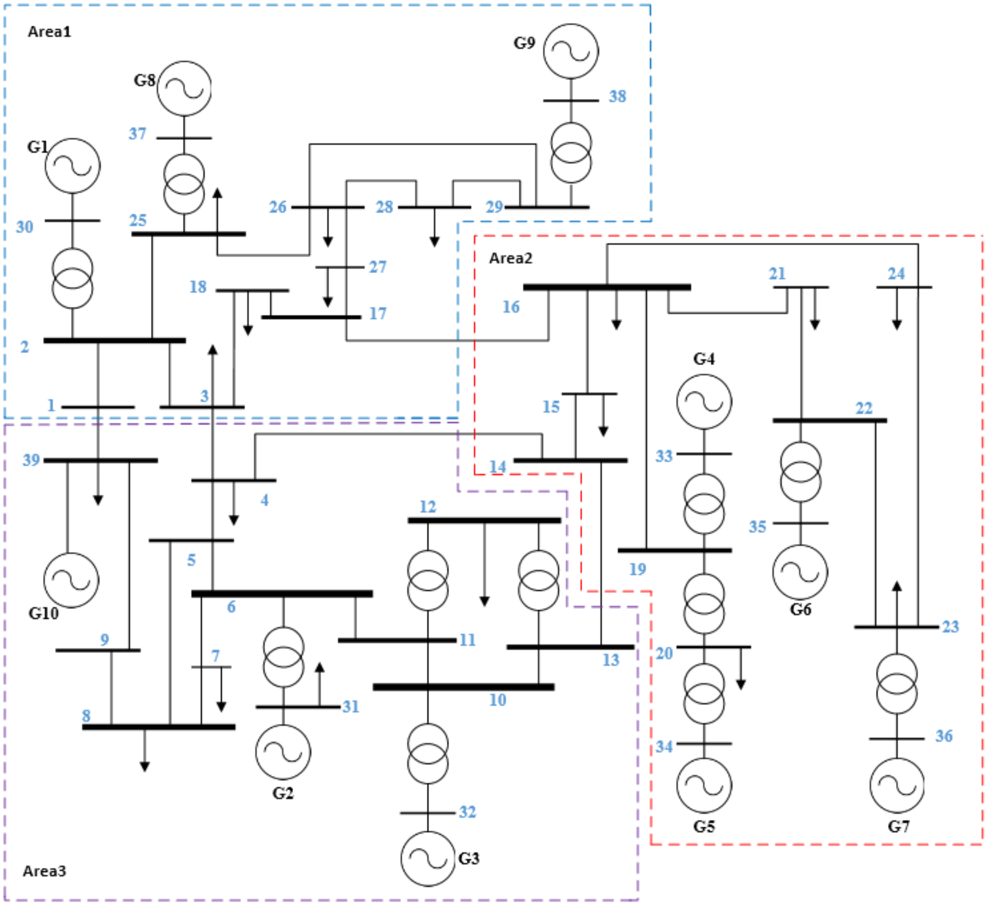
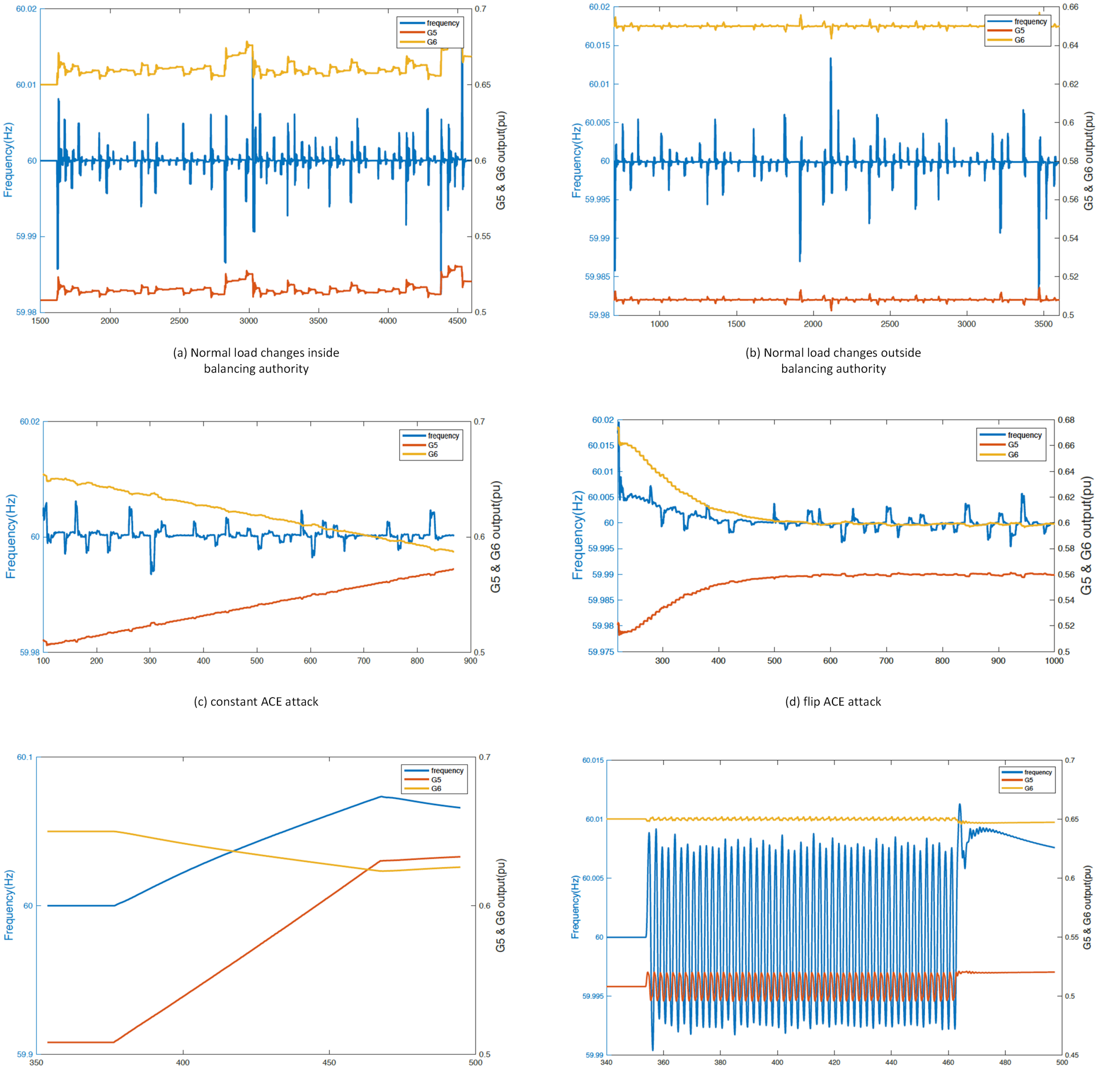


Figure 4.10 IEEE 39 bus model divided into 3 BAs



(a) Normal load changes inside balancing authority

(b) Normal load changes outside balancing authority

(c) constant ACE attack

(d) flip ACE attack

(e) ramp attack

(f) switching attack

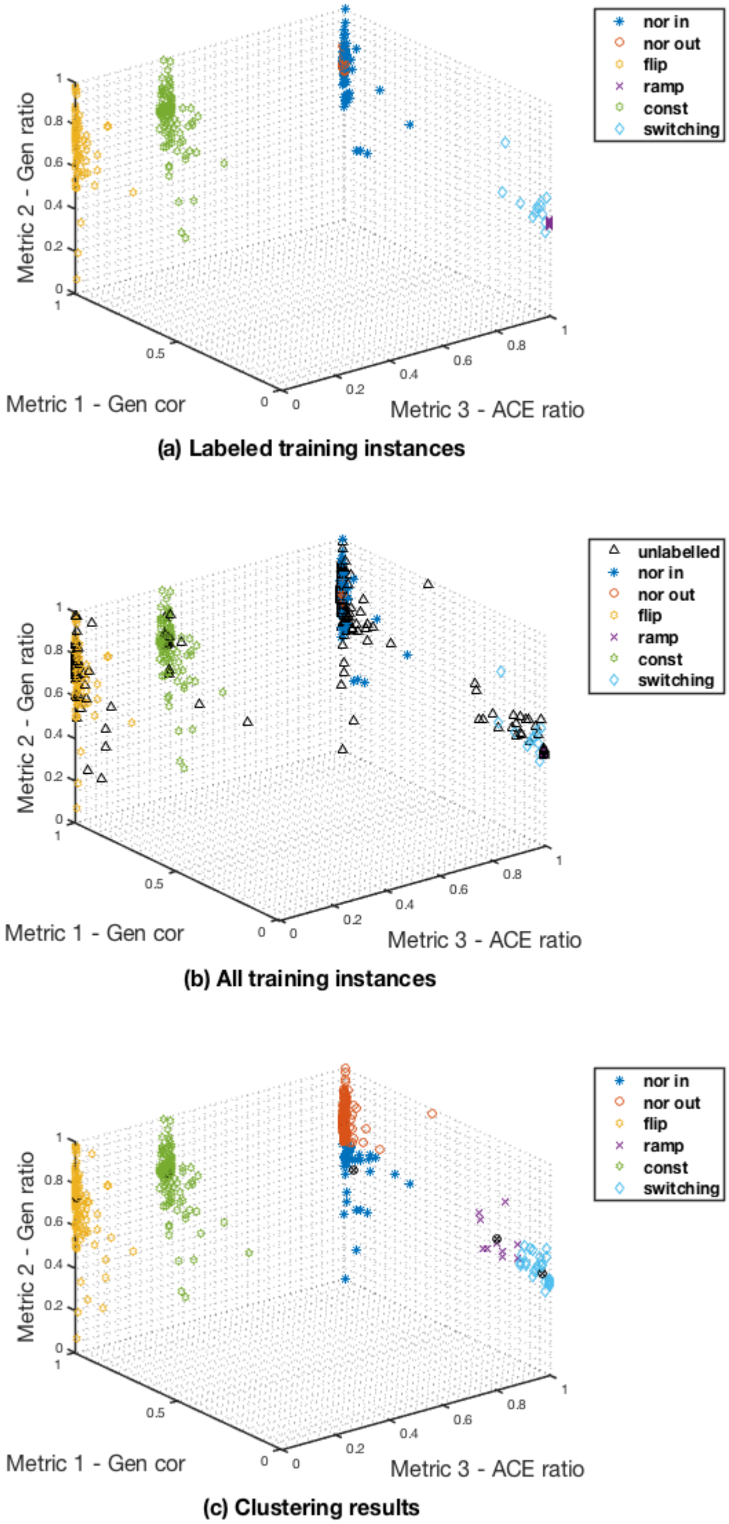
Figure 4.11 Six training scenarios

different simulation parameters and collecting data as test instances. For all scenarios with attacks, G5 on bus 34 will be chosen as the target, and G6 would stay uncompromised.

Figure 4.11 shows the simulation results under each scenario. Scenario (1) shows less excursion than (2) although all the events are exactly the same in terms of load change amount and time. It makes sense because inside events will result in both governor action and AGC operation, but the outside events ideally should only activate the governor. (3) is a scenario where the adversary will flip every real ACE and then forward to the generator, and for (4), a false constant ACE is injected each time. Both scenarios (3) and (4) are data integrity attacks that target ACE values. In these two scenarios, the internal load level has been changed but ACEs are not really fed into the generator since what we need is merely the data before each AGC operation. (5) and (6) are two scenarios with malicious generation manipulation via intrusion. Ramp attack keeps driving the generation level up whereas the switching attack causes oscillation in the power system.

The number of training and test instances are summarized in Table 4.2. Data instances are plotted in Figure 4.12 (a) - (c) with scaled conformity metrics, in which (c) shows the final clusters obtained from K-means clustering with centroids as black cross inside a circle. Test confusion matrix with this clustering model is provided in Table 4.4.

From the test results, we can see that the k-means based methodology performs well in detection of data integrity attacks on ACE values (flip attack and scaling attack), the false negative and false positive rates are both 0. However, based on the confusion matrix and Figure 4.12, it can be seen that the 3 metrics defined are not sufficient to separate the two normal scenarios or the two attack scenarios after intrusion for the ramp and switching attacks. For the two normal scenarios (nor in and nor out), it does not matter much as long as most of normal instances can be separated from the abnormal ones. The de-facto false positive rate, if we simplify the problem into a binary classification, is as low as 1.04% (equal to 1/(73+23) based on the scenario distribution). As for the ramp and switching attack scenarios, when malicious generation manipulation happens, no matter how the adversary changes the generation set point of the target, other generators that are not under attack will work against the targeted generator. This is why the C_{56} is negative for both ramp and switching attacks (close to 0 after being scaled). If the two patterns are allowed to be merged as one, the fake diagonal highlighted with light purple of the confusion matrix shows good detection accuracy.



(a) Labeled training instances

(b) All training instances

(c) Clustering results

Figure 4.12 Scenario distribution and clustering results with K-means clustering

Table 4.2 Instances summary

	Training		Test
	Labeled	Unlabeled	
nor IN	102	-	73
nor Out	91	-	23
flip attack	279	-	13
const attack	275	-	20
ramp attack	12	-	12
switching attack	14	-	16
Total	773	470	157

Table 4.3 Centroids of clusters

	C_{56}	D_{56}	R_{56}
nor in	0.9473	0.5379	1.0000
nor out	0.9946	0.7510	1.0000
flip attack	0.9912	0.7386	0.0000
const attack	0.9882	0.7388	0.3394
ramp attack	0.2536	0.5296	1.0000
switching attack	0.0334	0.4679	1.0000

From Figure 4.12 (a), it can be seen that the labeled switching attack data and ramp attack data have different density features that are not well-captured by the K-means clustering, and this is the reason why HDBSCAN based algorithm is applied. We can also notice that both normal scenarios highly overlap with each other and there is no practical needs to distinguish them in anomaly detection, thus they are merged as the same class during application of HDBSCAN. Figure 4.13 illustrates the process of the first recursion of the clustering. $\{m_{pts}\}$ is selected as $\{10, 11, 12\}$ and the cut depth as 4. With the dendrogram obtained with $m_{pts} = 10$, four different cuts are carried out with corresponding ϵ values. Those cuts are approximately marked on Figure 4.13. The cluster entropies resulted from each cut are also annotated on the left of Figure 4.13. Entropy tuples are corresponding to each cut from top to bottom and the entropy values in a tuple corresponds to

Table 4.4 Test confusion matrix with K-means clustering

	Prediction (%)					
	nor in	nor out	flip	const	ramp	switching
nor in	31.51	67.12	0	1.37	0	0
nor out	0	100	0	0	0	0
flip	0	0	100	0	0	0
const	0	0	0	100	0	0
ramp	0	0	0	0	0	100
switching	0	0	0	0	6.25	93.75

clusters from left to right. We need to pay attention that the right cluster of the second cut has the entropy as 0.82, but after a further split, the left child cluster has a much larger entropy as 1.35. That's due to the fact that the instance number of its left child cluster is much smaller than that of the right child cluster. In the end of this layer of recursion, it selects the 4th cut with $m_{pts} = 10$ and cut value $\epsilon = 0.31$. Since the 3rd cluster from the left (which contains all ramp and switching attack instances) still has large entropy, it will go to the second layer of recursion.

With the proposed clustering methodology, switching and ramp attack clusters are better separated, and this could be observed from Figure 4.14. As a bonus, HDBSCAN also points out the potential outliers existing in the data set, which should be further investigated by experts. For the on-line detection, KNN is applied to the test dataset and the results are summarized in Table 4.5. Compared with Table 4.4, the overall misclassification rate is deduced except the misclassification of switching attack to ramp attack.

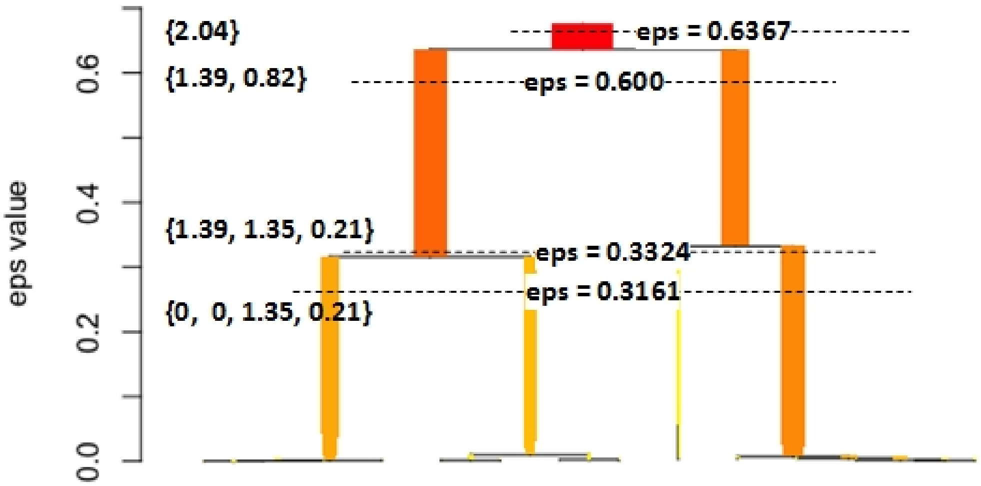


Figure 4.13 Dendrogram of HDBSCAN

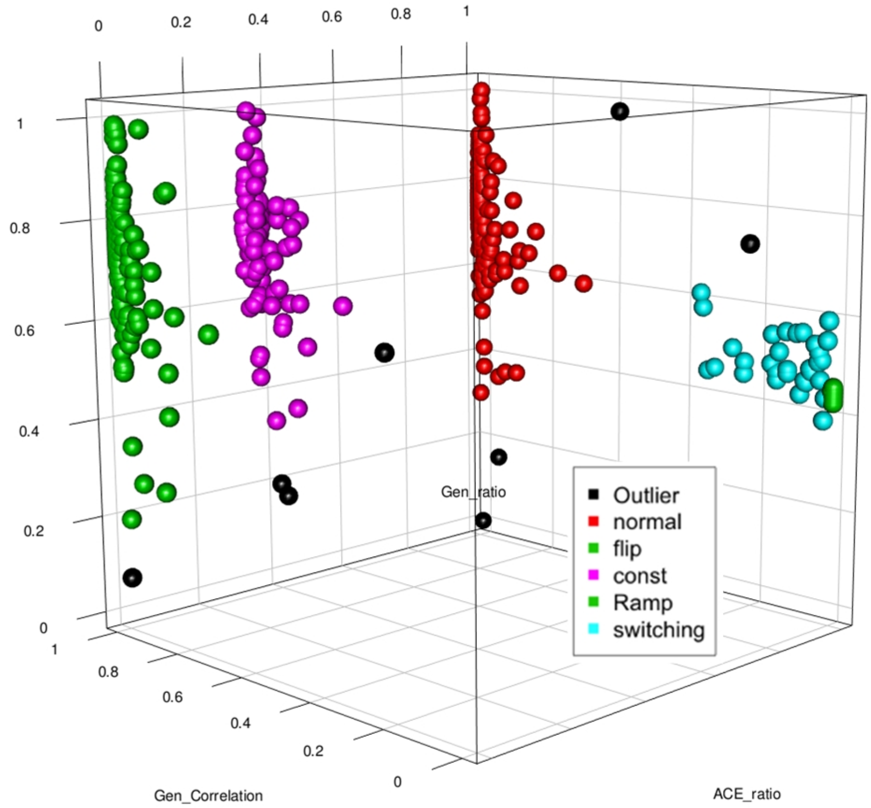


Figure 4.14 Clustering results with HDBSCAN

Table 4.5 Test confusion matrix of proposed clustering methodology

	Prediction (%)				
	normal	flip	const	ramp	switching
normal	100	0	0	0	0
flip	0	100	0	0	0
const	0	0	100	0	0
ramp	0	0	0	100	0
switching	0	0	0	18.75	81.25

4.7.3 AGC Attacks Mitigation Evaluation

The IEEE 39 bus model is treated as one balancing authority in the evaluation of the peer-assisted AGC attack mitigation. All the 10 generators in the system are configured to participate in the AGC operation. The evaluation focuses on the reputation convergence of different generating units when some of the units' ACEs are under cyber attacks.

Figure 4.15 shows an example of the reputation update containing the first 2 rounds of mitigation, i.e. the first two AGC cycles, when unit 1 is under cyber attack. Reputation vectors before peer selection, selected peers, and 1-d clustering for all the 10 generating units are provided. Figure 4.15 (a) shows the reputation vectors of all units at the beginning. In the first round, unit 1 is randomly selected as the peer by unit 4, 5, 6, 7, 8 and 9. But the estimated ACE valued based on unit 1's info is labeled as an outlier in the following 1-d clustering, as depicted in (c). In the second round, all the units that had selected unit 1 in the first round does not select it again (see (e)) due the reputation of unit 1 has been decreased (see (d)). But for those who does not select unit 0 in the first round, unit 0 still gets a good chance to be selected in the second round (see (f)). If such procedures continue, the generating unit that is under attack will have low reputation from the perspective of all the rest units, given that less than half of the units are attacked.

To further analyze the reputation vector evolution under various scenarios and how the proposed strategy works as an effective mitigation, five different attack scenarios are simulated and the results are provided from Figure 4.16 to Figure 4.20.

1. **Scenario 1:** The communication path between the first generating unit and the control center is compromised after the linear regression takes place. The attacker changes the ACE values to unit 1 to a constant value, which is made to be zero in the simulation. From Figure 4.16 we can tell that all the generating units but unit 1 recognizes unit 1 as the low reputation unit. When it comes to unit 1, since all the peer candidates are intact from any attacks, thus the reputation of all the peers does not get changed. In this scenario, the impact of the attack can be mitigated by ACE estimation being carried out at unit 1.

2. **Scenario 2:** The ACEs sent to unit 2 and 3 from the control center are continuously flipped after the linear regression models are obtained. Similarly to scenario 1, the units under attack are recognized by all the generating units and attack impacts are mitigated via ACE estimation. See Figure 4.17.

3. **Scenario 3:** ACEs sent to unit 2 and 3 from the control center are manipulated to be 5 MW continuously after the linear regression. The analysis is the same as that for scenario 2. See Figure 4.18.

4. **Scenario 4:** Constant ACE attacks are launched at unit 1, 3 and 5. But this time, the attacks take place before the generating units obtain the linear regression models, which means that some of the LR models stored at different generating units are incorrect. However, from Figure 4.19 we can see that since less than half of the randomly selected peers are compromised, the units under attack are still detected by the other generating units and therefore, the attack will not influence the other units. For the units that are under attack, it's not possible for them to get the correct ACE values even through the ACE estimation because the linear regression model that they have are incorrect. But they can be excluded from the AGC temporarily after being reported by the other units.

5. **Scenario 5:** Scaling attacks targeting unit 1 - 5 are simulated before linear regression. This is the most severe case and cannot be effectively handled by the proposed mitigation method. For the units under attack, each one of them is able to distinguish the other units under

attack, but the ACE estimation will not work since the model obtained are incorrect. The attack impacts cannot be mitigated by low reputation, that is because for the units which are not attacked, when they trying to find the majority from the randomly selected peers, the majority will be the attacked units. Thus, the attacked estimation value will outrun the true ACE value. The reputation evolution is depicted in Figure 4.20. Even they are able to report the low reputation units to the control center, the nomination itself will be incorrect. The only way that this scenario can be mitigated is that the additional indicator function in Algorithm.4 is set to -1.

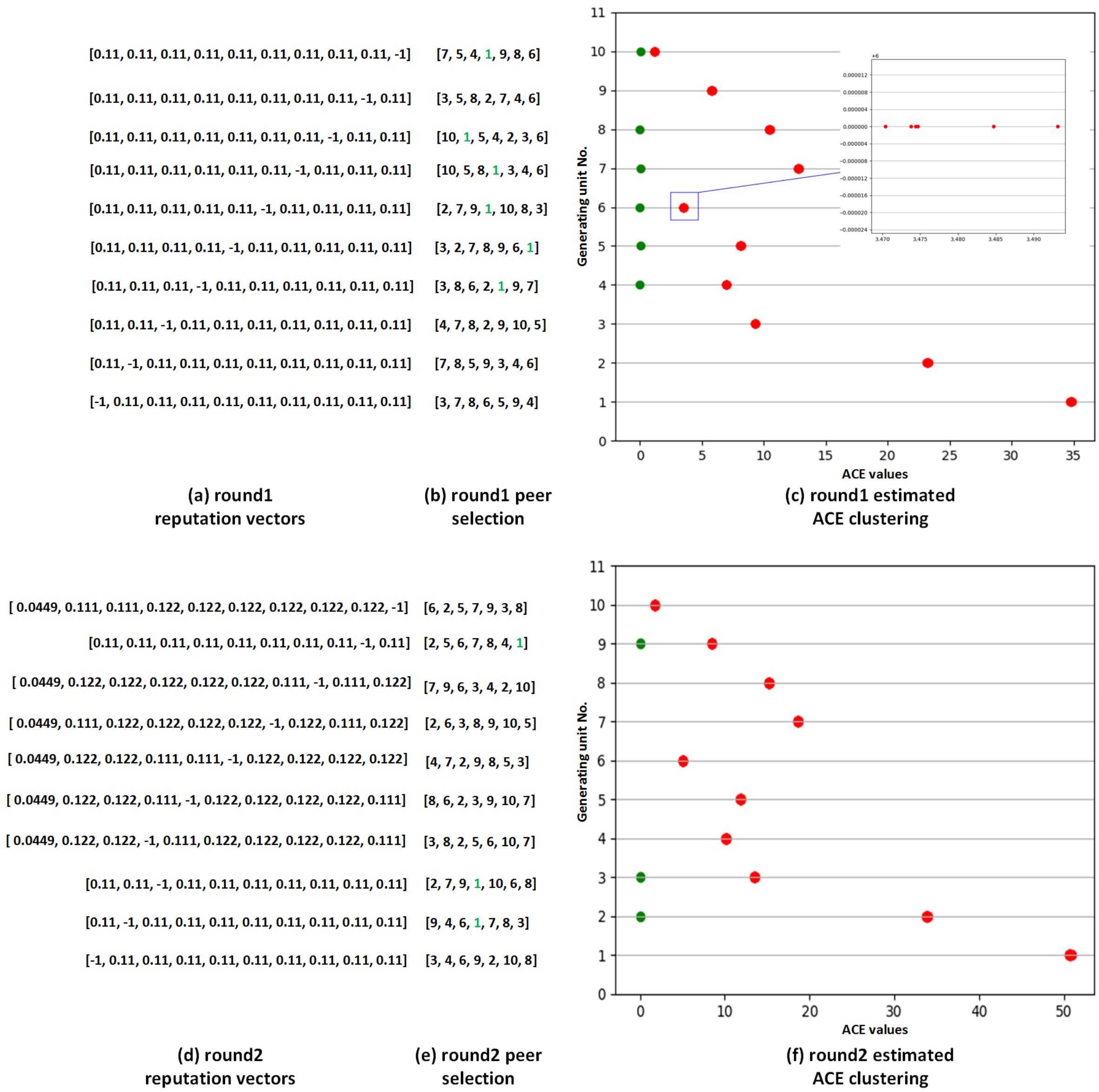
Figure 4.21 shows the output of the first three generators in the 39-bus model under scenario 1 above, with and without proposed mitigation strategy in place. The ACE values sent to generator 30 (unit 1) is manipulated to 0 when under attack, therefore, the generation output of generator 30 when it is under attack only contains the fluctuation due to its governor function (see (b)). On the other hand, when the proposed mitigation is present, although the ACE from control center to generator 30 is still 0, it can estimate the correct ACE value based on the peer information. Therefore, gen 30 still actively participates in the AGC (see (a)). It is worth to note that since all the ten units are involved in the AGC, it is not very likely that cyber attacks might induce system stability issues even the mitigation is not adopted, but the economic performance of the generation control will surely get impaired without mitigation.

4.8 Summary

Data-driven anomaly detection of generation control is smart grids is investigated in this chapter. Two different clustering techniques are adopted - K-means and HDBSCAN. HDBSCAN considers the density information of the data set, and thus it results in more satisfactory clustering results.

A peer-assisted ACE estimation strategy is proposed to mitigate the data availability and integrity attacks targeting ACE values.

The countermeasure proposed in this chapter is implemented at the station level and it serves as the last line of defense to counteract the attacks targeting the generation control. The anomaly



[0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, -1]	[7, 5, 4, 1, 9, 8, 6]
[0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, -1, 0.11]	[3, 5, 8, 2, 7, 4, 6]
[0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, -1, 0.11, 0.11]	[10, 1, 5, 4, 2, 3, 6]
[0.11, 0.11, 0.11, 0.11, 0.11, 0.11, -1, 0.11, 0.11, 0.11]	[10, 5, 8, 1, 3, 4, 6]
[0.11, 0.11, 0.11, 0.11, 0.11, -1, 0.11, 0.11, 0.11, 0.11]	[2, 7, 9, 1, 10, 8, 3]
[0.11, 0.11, 0.11, 0.11, -1, 0.11, 0.11, 0.11, 0.11, 0.11]	[3, 2, 7, 8, 9, 6, 1]
[0.11, 0.11, 0.11, -1, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11]	[3, 8, 6, 2, 1, 9, 7]
[0.11, 0.11, -1, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11]	[4, 7, 8, 2, 9, 10, 5]
[0.11, -1, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11]	[7, 8, 5, 9, 3, 4, 6]
[-1, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11]	[3, 7, 8, 6, 5, 9, 4]

(a) round1 reputation vectors

(b) round1 peer selection

(c) round1 estimated ACE clustering

[0.0449, 0.111, 0.111, 0.122, 0.122, 0.122, 0.122, 0.122, 0.122, -1]	[6, 2, 5, 7, 9, 3, 8]
[0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, -1, 0.11]	[2, 5, 6, 7, 8, 4, 1]
[0.0449, 0.122, 0.122, 0.122, 0.122, 0.122, 0.111, -1, 0.111, 0.122]	[7, 9, 6, 3, 4, 2, 10]
[0.0449, 0.111, 0.122, 0.122, 0.122, 0.122, -1, 0.122, 0.111, 0.122]	[2, 6, 3, 8, 9, 10, 5]
[0.0449, 0.122, 0.122, 0.111, 0.111, -1, 0.122, 0.122, 0.122, 0.122]	[4, 7, 2, 9, 8, 5, 3]
[0.0449, 0.122, 0.122, 0.111, -1, 0.122, 0.122, 0.122, 0.122, 0.111]	[8, 6, 2, 3, 9, 10, 7]
[0.0449, 0.122, 0.122, -1, 0.111, 0.122, 0.122, 0.122, 0.122, 0.111]	[3, 8, 2, 5, 6, 10, 7]
[0.11, 0.11, -1, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11]	[2, 7, 9, 1, 10, 6, 8]
[0.11, -1, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11]	[9, 4, 6, 1, 7, 8, 3]
[-1, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11]	[3, 4, 6, 9, 2, 10, 8]

(d) round2 reputation vectors

(e) round2 peer selection

(f) round2 estimated ACE clustering

Figure 4.15 Reputation update based on 1-d clustering

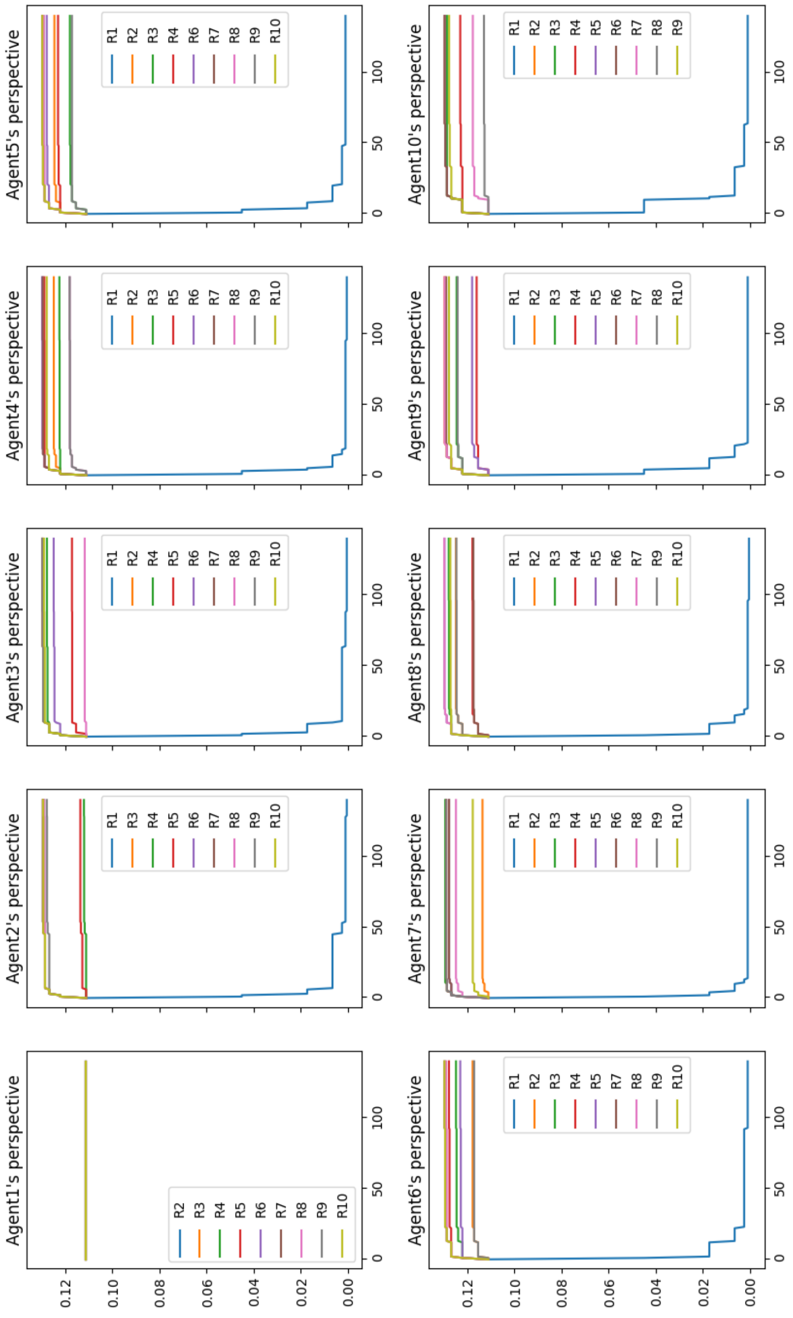


Figure 4.16 Reputation when agent 1 under constant ACE attack

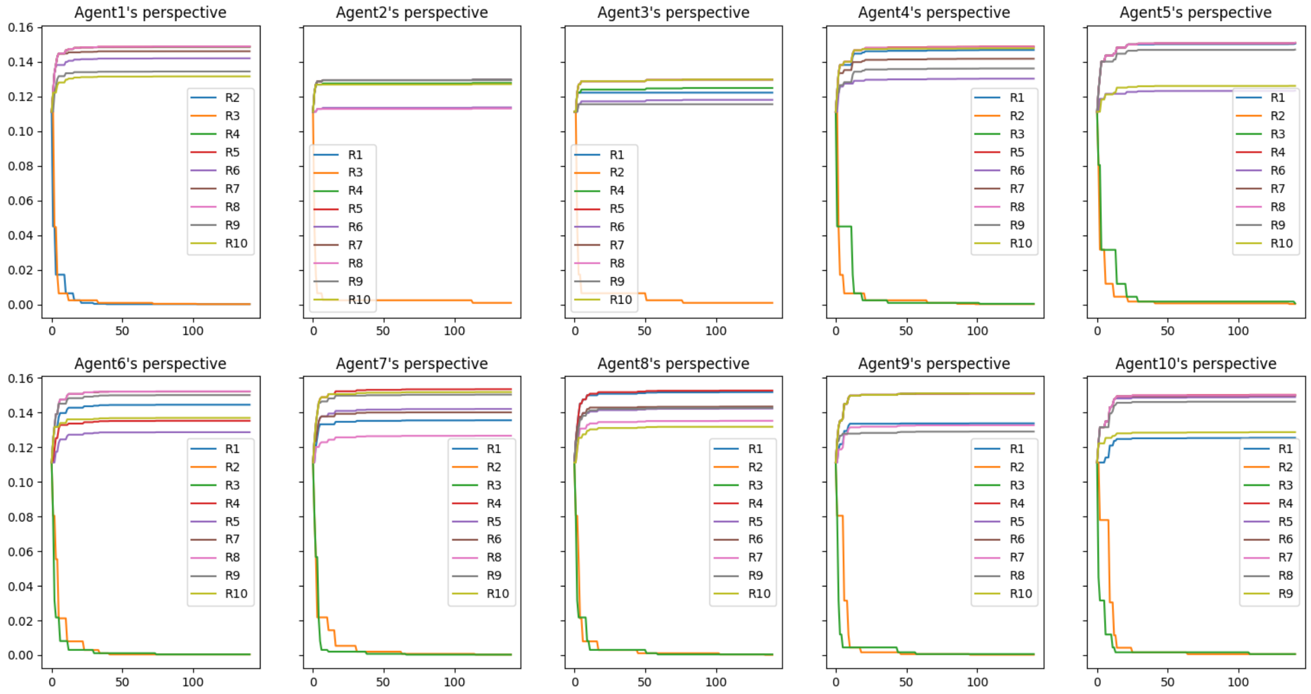


Figure 4.17 Reputation when agent 2 and 3 under flip ACE attack

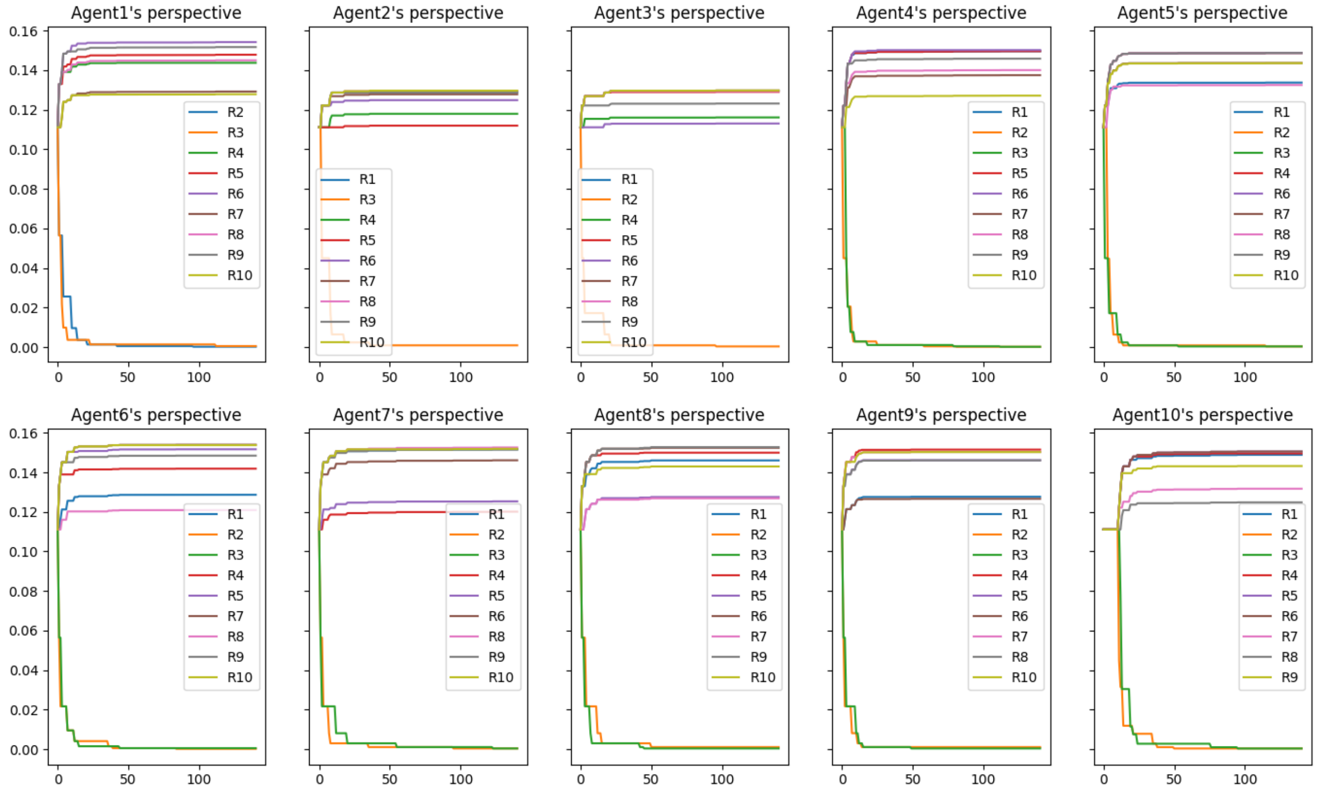


Figure 4.18 Reputation when agent 2 and 3 under constant ACE attack

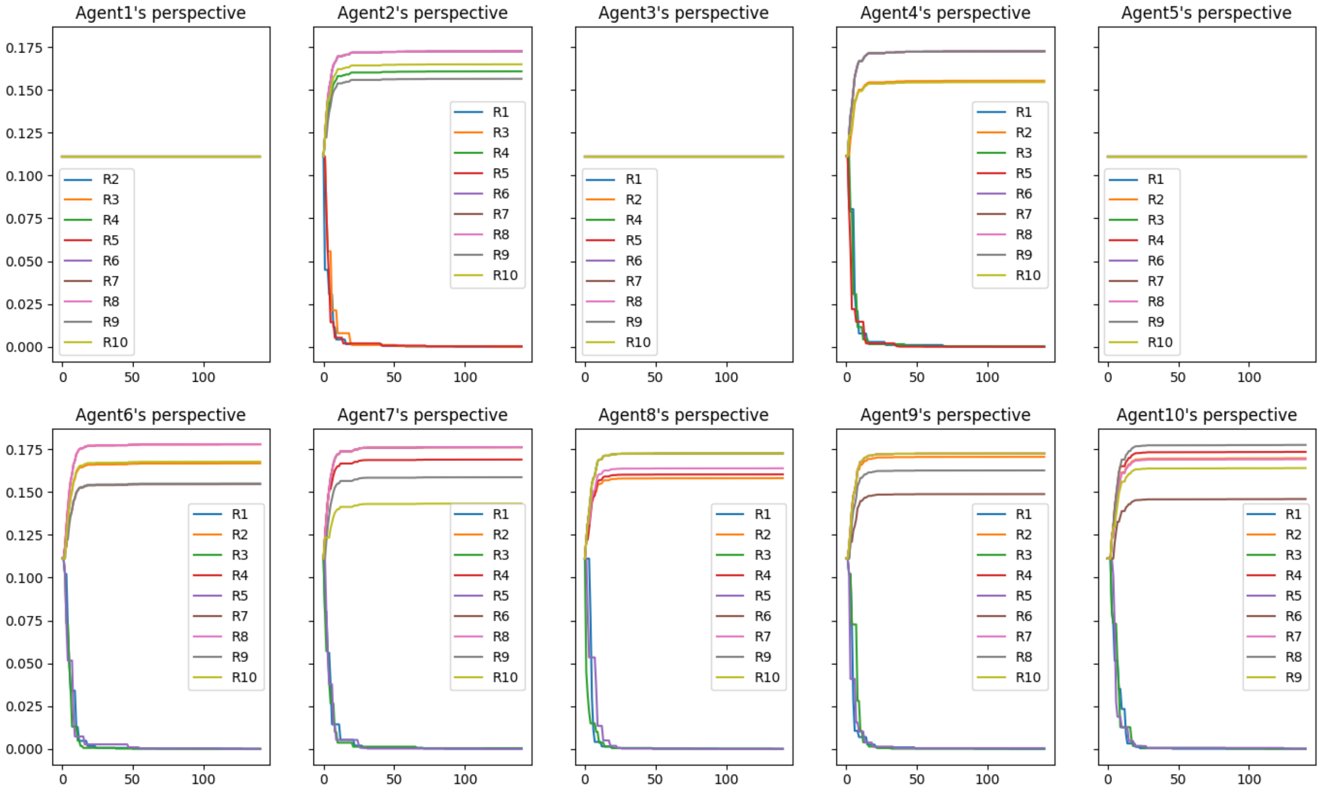


Figure 4.19 Reputation when agent 1,3 and 4 under constant ACE attack

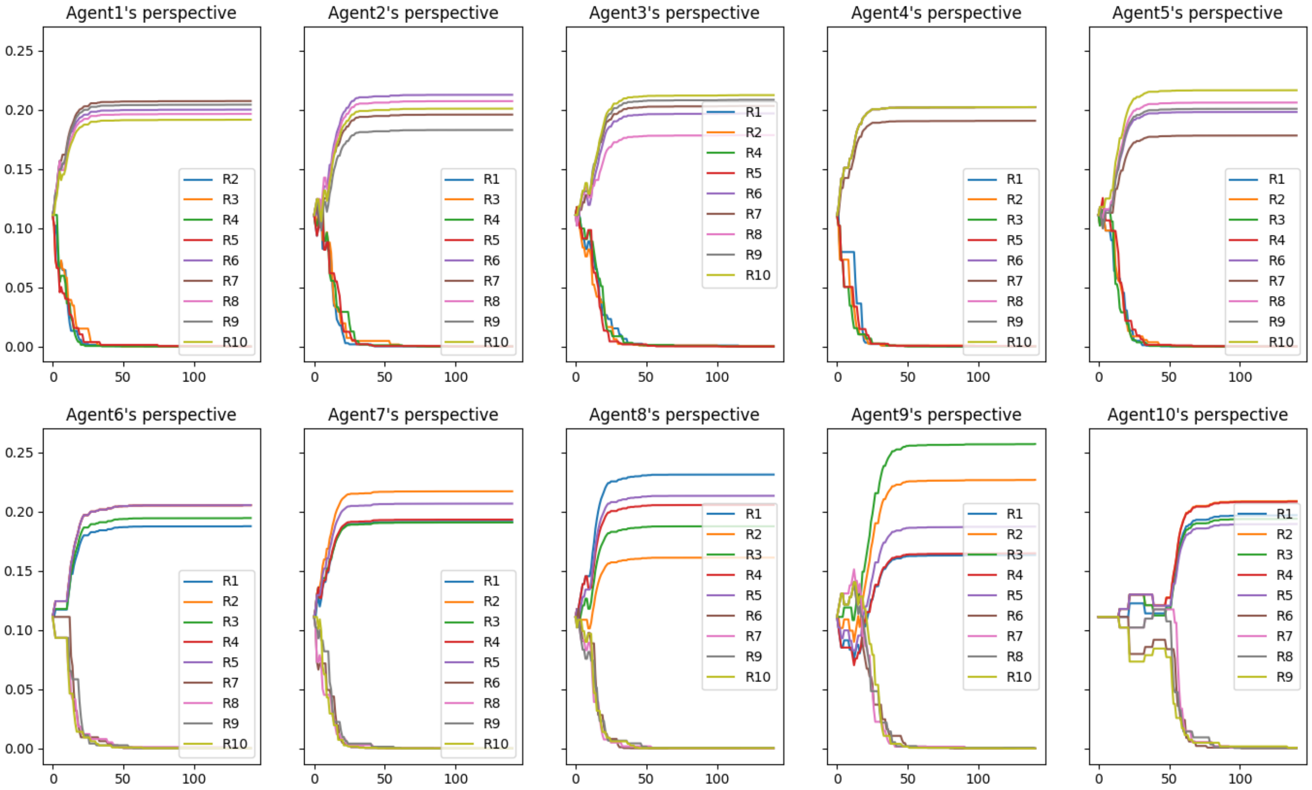
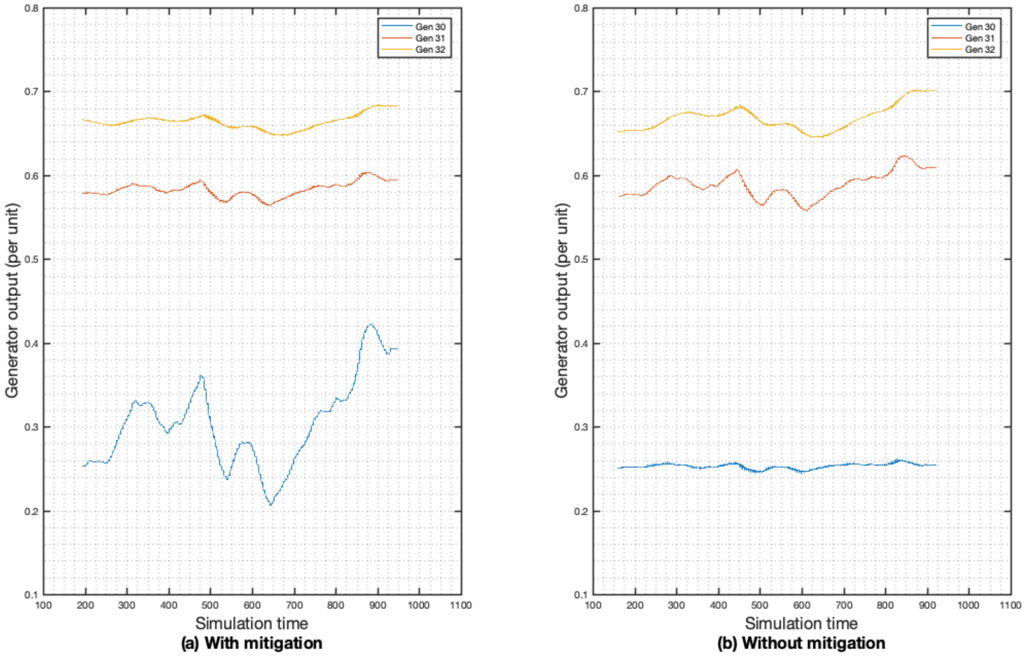


Figure 4.20 Reputation when agent 1-5 under scaling ACE attack



(a) With mitigation

(b) Without mitigation

Figure 4.21 Generation output with and without mitigation under scenario 1

detection method proposed is able to detect the attacks targeting both the measurements and ACEs as long as not all the data are compromised. For the peer-assisted mitigation, it is only effective in mitigating ACE attacks. Besides, the performance of the mitigation will be improved if the communication infrastructure interconnecting generator agents possesses certain level of redundancy and heterogeneousness.

CHAPTER 5. CONCLUSION

This dissertation focuses on the cyber attack detection and mitigation for wide area protection and control functions in smart grids in order to improve the situational awareness and self-adaptiveness of these critical functions and hence enhance the overall power system resilience.

5.1 Problems Resolved and Proposed Solution

Wide-area protection and control schemes in power systems are often centralized functions which are prone to single point failure resulted from malicious cyber attacks. To enhance the robustness of these critical cyber layer functions on which the physical layer's secure and reliable operation highly relies is of the most importance. Except for traditional IT security techniques, empirical best practices, and the solutions that can only be applied in the control center, cyber attack detection and mitigation based on a peer-to-peer and decentralized functional structure provides another promising research direction.

First part of the dissertation aims to make the SIPS resilient to data availability attacks. Decentralization based on MAS is adopted as an alternative to the centralized SIPS architecture. Anomaly detection and adaptive load shedding are realized with machine learning and optimization techniques respectively.

- For anomaly detection, a data-driven algorithm with high interpretbility, SVMLDT, is presented.
- For adaptive load shedding, an optimal load shedding scheme leveraging the historic knowledge is designed for the agents when the communication is compromised.

A realistic load rejection SIPS is mapped to IEEE 39 bus system as case study to validate the anomaly detection algorithm. Two synthetic data sets considering different events are generated with real time simulation. Detection results are collected with 5 different machine learning algorithms. Adaptive load shedding scheme is tested for scenarios when the system is under different attacks. Dynamic load profile (including load variation and local marginal price) is generated according to the statistics from MISO. Agent prototype has been programmed in Java and the data propagation protocol is verified.

Second part of the dissertation presents a data-driven anomaly detection algorithm for power system generation control and a mitigation of ACE attacks based on peer-assisted ACE estimation.

As a dimensionality reduction method for temporal-spatial data set, three conformity metrics are defined based on the generation control signals from adjacent peers. Semi-supervised K-means clustering is utilized for the anomaly detection. Considering the density information is ignored by K-means clustering, an improved version of HDBSCAN is developed and it results in better clustering performance. The mitigation proposed aims at mitigating the attacks targeting the ACE being sent from control center to generator units. A generating unit randomly chooses which accessible peers to request data from, and this reduces the overhead of mitigation and serves as a moving target defense against the cyber attacks. Peer selection is achieved according to the reputation of peers, which keeps being updated. With the information from peers, a generating unit can estimate its ACE value and use it to replace the abnormal values received from control center. Besides, a low reputation report strategy is proposed such that generating units can inform the control center about the peers with low reputation, which will be temporarily excluded from the AGC.

For detection evaluation, IEEE 39 bus system which is divided into 3 areas with AGC implemented individually. Synthetic data set is generated considering normal load change and common generation attacks such as constant ACE attack, generatin ramp attack. It's shown that the K-means clustering can successfully distinguish between the normal states and abnormal states, but its performance is not very satisfactory in telling apart two attack scenarios. In contrast, density sensitive method provides better performance. For mitigation, all the 10 generating units in IEEE

39 bus model are configured in AGC and experiments suggest that the solution proposed achieves the goal to rid the impacts of ACE attack when less than half of the agents are compromised.

5.2 Future Work

1. **MAS organization and protocol:** The communication protocol of MAS needs to be further explored in order to improve the communication efficiency and the timing performance for on-line anomaly detection.

2. **Data-driven algorithms:** Each detection and mitigation strategy proposed in this work can be replaced by alternatives in a plug-and-play manner, so more efficient algorithms should be developed and tested for comparison.

3. **Feature extraction techniques:** It is necessary to investigate and utilize more general feature reduction techniques to avoid the high requirements of domain knowledge in feature extraction.

4. **Data integrity attack targeting adaptive control:** Chapter III only considers DoS attack while investigating the optimal adaptive load shedding scheme. One future work is to improve the solution and to ensure that the adaptive control can stay effective under other attacks such as data integrity attack.

5. **Comprehensive mitigation solution:** The mitigation strategy proposed in chapter IV is not able to handle false injection attacks targeting the measurements of AGC. This can be resolved by adopting available model-based methods or developing a countermeasure that also counteracts such attacks based on the peer-to-peer data exchange.

6. **Solution application in other WAMPAC functions:** The detection and mitigation solution proposed in this dissertation can be potentially applied to other functions such as wide area damping control and wide area voltage control, i.e., the same MAS could and should be used to implement various WAMPAC functions.

BIBLIOGRAPHY

[1] "Nist framework and roadmap for smart grid interoperability standards, release 3.0," Tech. Rep. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1108r3.pdf

[2] A. F. Zobaa and A. Vaccaro, *Computational Intelligence Applications in Smart Grids: Enabling Methodologies for Proactive and Self-Organizing Power Systems.* London, UK, UK: Imperial College Press, 2015.

[3] A. Ashok, M. Govindarasu, and J. Wang, "Cyber-physical attack-resilient wide-area monitoring, protection, and control for the power grid," *Proceedings of the IEEE*, vol. 105, no. 7, pp. 1389–1407, 7 2017.

[4] Electric grid cybersecurity. [Online]. Available: https://fas.org/sgp/crs/homesec/R45312.pdf

[5] I. N. Lab, "Cyber threat and vulnerability analysis of the u.s. electric sector," 2016. [Online]. Available: https://www.energy.gov/sites/prod/files/2017/01/f34/Cyber%20Threat%20and%20Vulnerability%20Analysis%20of%20the%20U.S.%20Electric%20Sector.pdf

[6] S. Ward, J. O'Brien, B. Beresh, G. Benmouyal, D. Holstein, J. T. Tengdin, K. Fodero, M. Simon, M. Carden, M. V. V. S. Yalla, T. Tibbals, V. Skendzic, S. Mix, R. Young, T. Sidhu, S. Klein, J. Weiss, A. Apostolov, D. Bui, S. Sciacca, C. Preuss, S. Hodder, and G. Seifert, "Cyber security issues for protective relays; c1 working group members of power system relaying committee," in *2007 IEEE Power Engineering Society General Meeting*, 6 2007, pp. 1–8.

[7] N. A. of Sciences, "Terrorism and the electric power delivery system," 2012. [Online]. Available: https://www.nap.edu/catalog/12050/terrorism-and-the-electric-power-delivery-system

[8] J. Dougherty, "Biggest u.s. power grid operator suffers thousands of attempted cyber attacks per month," https://forwardobserver.com/2017/08/biggest-u-s-power-grid-operator-suffers-thousands-of-attempted-cyber-attacks-per-month/, 8 2017, [Online; accessed 19-April-2018].

[9] S. I. E-ISAC, "Analysis of the cyber attack on the ukrainian power grid," 3 2016.

[10] L. O. M. Nicolas Falliere and E. Chien, "W32.stuxnet dossier," 2 2011.

[11] Attackers deploy new ics attack framework triton and cause operational disruption to critical infrastructure. [Online]. Available: https://www.fireeye.com/blog/threat-research/2017/12/attackers-deploy-new-ics-attack-framework-triton.html

[12] R. Liu, C. Vellaithurai, S. S. Biswas, T. T. Gamage, and A. K. Srivastava, "Analyzing the cyber-physical impact of cyber events on the power grid," *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2444–2453, 9 2015.

[13] S. Liu, B. Chen, T. Zourntos, D. Kundur, and K. Butler-Purry, "A coordinated multi-switch attack for cascading failures in smart grid," *IEEE Transactions on Smart Grid*, vol. 5, no. 3, pp. 1183–1195, 5 2014.

[14] S. Liu, S. Mashayekh, D. Kundur, T. Zourntos, and K. Butler-Purry, "A framework for modeling cyber-physical switching attacks in smart grid," *IEEE Transactions on Emerging Topics in Computing*, vol. 1, no. 2, pp. 273–285, 12 2013.

[15] A. Ashok, Wang, M. Brown, and M. Govindarasu, "Experimental evaluation of cyber attacks on automatic generation control using a cps security testbed," in *2015 IEEE Power Energy Society General Meeting*, 7 2015, pp. 1–5.

[16] A. Ashok, S. Sridhar, A. D. McKinnon, P. Wang, and M. Govindarasu, "Testbed-based performance evaluation of attack resilient control for agc," in *2016 Resilience Week (RWS)*, Aug 2016, pp. 125–129.

[17] V. Venkataramanan, P. Wang, A. Srivastava, A. Hahn, and M. Govindarasu, "Interfacing techniques in testbed for cyber-physical security analysis of the electric power grid," in *2017 Workshop on Modeling and Simulation of Cyber-Physical Energy Systems (MSCPES)*, April 2017, pp. 1–6.

[18] V. Kumar Singh, A. Ozen, and M. Govindarasu, "Stealthy cyber attacks and impact analysis on wide-area protection of smart grid," in *2016 North American Power Symposium (NAPS)*, Sep. 2016, pp. 1–6.

[19] E. E. Bernabeu and F. Katiraei, "W32.stuxnet dossier," https://www.smartgrid.gov/files/Aurora_Vulnerability_Issues_Solution_Hardware_Mitigation_De_201102.pdf, 2 2011, [Online; accessed 1-Aug-2017].

[20] P. Wang, A. Ashok, and M. Govindarasu, "Cyber-physical risk assessment for smart grid system protection scheme," in *2015 IEEE Power Energy Society General Meeting*, 7 2015, pp. 1–5.

[21] Y. Jiang, S. Chen, C.-C. Liu, W. Sun, X. Luo, S. Liu, N. Bhatt, S. Uppalapati, and D. Forcum, "Blackstart capability planning for power system restoration," *International Journal of Electrical Power and Energy Systems*, vol. 86, pp. 127 – 137, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0142061516307220

[22] W. Wang and Z. Lu, "Cyber security in the smart grid: Survey and challenges," *Computer Networks*, vol. 57, no. 5, pp. 1344 – 1371, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128613000042

[23] R. Amoah, S. Camtepe, and E. Foo, “Securing dnp3 broadcast communications in scada systems,” *IEEE Transactions on Industrial Informatics*, vol. 12, no. 4, pp. 1474–1485, 8 2016.

[24] R. Schlegel, S. Obermeier, and J. Schneider, “A security evaluation of iec 62351,” *Journal of Information Security and Applications*, vol. 34, pp. 197 – 204, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2214212616300771

[25] Y. Yan, Y. Qian, H. Sharif, and D. Tipper, “A survey on cyber security for smart grid communications,” *IEEE Communications Surveys Tutorials*, vol. 14, no. 4, pp. 998–1010, 4 2012.

[26] V. K. Singh, H. Ebrahem, and M. Govindarasu, “Security evaluation of two intrusion detection systems in smart grid scada environment,” in *2018 North American Power Symposium (NAPS)*, Sep. 2018, pp. 1–6.

[27] DOE, “Critical infrastructure protection.” [Online]. Available: https://www.energy.gov/sites/prod/files/Energy%20Delivery%20Systems%20Cybersecurity%20Roadmap_finalweb.pdf

[28] NERC, “Critical infrastructure protection.” [Online]. Available: https://www.nerc.com/pa/Stand/Pages/CIPStandards.aspx

[29] N. I. of Standards and Technology, *Framework for Improving Critical Infrastructure Cybersecurity*. USA: CreateSpace Independent Publishing Platform, 2014.

[30] Virginia electric utilities wiring rural areas for broadband. [Online]. Available: https://apnews.com/6de7257530a546849c6b090d70053dd1

[31] Big data best practices. [Online]. Available: https://www.wecc.org/Administrative/08%202017-05-JSIS-%20Big%20Data%20Best%20Practices.Murphy.pdf

[32] Z. Huang, H. Luo, D. Skoda, T. Zhu, and Y. Gu, “E-sketch: Gathering large-scale energy consumption data based on consumption patterns,” in *2014 IEEE International Conference on Big Data (Big Data)*, Oct 2014, pp. 656–665.

[33] J. Hu and A. V. Vasilakos, “Energy big data analytics and security: Challenges and opportunities,” *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2423–2436, Sep. 2016.

[34] V. Madani, D. Novosel, M. Begovic, and M. Adamiak, “Application considerations in system integrity protection schemes (sips),” *GE Magazine*, pp. 25–30, 2008.

[35] J. Sykes, Y. Hu, M. Adamiak, A. Apostolov, B. Dac-Phuoc, A. Deronja, J. Ebrecht, G. Henneberg, S. Imai, V. Madani, D. Miller, A. D. L. Quintana, B. Vandiver, R. Whittaker, M. Zubair, and S. Ward, “Ieee/pes psrc report on design and testing of selected system integrity protection schemes,” in *2014 67th Annual Conference for Protective Relay Engineers*, 3 2014, pp. 738–742.

[36] M. Adamiak, A. Apostolov, M. Begovic, C. Henville, K. Martin, G. Michel, A. Phadke, and J. Thorp, "Wide area protection-technology and infrastructures," *Power Delivery, IEEE Transactions on*, vol. 21, no. 2, pp. 601–609, 4 2006.

[37] V. Madani, J. Sykes, and M. Adamiak, "Wide area protection schemes - design and implementation," *PAC World*, 2009.

[38] NERC, "Remedial action scheme definition development," 6 2014. [Online]. Available: https://www.nerc.com/pa/Stand/Prjct201005_2SpclPrtctnSstmPhs2/FAQ_RAS_Definition_0604_final.pdf

[39] V. Madani, D. Novosel, S. Horowitz, M. Adamiak, J. Amantegui, D. Karlsson, S. Imai, and A. Apostolov, "Ieee psrc report on global industry experiences with system integrity protection schemes (sips)," *IEEE Transactions on Power Delivery*, vol. 25, no. 4, pp. 2143–2155, 10 2010.

[40] J. Sykes, M. Adamiak, and G. Brunello, "Implementation and operational experience of a wide area special protection scheme on the srp system," in *Power Systems Conference: Advanced Metering, Protection, Control, Communication, and Distributed Resources, 2006. PS '06*, 3 2006, pp. 145–158.

[41] M. S. Rahman, M. A. Mahmud, A. M. T. Oo, and H. R. Pota, "Multi-agent approach for enhancing security of protection schemes in cyber-physical energy systems," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 436–447, 4 2017.

[42] J. Jung, C.-C. Liu, S. Tanimoto, and V. Vittal, "Adaptation in load shedding under vulnerable operating conditions," *Power Systems, IEEE Transactions on*, vol. 17, no. 4, pp. 1199–1205, 11 2002.

[43] K. J. Ross, K. M. Hopkinson, and M. Pachter, "Using a distributed agent-based communication enabled special protection system to enhance smart grid security," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 1216–1224, 6 2013.

[44] S. D. J. McArthur, E. M. Davidson, V. M. Catterson, A. L. Dimeas, N. D. Hatziargyriou, F. Ponci, and T. Funabashi, "Multi-agent systems for power engineering applications part i: Concepts, approaches, and technical challenges," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1743–1752, 11 2007.

[45] ——, "Multi-agent systems for power engineering applications part ii: Technologies, standards, and tools for building multi-agent systems," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1753–1759, 11 2007.

[46] A. Ashrafi and S. Shahrtash, "Dynamic wide area voltage control strategy based on organized multi-agent system," *Power Systems, IEEE Transactions on*, vol. 29, no. 6, pp. 2590–2601, 11 2014.

[47] Z. Liu, Z. Chen, H. Sun, and Y. Hu, “Multiagent system-based wide-area protection and control scheme against cascading events,” *Power Delivery, IEEE Transactions on*, vol. 30, no. 4, pp. 1651–1662, 8 2015.

[48] A. Manickam, S. Kamalasadan, D. Edwards, and S. Simmons, “A novel self-evolving intelligent multiagent framework for power system control and protection,” *IEEE Systems Journal*, vol. 8, no. 4, pp. 1086–1095, 12 2014.

[49] X. Tong, X. Wang, R. Wang, F. Huang, X. Dong, K. M. Hopkinson, and G. Song, “The study of a regional decentralized peer-to-peer negotiation-based wide-area backup protection multi-agent system,” *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 1197–1206, 6 2013.

[50] A. Parmar, J. Gnanadhas, T. T. Mini, G. Abhilash, and A. C. Biswal, “Multi-agent approach for anomaly detection in automation networks,” in *International Conference on Circuits, Communication, Control and Computing*, 11 2014, pp. 225–230.

[51] C. Rieger and Q. Zhu, “A hierarchical multi-agent dynamical system architecture for resilient control systems,” in *2013 6th International Symposium on Resilient Control Systems (ISRCS)*, 8 2013, pp. 6–12.

[52] C. G. Rieger, K. L. Moore, and T. L. Baldwin, “Resilient control systems: A multi-agent dynamic systems perspective,” in *IEEE International Conference on Electro-Information Technology , EIT 2013*, 5 2013, pp. 1–16.

[53] V. K. Singh, A. Ozen, and M. Govindarasu, “A hierarchical multi-agent based anomaly detection for wide-area protection in smart grid,” in *2018 Resilience Week (RWS)*, Aug 2018, pp. 63–69.

[54] M. Wu and L. Xie, “Online detection of low-quality synchrophasor measurements: A data-driven approach,” *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2817–2827, 7 2017.

[55] M. Zhou, Y. Wang, A. K. Srivastava, Y. Wu, and P. Banerjee, “Ensemble based algorithm for synchrophasor data anomaly detection,” *IEEE Transactions on Smart Grid*, pp. 1–1, 2018.

[56] T. Guo and J. V. Milanovi, “Online identification of power system dynamic signature using pmu measurements and data mining,” *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 1760–1768, 5 2016.

[57] B. Wang, B. Fang, Y. Wang, H. Liu, and Y. Liu, “Power system transient stability assessment based on big data and the core vector machine,” *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2561–2570, 9 2016.

[58] S. Pan, T. Morris, and U. Adhikari, “Developing a hybrid intrusion detection system using data mining for power systems,” *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 3104–3113, 11 2015.

[59] M. Rafferty, X. Liu, D. M. Laverty, and S. McLoone, “Real-time multiple event detection and classification using moving window pca,” *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2537–2548, 9 2016.

[60] V. K. Singh and M. Govindarasu, “Decision tree based anomaly detection for remedial action scheme in smart grid using pmu data,” in *2018 IEEE Power Energy Society General Meeting (PESGM)*, Aug 2018, pp. 1–5.

[61] S. Waghmare, F. Kazi, and N. Singh, “Data driven approach to attack detection in a cyber-physical smart grid system,” in *2017 Indian Control Conference (ICC)*, 1 2017, pp. 271–276.

[62] N. Lu, P. Du, F. L. Greitzer, X. Guo, R. E. Hohimer, and Y. G. Pomiak, “A multi-layer, data-driven advanced reasoning tool for intelligent data mining and analysis for smart grids,” in *2012 IEEE Power and Energy Society General Meeting*, 7 2012, pp. 1–7.

[63] S. Manson, G. Zweigle, and V. Yedidi, “Case study: An adaptive underfrequency load-shedding system,” in *Industry Applications Society 60th Annual Petroleum and Chemical Industry Conference*, 9 2013, pp. 1–9.

[64] Y. Xu, W. Liu, and J. Gong, “Stable multi-agent-based load shedding algorithm for power systems,” *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2006–2014, 11 2011.

[65] P. Wang and M. Govindarasu, “Multi intelligent agent based cyber attack resilient system protection and emergency control,” in *2016 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 9 2016, pp. 1–5.

[66] L. H. Fink and K. Carlsen, “Operating under stress and strain [electrical power systems control under emergency conditions],” *IEEE Spectrum*, vol. 15, no. 3, pp. 48–53, 3 1978.

[67] F. Takahashi and S. Abe, “Decision-tree-based multiclass support vector machines,” in *Neural Information Processing, 2002. ICONIP ’02. Proceedings of the 9th International Conference on*, vol. 3, 11 2002, pp. 1418–1422 vol.3.

[68] M. A. Kumar and M. Gopal, “A hybrid svm based decision tree,” *Pattern Recognition*, vol. 43, no. 12, pp. 3977 – 3987, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320310003067

[69] Package c50. [Online]. Available: https://cran.r-project.org/web/packages/C50/C50.pdf

[70] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R.* Springer Publishing Company, Incorporated, 2014.

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

[72] [Online]. Available: https://www.misoenergy.org/MARKETSOPERATIONS/REALTIMEMARKETDATA/Pages/ACEChart.aspx

[73] NERC. (2011, 1) Balancing and frequency control a technical document prepared by the nerc resources subcommiittee.

[74] C.-W. Ten, C.-C. Liu, and G. Manimaran, “Vulnerability assessment of cybersecurity for scada systems,” *Power Systems, IEEE Transactions on*, vol. 23, no. 4, pp. 1836–1846, 11 2008.

[75] A. Ameli, A. Hooshyar, E. F. El-Saadany, and A. M. Youssef, “Attack detection and identification for automatic generation control systems,” *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 4760–4774, 9 2018.

[76] 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, 7 2017.

[77] S. Liu, X. P. Liu, and A. E. Saddik,, “Denial-of-service (dos) attacks on load frequency control in smart grids,” in *2013 IEEE PES Innovative Smart Grid Technologies Conference (ISGT)*, Feb 2013, pp. 1–6.

[78] 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, 3 2014.

[79] J. D. D. Glover and M. S. Sarma, *Power System Analysis and Design*, 3rd ed. Pacific Grove, CA, USA: Brooks/Cole Publishing Co., 2001.

[80] [Online]. Available: https://www.mathworks.com/help/signal/ref/sgolay.html

[81] P. Wang, M. Govindarasu, A. Ashok, S. Sridhar, and D. McKinnon, “Data-driven anomaly detection for power system generation control,” in *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 11 2017, pp. 1082–1089.

[82] P. Wang and M. Govindarasu, “Anomaly detection for power system generation control based on hierarchical dbscan,” in *2018 North American Power Symposium (NAPS)*, 9 2018, pp. 1–5.

[83] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise,” in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, ser. KDD’96. AAAI Press, 1996, pp. 226–231. [Online]. Available: http://dl.acm.org/citation.cfm?id=3001460.3001507

[84] R. J. G. B. Campello, D. Moulavi, A. Zimek, and J. Sander, “Hierarchical density estimates for data clustering, visualization, and outlier detection,” *ACM Trans. Knowl. Discov. Data*, vol. 10, no. 1, pp. 5:1–5:51, Jul. 2015. [Online]. Available: http://doi.acm.org/10.1145/2733381

[85] X. Wang, D. Shi, Z. Wang, C. Xu, Q. Zhang, X. Zhang, and Z. Yu, “Online calibration of phasor measurement unit using density-based spatial clustering,” *IEEE Transactions on Power Delivery*, vol. PP, no. 99, pp. 1–1, 2017.

[86] A. J. Wood and B. F. Wollenberg, *Power Generation, Operation, and Control*, 2nd ed.