Skoltech

Skolkovo Institute of Science and Technology

Skolkovo Institute of Science and Technology

Advancements in Power System State Estimation: Innovative Algorithms and Solutions for Enhanced Reliability and Efficiency

Doctoral Thesis
by

Sajjad Asefi

Doctoral Program in Engineering Systems

Supervisor
Assistant professor, Elena Gryazina

Moscow - 2023

I hereby declare that the work presented in this thesis was carried out by myself at Skolkovo Institute of Science and Technology, Moscow, except where due acknowledgement is made, and has not been submitted for any other degree.

Sajjad Asefi (Candidate)

Prof. Elena Gryazina (Supervisor)

Advancements in Power System State Estimation: Innovative Algorithms and Solutions for Enhanced Reliability and Efficiency

by

Sajjad Asefi

Tuesday, 26 September 2023

Abstract

This thesis explores and introduces significant advancements in the field of power system state estimation, aiming to enhance the reliability and efficiency of modern power grids. The power system serves as the backbone of our interconnected world, ensuring the continuous and secure supply of electricity. However, the increasing complexity and vulnerability of this infrastructure pose significant challenges. This research addresses these challenges through a multifaceted approach.

First, a novel algorithm for the detection and classification of single/multi-bus sudden load change and single/multi-state false data injection attack has been developed. This algorithm leverages an anomaly detection index and utilizes supervised machine learning techniques for accurate classification. The proposed method is a topology-resilient classification solution, mitigating the adverse effects of topology changes by focusing on features associated solely with buses. Also, the proposed method accelerates the training speed of machine learning algorithms by applying an optimal feature selection method, "maximum relevance - minimum redundancy", leading to more efficient state estimation processes.

Additionally, the study evaluates various distributed state estimation methods using a modified convergence criterion and IEEE standard test systems. This evaluation results in the selection of the most effective method based on metrics such as data transfer requirements, computation time, solution accuracy, and convergence rates. Furthermore, an optimal power system partitioning method has been introduced to reduce communication overhead in distributed state estimation, significantly decreasing the number of iterations required to attain a reasonable solution.

The research also explores the integration of blockchain technology to enhance the security of data transfers within the distributed power system state estimation framework, ensuring the integrity of critical information. Moreover, the study considers the challenges posed by asynchronous and delayed data transfer within distributed state estimation, contributing to the development of robust algorithms capable of handling real-world communication constraints.

Collectively, this thesis presents a comprehensive and innovative approach to power system state estimation, offering practical solutions to enhance the reliability and efficiency of power grid operations, ultimately contributing to the resilience and sustainability of modern power systems.

Publications

Main

1. **S. Asefi**, M. Mitrovic, D. Ćetenović, V. Levi, E. Gryazina, and V. Terzija, "Anomaly detection and classification in power system state estimation: Combining model-based and data-driven methods," *Sustainable Energy, Grids and Networks*, vol. 35, p. 101116, 2023

2. **S. Asefi**, M. Mitrovic, D. Ćetenović, V. Levi, E. Gryazina, and V. Terzija, "Anomaly detection, classification and identification tool (ADCIT)," *Software Impacts*, vol. 15, p. 100465, 2023

3. **S. Asefi**, Y. Madhwal, Y. Yanovich, and E. Gryazina, "Application of blockchain for secure data transmission in distributed state estimation," *IEEE Transactions on Control of Network Systems*, 2021

4. **S. Asefi**, E. Gryazina, and H. Leite, "Optimal partitioning in distributed state estimation considering a modified convergence criterion," in *2021 International Conference on Smart Energy Systems and Technologies (SEST)*. IEEE, 2021, pp. 1–6

Others

1. **S. Asefi**, M. Ali, and E. Gryazina, "Optimal energy management for off-grid hybrid system using hybrid optimization technique," in *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*. IEEE, 2019, pp. 1–5

2. T. Amara, **S. Asefi**, O. B. Adewuyi, M. Ahmadi, A. Yona, and T. Senjyu, "Technical and economic performance evaluation for efficient capacitors sizing and placement in a real distribution network," in *2019 IEEE Student Conference on Research and Development (SCOReD). IEEE, 2019*, pp. 100–105

To my beloved parents, whose love, support, and countless sacrifices have been the pillars of my strength and perseverance throughout my academic journey. Thank you for believing in me and for encouraging me to pursue my dreams.

To my dear sister, brother, and brother-in-law, Sima, Soheil, and Masoud, whose companionship and genuine care have enriched my life beyond measure. Thank you for being my constant source of inspiration.

To my precious niece, Sania, and nephew, Samiar, whose infectious joy, boundless curiosity, and innocence have brought so much light and laughter into my life. Thank you for reminding me of the simple joys in life and for inspiring me to always strive for excellence.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my supervisor, Elena Gryazina, whose guidance and assistance have been highly beneficial in shaping my research journey.

I express my heartfelt gratitude to Henni Ouerdane, apart from being the chairman of my doctoral thesis defense session, for his unforgettable kindness and support which have been highly important in my academic journey. I am also deeply grateful to the rest of the jury members (in alphabetic order), Ramesh Bansal, Oleg O. Khamisov, Dmitry Shatov, and Qiuwei Wu for their insightful feedback, constructive criticism, and perspectives that have significantly enhanced my work.

My academic journey would not have been possible without the exceptional guidance of Vladimir Terzija. His profound knowledge, generous personality, belief in my abilities, and unwavering support have profoundly impacted my academic growth and aspirations.

I am profoundly grateful to Dragan Ćetenović for his invaluable technical contribution and dedication throughout my studies. His expertise, willingness to share his knowledge, and patience have been a great advantage in overcoming technical challenges and advancing my research.

A special acknowledgment goes to Helder Leite, who, despite his untimely passing, left an indelible mark on my academic life. His passion for research and enthusiasm continue to inspire me.

The last but not the least, I am deeply indebted to my colleagues and friends who have enriched my PhD journey with their camaraderie, encouragement, and intellectual stimulation. Their presence has made my academic experience truly fulfilling: Hekmat Taherinejad, Ishita Jain, Puskar Pathak, Kripa Adhikari, Mohammad Ebadi, Arman Alahyari, Maryam Majidi, Bakhodur Abdusatorov, Rahim Samanbakhsh, Fahimeh Shiravani, Hamid Zaker, Mile Mitrovic, Behnam Mohseni, Ilia Kamyshev, Sahar Moghimian, Mazhar Ali, Ahmed Baza, Yash Madhwal, Yury Yanovich, and many others that I would not be able to list their names for the sake of brevity.

Contents

List of Figures

List of Tables

Glossary

ADCIT	Anomaly Detection, Classification and Identification Tool
ADI	Anomaly Detection Index
ADMM	Alternating Direction Method of Multipliers
AN	Area Number
API	Application Programming Interface
BC	Blockchain
BD	Bad Data
BDD	Bad Data Detection
CB	Computational Burden
CPPS	Cyber-Physical Power System
CSE	Centralized State Estimation
DApps	Decentralized Applications
DoS	Denial of Service
DSE	Distributed State Estimation
EKF	Extended Kalman Filter
EMS	Energy Management System
EOA	Externally Owned Account
EVM	Ethereum Virtual Machine
FASE	Forecasting Aided State Estimation
FDIA	False Data Injection Attack
FN	False Negatives
FP	False Positives
HVDC	High Voltage Direct Current
ICT	Information And Communication Technology
IEEE	Institute of Electrical And Electronics Engineering
Iter	Number of Iterations
KNN	K-Near Neighborhood
LNR	Largest Normalized Residual
LR	Logistic Regression
ML	Machine Learning
MRMR	Maximum Relevance Minimum Redundancy
OPF	Optimal Power Flow
OT	Overall Elapsed Time
OV	Objective Function Value
PMUs	Phasor Measurement Units
PoS	Proof of Stake

PoW	Proof of Work
RF	Random Forest
RTO	Regional Transmission Organization
SCADA	Supervisory Control and Data Acquisition
SE	State Estimation
SGC	Sudden Generation Change
SLC	Sudden Load Change
SN	State Number
SQP	Sequential Quadratic Programming
SSE	Static State Estimation
tdelay	Data Transmission Delay
TP	True Positives
UKF	Unscented Kalman Filter
WCC	With Modified Convergence Criterion
WLS	Weighted Least Squares
WOCC	Without Modified Convergence Criterion
XGB	Extreme Gradient Boosting

Chapter 1

Introduction

1.1 Challenges of power system state estimation

The power system is a complex and critical infrastructure that ensures the reliable and efficient delivery of electricity to consumers. Power system state estimation plays a pivotal role in monitoring and controlling this infrastructure. However, the power system has never been without challenges. Power systems are becoming increasingly large and interconnected, making accurate state estimation a daunting task. Factors such as the integration of renewable energy sources, the aging of infrastructure, and the potential for cyberattacks further complicate the estimation process. These challenges necessitate the development of advanced techniques and methodologies to enhance the accuracy and robustness of state estimation, ensuring the reliability of the power grid.

1.2 Research objectives and scope

There have been attempts to present solutions for false data injection attack (FDIA) identification in the power system [4]. However, analyzing the effect of typical system events, such as sudden load change (SLC) on the performance of these identification algorithms has not been [5]. Taking into account the application of data-driven algorithms in the power system, in this research an effort has been made to propose a methodology to detect and classify FDIA and SLC, which leads to the following

hypothesis:

The proposed anomaly detection and classification algorithm which is based on combining model-based and data-driven methods are capable of detecting and classifying FDIA and SLC events.

The transmission system faces topology changes (i.e. changes in the network configuration). This will require updating the data-driven model with the new configuration. This process can be time-consuming and inefficient depending on the size of the system which has not been discussed in the literature. To eliminate the adverse effect of topology change, features related to the branches are excluded and only features associated with the buses are applied for the proposed algorithm training. This implies the following hypothesis:

The topology-resilient classification solution effectively mitigates the adverse effects of topology changes, leading to improved robustness in the classification of FDIA and SLC.

Expansion of the power system, communication bottleneck, increase in data size, and security/reliability results in concerns from a centralized state estimation perspective. One of the solutions that can be utilized to mitigate these obstacles is the application of distributed state estimation [3,6–9]. Considering the application of blockchain in the power system [10], it is expected that it improves the distributed scheme from the security perspective. However, it is important to consider that such a distributed solution is iterative. Nevertheless, there was no study analyzing the effects of convergence criterion, partitioning of the system, and delayed (or asynchronous) data transfer on the required number of iterations, and accuracy of the applied method. The mentioned findings give rise to the following hypotheses:

The modified convergence criterion for distributed state estimation methods significantly improves the convergence rates while maintaining the overall performance in an acceptable range compared to traditional convergence criteria.

The optimal partitioning method for distributed state estimation reduces communication overhead, data transfer requirements, and computational time.

The application of blockchain technology for data transfer security effectively safeguards sensitive information and ensures the integrity of data exchanged in the distributed power system state estimation environment.

This research aims to address the power system state estimation challenges through innovative approaches and methodologies. The primary objectives are to improve the accuracy of state estimation, enhance the resilience of the power system against disturbances, and enable the integration of new technologies. The scope of this study involves the development and testing of state-of-the-art algorithms, and the evaluation of these algorithms utilizing the conventional power system models available in the literature. By achieving these objectives, this research endeavors to contribute to the advancement of power system state estimation techniques.

1.3 Research overview and contributions

This research represents a significant contribution to the field of power system state estimation by addressing several key challenges and introducing innovative solutions. The accomplishments of this study are summarized as follows:

Development of Anomaly Detection and Classification Algorithm: A novel algorithm has been developed for the detection and classification of single/multi-bus SLC and single/multi-state FDIA. This algorithm leverages an anomaly detection index for detection and employs supervised machine learning (ML) techniques for the classification of these anomalies.

Topology-Resilient Classification Solution: This research proposes a solution for the classification of FDIA and SLC events that effectively mitigates the adverse effects of topology changes. By focusing on features associated solely with buses, such as nodal measurements, normalized measurement innovations, estimated and predicted values of measurements, and system states, this solution ensures robust classification.

Evaluation of Distributed SE Methods: The application of a modified convergence criterion to recent and well-known distributed state estimation methods,

using IEEE standard test systems, has been thoroughly examined. This evaluation encompasses various performance metrics, including data transfer requirements, computation time, solution accuracy compared to centralized methods, and convergence rates, ultimately leading to the selection of the most effective method.

Optimal Power System Partitioning: An optimal partitioning method has been introduced to reduce communication overhead and data transfer requirements in distributed SE. This optimization significantly decreases the number of iterations needed to achieve a reasonable solution.

Enhanced Data Transfer Security: The research explores the application of blockchain technology to enhance the security of data transfers within distributed power system SE. This novel approach safeguards sensitive information and ensures the integrity of data exchanged in the distributed environment.

Consideration of Asynchronous and Delayed Data Transfer: The study acknowledges the challenges induced by asynchronous and delayed data transfers within the context of distributed SE. By addressing these issues, the research contributes to the development of robust DSE algorithms capable of handling real-world communication constraints.

These accomplishments collectively represent a substantial advancement in the field of power system state estimation, offering practical solutions to address critical challenges and improve the reliability and efficiency of power grid operations.

1.4 Thesis outline

The remainder of this thesis is organized as follows. Chapter 2 provides a literature review, presenting an in-depth analysis of existing state estimation techniques and their limitations. Chapter 3 delves into the power system components modeling and details of the measurement data. Chapter 4 discusses various SE algorithms and methodologies employed in this research. Chapter 5 discusses the results and their implications, while chapter 6 focuses on showcasing the practical application of the proposed methods. Finally, Chapter 7 summarizes the key findings of this study and offers recommendations for future research directions.

1.5 Summary

In this chapter, an overview of the challenges in the power system state estimation was presented. Moreover, research objectives together with a brief background about the main problems of power system state estimation was stated. Then the research gaps were pointed out and the hypotheses were formulated. Finally, the contributions of the research work were highlighted.

Chapter 2

Literature review

This chapter provides a comprehensive review of the power system state estimation techniques. In this chapter, different state estimation algorithms, methodologies, and approaches are discussed. Amongst these literature the research gaps and the areas that can be improved are pointed out.

2.1 Power system state estimation

Power transmission systems contain large numbers of substations. These substations are connected to one another via transmission line and transformers. Moreover, various measurement devices and protection equipment are utilized within the transmission systems, for the purpose of control and protection of these systems [11]. These systems are categorized as a dynamic system, taking into account that the demand and the generation have intermittent nature and are affected by numerous factors. As an example for these factors, increase in utilization of the distributed energy sources and the non-stop escalating level of energy demand, can be stated [12]. The dynamic behaviour of the system, causes different challenges from operational prospective. It is clearly obvious that operating this bulk and dynamic system is not trivial, and it is essential to monitor the state of the system in real-time.

State estimation (SE) is the core component of the energy management system (EMS) for power grids and it plays an essential role in justifying and regulating system operator decisions like load forecasting, contingency analysis, optimal power

flow, etc. The idea of applying SE in the power systems was initiated by Fred Schweppe [13]. SE provides the most likely values of the voltage magnitudes and phase angles for all buses in the power system. The accuracy of these values is essential for achieving optimal and secure operation of the system [7]. It is to be noted that advanced SE can improve monitoring and controlling the power grid in case of a contingency. Especially for the smart grids in which bidirectional transfer of electrical energy and system/consumer data increases the complexity [10, 14]. Such a system can be divided into two integrated parts, i.e., physical equipment of a traditional power system (Physical part) and telecommunication equipment (Cyber part). Combining these two parts will lead to a cyber-physical power system (CPPS) [15]. Figure 2-1 demonstrates the connection between physical, communication and energy management systems.

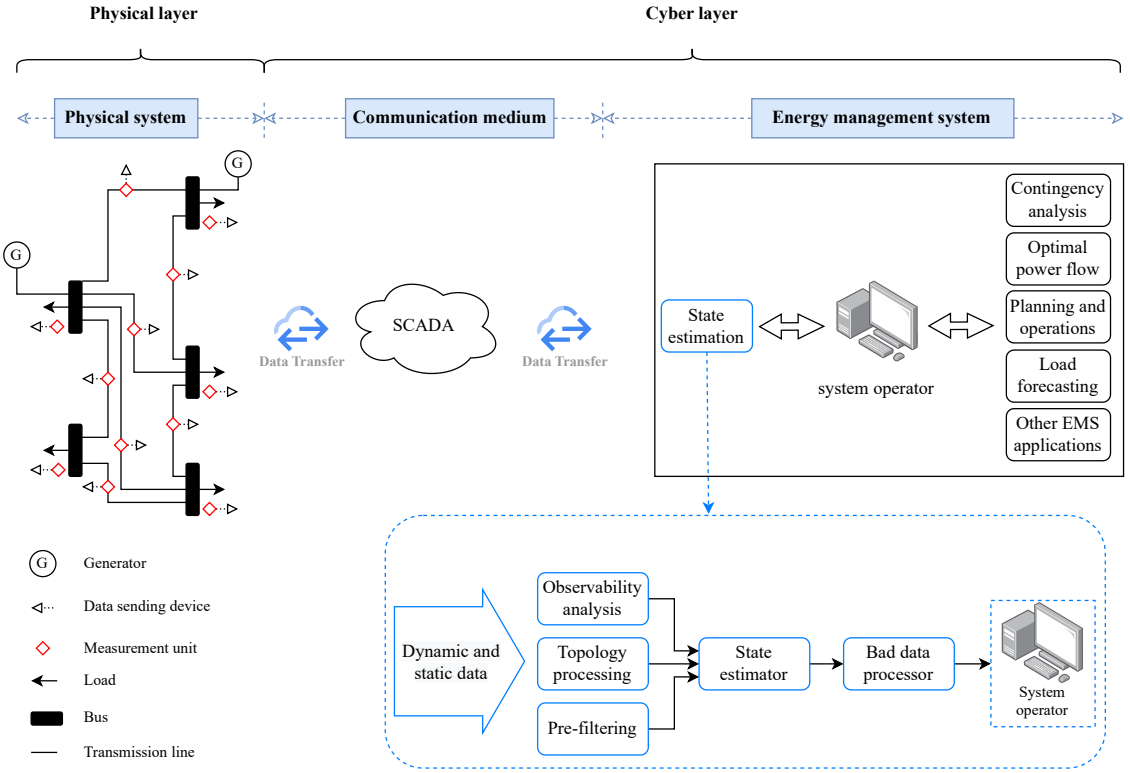

Figure 2-1: State estimator's role in the power system

As it is shown in the Fig. 2-1, typically include the following functions:

- *Dynamic and static data*: The first stage in the process of SE is receiving data from the network. These data can be dynamic, like measurement values

that can change with time. Or they can be static, like parameters of the transmission line.

- *Observability analysis*: Determines if the available number of measurements are sufficient to obtain the SE solution.
- *Topology processing*: Configures the system physical layout, taking into account the status of the switching devices such as circuit breakers.
- *Measurement pre-filtering*: Contains a set of elementary inspections to remove the measurement values that are obviously wrong (e.g., power flow values beyond the limits, voltage magnitudes with negative value, etc.)
- *State estimator*: Calculates the optimal network's state based on the obtained network parameters, measurement data, and network topology.
- *Bad data processing*: Detects existence of non-Gaussian errors within the measurements, based on the SE statistical properties.

Although SE is highly comparable to the conventional load flow, it considers the unpredictable errors that might originate due to unexpected system changes, meters or communication system, inaccuracy in equipment calibration, planned manipulation from a malicious attacker, etc. [10, 13]. Additionally, conventional load flow analysis does not consider redundancy and imprecision of the system's measurement data, whereas SE considers the mentioned features [13].

Considering a brief background of the evolution of the SE method and its application in power system, it is worth noting that as soon as Schweppe pointed out the application of SE in power system, it attracted industrial communities' attention [16]. Growth of the power system due to increase in the level of needed electricity consumption and propagation of the communication technology, bring in several problems assigned with power system operation, especially centralized SE such as:

- *Expansion of power system continent-wise*
- *Policy and privacy*

- *Dimension of the grid*
- *Communication bottleneck*
- *Data size*
- *Security/Reliability*

Expansion of the power grid over continents makes an interconnected system such that these continents can be affected by contingencies in other ones [9]. Although, in some regional expansion cases, e.g., the case with regional transmission organizations (RTOs) in Europe, operators are using high-voltage direct current (HVDC) technology for power transfer which is also another research area for considering hybrid HVDC/AC SE so that they can meet the characteristics of the new network regarding Supervisory Control and Data Acquisition (SCADA) system [17]. Vulnerability and inflexibility of centralized SE make it unsuitable for a multi-area (or multi RTO) estimator from policy and privacy point of view [9]. The grid's high dimension is another challenge that affects the computational difficulty [18,19]. Having only one central control unit, extensive network parameters and measurement unit's information, which needs to be transferred to this unit, may result in communication bottlenecks [6,20]. Another problem that has attracted the researcher's attention, especially in smart grids, is that the size and the speed of receiving data (so called big data) from measurement units might be infeasible to be stored and processed [10,21]. Moreover, in most of the literature, it is assumed that the central node is secure, though it can be the most vulnerable, insecure, and unreliable point in a network and prone to a single point of failure [10,22].

2.2 Power system anomalies

Being a cyber-physical system, the power system is vulnerable to various types of anomalies, such as cyberattacks. Adversaries can exploit vulnerabilities in the SE system to inject false data, which can lead to inaccurate state estimates. This can have a number of negative consequences, such as:

- Cascading power outages
- Economic losses
- Safety hazards

Three main security features for the data in a smart grid are *Confidentiality*, *integrity* and *availability*, which they refer to occasions when the data are accessible only to authorized users, the data are trustworthy in any operational circumstances, and the data are promptly and reliably available, respectively [23].

Cyberattacks, such as the denial of service (DoS) or false data injection (FDIA), aim to deteriorate such properties. Noting the case when a central control unit gets compromised, all data can either get lost or controlled by the attackers [24], and one of the potential solutions could be a distributed control scheme. However, the distributed grid can also be subjected to a cyber-attack, i.e., attack to measurement units, to control centers, to communication line between control centers and measurement units, to communication line between control centers (i.e., between areas).

Power system SE can be subjected to different types of anomalies that might spoil the accuracy of the estimated states. Inter alia, these can be anomalies like bad data (BD), sudden change in bus injections or false data injection attack (FDIA). BD is caused by unexpected errors in sensors or communication medium. Besides, network model parameters might contain BD. Sudden change in bus injections can be either sudden load change (SLC) or sudden generation change (SGC), depending on whether consumer or generator is connected to the bus. Severe SLC is usually caused by serious variations in industry load or by disconnection/reconnection of a large portion of the load. Although SLC was under the scope of many research work in the past, SGC emerges as a new challenge since penetration of uncertain renewable energy sources is increasing incessantly [25, 26]. Both SLC and SGC will further lead to the sudden change in the system operating point, i.e., it will cause a large and rapid change in most state variables. FDIA is a type of perfect interacting BD [27]. FDIA is amongst the most hazardous cyberattacks which targets data integrity. It has attracted industry and research community's attention recently [28–

31]. Regardless of the anomaly type, it is essential to detect the anomaly presence as soon as it occurs, as well as to classify (discriminate between) different types of anomalies and identify their origin in order to enable proper countermeasures against each of them in the correction phase (see Fig. 5-1).

Techniques for detection and classification of anomalies depend on which SE method is implemented. Assuming that normalized measurements' residuals follow a standard Gaussian distribution, their sum of squares are expected to follow a χ^2 distribution [11, 14]. This resulted in χ^2-test to be widely used as BD detection method in conventional weighted least square (WLS) power system static state estimation (SSE). Besides, largest normalized residual (LNR) test has shown good detecting capabilities but also the ability to identify measurement(s) corrupted with bad data [11]. LNR test found its role in bad data detection stage for hybrid Voltage Source Converter - HVDC/AC transmission systems, where it has been improved by integrating the Gaussian mixture model algorithm [32]. LNR method is also used to cope with bad data in multi-energy applications [33]. Other approaches suitable for WLS framework are recently studied to mitigate the impact of outliers, like Hampel's redescending and the Schweppe–Huber generalized M-estimators [34]. Unlike WLS, forecasting-aided state estimation (FASE) based on Kalman filtering utilizes process model yielding state forecasting. FASE combines state predictions with observed measurements which facilitates BD detection and classification. To detect BD in phasor measurement units (PMUs) a robust unscented Kalman filter (UKF) is presented by processing the predicted state vector and the received measurements simultaneously [35]. A fault tolerant second-order extended Kalman filter (EKF) based on discrete-time nonlinear Luenberger-type observer has been presented in [36] to mitigate the adverse effects of BD.

Considering WLS SSE employs only current snapshot of measurements, χ^2-test conducted over measurement residuals will not recognize the presence of the SLC. On the other hand, due to existence of the process model, SLC will affect FASE performance. This might be used to enable SLC detection. In [25, 26], normalized measurement innovations are tested against predefined threshold to detect anomaly presence, while skewness of distributions and χ^2-test of goodness of fit of normal-

ized innovations are proposed to discriminate between BD and SLC. In [37], same detection technique has been used whilst the logical check routine is applied for classification purpose. In [38], an UKF based on minimum error entropy is presented to avoid abnormalities such as SLC. A maximum-correntropy-based EKF estimator is presented in [39] for discrimination between BD and SLC, incorporating both supervisory control and data acquisition (SCADA) and PMU measurements. In [40], an UKF with generalized correntropy loss is introduced to suppresses the effect of outliers by utilizing inverse of the exponential function of innovations for update of measurements noise covariance matrix.

2.3 Machine learning in power system

As discussed above, determining an analytical approach for detection and classification of the discussed three anomalies is a difficult task. Nowadays, similar to other research areas, application of machine learning (ML) algorithms in power systems is increasing as well. First reason is their low mathematical dependency on system models; instead, ML algorithms extract knowledge directly from the data. Next reason is self-learning capability of these algorithms, enabling the algorithms to learn from experience and update their knowledge to give better results [41]. Additionally, proper accuracy and effectiveness makes ML based methods a suitable choice for detection and classification of anomalies [42]. Apart from supervised ML methods commonly utilized in the literature, such as logistic regression, k-near neighborhood, random forest and/or extreme gradient boosting, combination of artificial neural networks and ML, so called deep learning, is also broadly applied in the area of power systems. In [43], a deep neural network architecture that integrates a universal BD detection technique using a binary hypothesis testing scheme has been presented. A matrix completion approach is proposed for SE of distribution networks in [44], aiming to minimize the weighted sum of the measurement residuals to suppress the effects of BD. Bayesian BD detection method within a deep learning based SE scheme is used in [45]. An ensemble correlation based detector with adaptive statistics, presented in [46], compares squared Mahalanobis

distance of new measurement samples with an adaptive threshold in order to detect and classify FDIA. Authors in [47] have introduced two ML based algorithms for FDIA detection: if the measurements are labeled (i.e. normal and tampered measurements are specified), detection is performed using support vector machine algorithm; in the case of unlabeled measurements, detection is done utilizing statistical characteristics of historical measurements. A deep learning based method utilizing a feed-forward artificial neural network has been implemented for FDIA detection in [48]. In [43–48], detection and classification of SLC and multi-state FDIA, as well as analyzing the effects of network topology change on detection and classification of FDIA and SLC, have not been discussed. In another words, the algorithms in the mentioned literature are not trained to deal with such anomalies.

2.4 Power system distributed state estimation

One way to overcome centralized SE issues would be implementing the distributed state estimation. In distributed SE, the power system will be divided into several smaller areas or sub-systems, and the SE process will take place concurrently in each area. A low amount of information exchange at borders of the areas is required so that each area reaches convergence, i.e., the distributed network reaches a similar solution as the centralized one. The amount of information that must be exchanged depends on the method applied. In [49], a detailed comparison of the recent distributed SE methods regarding indices such as convergence rate and information exchange has been made, which clearly confirms that each method varies from one another considering information transfer between areas.

The distributed SE algorithms can be classified into two categories, having a global control center, i.e., hierarchical distributed SE [50–52] or fully distributed [6,8,49]. Both of them are successful in reaching to an acceptable solution compared to centralized algorithms. Alternating direction method of multipliers (ADMM) [53] that are in the category of distributed optimization [9], have been very popular recently. In [54], a fully decentralized adaptive SE scheme has been presented for the power system via applying the network gossiping method. The method enables

collaboration between areas to solve the global problem; however, there is still a significant performance error in comparison to centralized SE. Authors in [55] presented a distributed SE for wide area monitoring system, which does not need local observability of all areas. In [56], a new multi-area SE method is discussed that utilizes a central coordinator; however, there is no need to exchange topology information between areas or from areas to the central coordinator. The proposed approach in [57] is a new hierarchical multi-area power system SE, which shares the sensitivity function of local estimators instead of boundary measurements or state estimates. As stated by the authors in [57], the approach reduces the information exchange, as well as increases convergence speed. In [22], the authors have provided an ADMM based distributed SE. Also, in [3] and [6], a distributed SE process using matrix splitting method for DC and AC SE, respectively. For more details, we refer to [7] that presents a brief review of multi-area SE.

It is to be noted that mostly in the literature, the transmission system has been a matter of concern, which we have followed the same approach. To solve AC SE via centralized method or some of the distribution methods, such as matrix splitting or ADMM, would need linearization of the problem using Newton's method. However, by applying the decomposition method [9] and the available solvers, there would be no further need for linearization of the problem.

2.5 Blockchain

Blockchain (BC) is a peer-to-peer distributed ledger technology that stores data on multiple servers globally. In 2008, Satoshi Nakamoto's whitepaper on Bitcoin [58] pioneered the use of BC technology in financial application [59]. BC technology was primarily used in the financial domain, so as to provide trustfulness and a secure environment without central authority where digital assets like cryptocurrency can be prevented from double-spending attacks [60, 61]. Since then, the technology's potential has moved beyond financial domains to different sectors like supply chain management, healthcare, etc. [62–66].

BC is a distributed ledger of chronologically generated blocks containing cryp-

tography linked blocks to the previous block forming a chain. Any modification to the previous existing block will be reflected on every subsequent block, making it secure and immutable to modification. If an attacker changes data on any of the previous blocks, the following block's data will also change, and the ledger can be compared with another copy to track the point at which the data was manipulated and later rectified. In cryptocurrency, it is computationally hard to take control over the BC network because the attacker will require 51% of the network's computing power, i.e., it will be difficult for an attacker to fork from a past block and mine blocks faster, surpassing current (honest chain) block height. This will create double spending, which computationally hard [67]. BC-based applications can provide security, trust, economic, and auditability [68].

Since Bitcoin, many alternative cryptocurrencies (altcoins) have emerged. Ethereum [69] is the most popular cryptocurrency after Bitcoin, which provides an open-source platform to develop BC-based decentralized applications (DApps). DApps are application programs that runs on decentralized BC applications using Ethereum Virtual Machine (EVM). For example, smart contracts can specify the functionality and condition, under which circumstances payment can occur between two individuals. These conditions are programmed and deployed on the BC, and individuals can abide by these conditions and transact in a secure environment without intermediaries. EVM is one big computer that is made of small individual computers located globally. These computers are nodes connected, having a copy of the Ethereum BC. The transactions are broadcasted to the network via a node which is replicated across the network. For feasibility demonstration of BC for secure data transmission, we have developed a prototype on the Ethereum platform using truffle framework [70] which can be deployed on local machines. We have created a smart contract, specified conditions and deployed it on a BC network running on local devices.

Aside from financial applications of BC, it has been developed in other fields. For example, in [71], the authors propose a BC based method to preserve security of the spectrum sharing between aerial (unmanned aerial vehicle as a component of next generation cellular network) and terrestrial communication systems. Application of BC in the smart grid mainly has been investigated in the area of power markets, i.e.,

the issues related to the secure energy transactions [12]. In [72], a proof of concept (PoC) for decentralized energy trade using BC has been proposed, to enable peer to peer energy transactions. In [73] a BC based platform for solar energy trade amongst prosumers has been implemented in laboratory scale. However, a few works have been applied BC in power system for security purposes, [74, 75]. These studies consider storing system wide measurement data in each measurement, which seems inefficient due to low memory of measurement units and time delay caused by encryption/decryption of the data.

2.6 Summary

A comprehensive literature review of the power system state estimation concept and its techniques (i.e., centralized and distributed) was provided in this chapter. Furthermore, various types of anomalies that may occur in the power system was discussed. Additionally, recent applications of contemporary methods in power systems, namely machine learning and blockchain were investigated. Table 2.1 presents an overview of the literature together with their challenges.

Table 2.1: Summary of the literature overview

Category	Common challenge	Citations
Power system state estimation	**Centralization**: Centralized SE is prone to single point of failure and communication bottlenecks. **Data size**: Handling large amount of data from sensors and measurement units can be challenging. **Security**: Centralized SE can be vulnerable to cyber-attacks.	[9–11, 22, 76]

Power system anomalies	**Detection and classification**: Accurately detecting and classifying different types of anomalies, such as bad data, sudden change in bus injections, and false data injection attacks, despite its importance can be difficult.	[11, 48, 77, 78]
Machine learning in power system	**Data availability and quality**: The availability and quality of data for training machine learning models can be limited or inaccurate. **Model complexity**: Designing complex machine learning models that can handle the complex nature of power systems is challenging. **Interpretability**: Explaining the decisions made by complex machine learning models can be difficult.	[4, 14, 40, 43, 44]
Power system distributed state estimation	**Convergence speed and accuracy**: Distributed SE algorithms need to converge quickly and accurately to provide reliable state estimates. **Information exchange**: The amount of information exchanged between areas in distributed SE needs to be carefully controlled to optimize communication and computational resources.	[6, 22, 49, 55, 79]

Blockchain	**Scalability**: The scalability of blockchain technology needs to be improved to handle the large amount of data and transactions in the power system. **Performance**: The performance of blockchain-based applications needs to be optimized to ensure real-time operation and response times. **Security**: Blockchain technology needs to be further hardened against cyberattacks and vulnerabilities.	[65, 74, 75, 80]

Chapter 3

Power system modeling and measurement data

Power system state estimation refers to the procedure of obtaining the voltage phasors of all buses in the system using a set of redundant measurements. The redundancy in the measurement units will avoid vulnerability to error in measurement units and in telecommunication system. Apart from conventional power and voltage measurements, the measurement set may include current or synchronized voltage phasor measurements as well. A certain amount of time skew between measurements is commonly tolerated. Because it is practically impossible to obtain the measurements from different parts of the network simultaneously.

Another requirement for the implementation of state estimation is that the network topology and parameters must be perfectly known. This chapter discusses the component model that represents the entire network.

3.1 Modeling of the network components

3.1.1 Transmission line

A two port π-model is utilized to characterize the positive sequence of the transmission line. It is assumed that the power system is operating in steady state mode and under balanced condition. This requires that all transmission lines are fully

transposed, all bus loads and branch power flows will be three phase and balanced, and all other series or shunt devices will be symmetrical in the three phases. The equivalent circuit presented in Fig. 3-1 which demonstrates a transmission line with a positive sequence series impedance of $R + jX$ and total line charging susceptance of $j2B$.

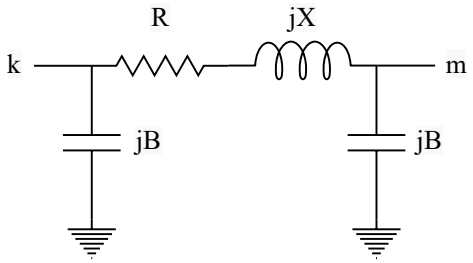

Figure 3-1: Two port π-model equivalent circuit of a transmission line

3.1.2 Shunt capacitors or reactors

To control the voltage or reactive power in the network, shunt capacitors or reactors might be installed. The type of the shunt element can be determined by the sign of the susceptance value. For a shunt capacitor the value will negative and for a reactor the value will be positive.

3.1.3 Loads and Generators

Power injections, such as loads and generators, are modeled as complex power injections, which do not affect the network model. However, constant impedance type loads are included as shunt admittances at the corresponding buses.

3.2 Network modeling

The component models described above can be used to create a model of the entire power system. This is done by writing a set of equations for each node in the system,

using Kirchhoff's current law. These equations can be written in the following form:

$$\boldsymbol{I} = \begin{bmatrix} i_1 \\ i_2 \\ \vdots \\ i_N \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{12} & \dots & Y_{1N} \\ Y_{21} & Y_{22} & \dots & Y_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ Y_{N1} & Y_{N2} & \dots & Y_{NN} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix} = \boldsymbol{Y}.\boldsymbol{V} \tag{3.1}$$

where

$\boldsymbol{I}$ is the net current injection phasors vector

$\boldsymbol{V}$ is the bus voltage phasors vector

Y_{km} is the (k,m)th element of the $\boldsymbol{Y}$ matrix

N is the number of bus

i_k is the current injection phasor in bus k

v_k is the voltage at bus k

3.3 Measurement data

Supervisory control and data acquisition (SCADA) systems and phasor measurement units (PMUs) are two types of measurement systems that are commonly used for state estimation in power systems. SCADA systems collect data from various sensors and devices in a power system, such as voltage and current sensors, circuit breakers, and switchgear. The data collected by SCADA systems is typically used to monitor the state of the power system and to control and operate the system. SCADA systems can also be used to collect data for state estimation.

PMUs are specialized devices that measure the phasor of voltage and current at a specific location in a power system. Phasors are complex numbers that represent the magnitude and phase angle of a sinusoidal signal. PMUs are very accurate and can measure phasors at a high sampling rate. This makes PMUs ideal for state estimation applications.

SCADA and PMU measurements can be used together to improve the accuracy and reliability of state estimation. SCADA measurements are typically more widely available than PMU measurements, but PMU measurements are more accurate and can provide more information about the state of the power system.

Table 3.1 summarizes the key differences between SCADA and PMU measurements:

Table 3.1: summarization of the SCADA and PMU measurements

Characteristic	**SCADA**	**PMU**
Accuracy	Less accurate	More accurate
Sampling rate	Lower sampling rate	Higher sampling rate
Availability	More widely available	Less widely available
Cost	Lower cost	Higher cost

In general, SCADA measurements can provide a wide range of data about the state of the power system, while PMU measurements can provide more accurate and detailed information about the state of the power system at specific locations. In this study, only SCADA measurements have been considered. These measurements are composed of active and reactive power flows, active and reactive power injections, and voltage magnitudes.

While using SCADA measurements for state estimation, it is important to consider the following sources of error.

Measurement noise is any error in a measurement that is not due to the true value of the quantity being measured. Measurement noise can be caused by a variety of factors, including:

- Sensor errors: Sensors are not perfect and can introduce errors into measurements. Sensor errors can be caused by a variety of factors, such as manufacturing defects, environmental factors, and aging.

- Communication errors: Measurements are often transmitted from sensors to the state estimator over communication channels. Communication errors can introduce errors into measurements. Communication errors can be caused by a variety of factors, such as noise on the communication channel, interference, and cyberattacks.

- Environmental factors: Environmental factors, such as temperature, humidity, and vibration, can also introduce errors into measurements.
- Human errors: Human errors can also introduce errors into measurements. Human errors can occur during the installation, calibration, and operation of sensors.

Measurement noise can have a significant impact on the accuracy of state estimation. By understanding the different factors that can cause measurement noise, it is possible to take steps to mitigate its impact on state estimation.

Here are some specific things that can be done to reduce the impact of measurement noise on state estimation:

- Use high-quality sensors and communication equipment.
- Calibrate sensors regularly.
- Use a state estimation model that is as accurate as possible.
- Use an error modeling technique that is appropriate for the application.
- Detect and correct bad data as quickly as possible.

In addition to the above measures, it is also important to be aware of the specific sources of measurement noise in the power system being monitored. For example, if the power system is subject to a lot of lightning strikes, then it is important to use sensors that are resistant to lightning strikes.

By taking these steps, it is possible to reduce the impact of measurement noise on state estimation and improve the accuracy of state estimates.

3.4 Studied networks

In this study, two IEEE standard networks have been utilized. The IEEE 14-bus system and the IEEE 118-bus system which are presented in Fig. 3-2 and Fig. 3-3, respectively. The IEEE 14-bus system is a small power system that is often used for

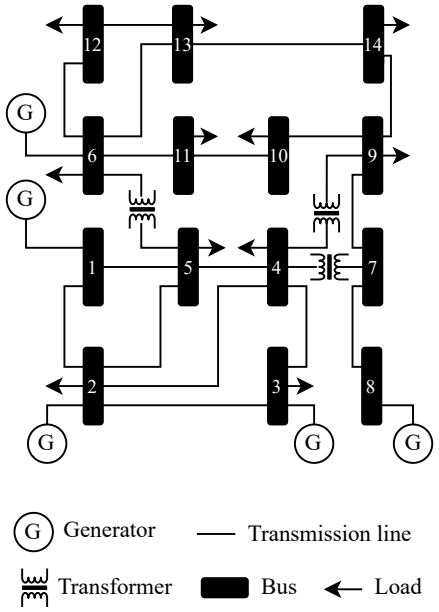

Figure 3-2: Single line diagram of IEEE 14 bus system [1]

educational and research purposes. The IEEE 118-bus system is a larger and more complex power system that is more representative of real-world power systems [1,2].

Both the IEEE 14-bus system and the IEEE 118-bus system are commonly used for state estimation studies. The IEEE 14-bus system is a good choice for simple state estimation studies, while the IEEE 118-bus system is a good choice for more complex state estimation studies. Table 3.2 summarizes the key features of the IEEE 14-bus system and the IEEE 118-bus system:

Table 3.2: IEEE 14 and IEEE 118 bus system comparison

Characteristic	**IEEE 14-bus system**	**IEEE 118-bus system**
Number of buses	14	118
Number of generators	5	19
Number of loads	11	91
Number of lines	20	177
Complexity	Simple	Complex

3.5 Weighted Least Squares

The SSE results from a further simplification in which the state transition information is completely disregarded and only the nonlinear measurement function is

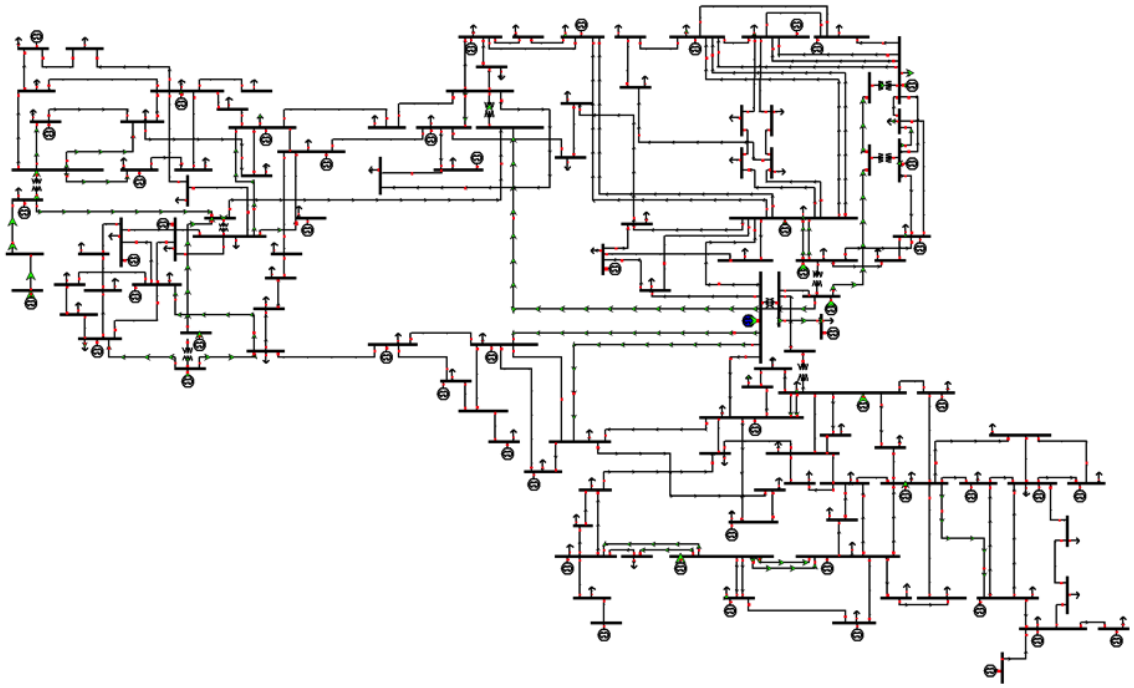

Figure 3-3: Single line diagram of IEEE 118 bus system [2]

kept [4]. As a result, SSE has no recollection of the states from the previous time steps and, unlike FASE, it lacks the capacity to track the system state transition [81]. Thus, SSE implies that the state vector is estimated using the most recent set of available measurements.

Based on the concept of maximum likelihood, and assuming the measurement errors are independent and follow Gaussian distribution [11, 14], SSE problem can be written in the following format:

$$min\ [\boldsymbol{z}_t - \boldsymbol{h}(\boldsymbol{x}_t)]^T \boldsymbol{R}_t^{-1} [\boldsymbol{z}_t - \boldsymbol{h}(\boldsymbol{x}_t)] \tag{3.2}$$

$$subject\ to \qquad \boldsymbol{z}_t = \boldsymbol{h}(\boldsymbol{x}_t) + r_t$$

where $\boldsymbol{z}_t$ is the vector of measurements at time step t; $\boldsymbol{h}$ is the nonlinear function relating measurements to the state vector $\boldsymbol{x}_t$; $\boldsymbol{R}_t$ indicates measurement noise covariance matrix which is a diagonal matrix and its elements are composed of each measurement's standard deviation (σ), $\boldsymbol{R}_t = diag\{\sigma_{1,t}^2, \sigma_{1,t}^2, \ldots, \sigma_{m,t}^2\}$ The optimization problem (3.2) is known as weighted least squares (WLS) estimation and can be

solved using Gauss-Newton iterative process [14]:

$$\boldsymbol{x}_t^{k+1} = \boldsymbol{x}_t^k + [[\boldsymbol{H}_t^k]^T \boldsymbol{R}_t^{-1} \boldsymbol{H}_t^k]^{-1} [\boldsymbol{H}_t^k]^T \boldsymbol{R}_t^{-1} [\boldsymbol{z}_t - \boldsymbol{h}(\boldsymbol{x}_t^k)] \tag{3.3}$$

where k is iteration counter and $\boldsymbol{H}^k$ represents Jacobian of $\boldsymbol{h}$ evaluated at $\boldsymbol{x}^k$. After convergence, vector of measurement residuals $\boldsymbol{r}$ and its covariance matrix $\boldsymbol{\Omega}$ can be calculated as:

$$\boldsymbol{r}_t = \boldsymbol{z}_t - \boldsymbol{h}(\hat{\boldsymbol{x}}_t) \tag{3.4}$$

$$\boldsymbol{\Omega}_t = \boldsymbol{R}_t - \hat{\boldsymbol{H}}_t \left[\hat{\boldsymbol{H}}_t^T \boldsymbol{R}_t^{-1} \hat{\boldsymbol{H}}_t \right]^{-1} \hat{\boldsymbol{H}}_t^T \tag{3.5}$$

where $\hat{\boldsymbol{H}}$ represents Jacobian of $\boldsymbol{h}$ evaluated at $\hat{\boldsymbol{x}}$. More detailed explanation regarding WLS is provided in chapter 4.

3.6 Bad data

If the normalized measurement residuals follow Standard Gaussian (Normal) distribution, their sum of squares will have a χ^2 distribution. Statistical properties of the normalized measurement residuals have been considered in Appendix A. Combination of χ^2-test and LNR test is widely used within the WLS SSE framework for bad data detection (BDD) and identification of BD's origin. Steps are given as follows [14]:

- Calculate the following objective function after solving the SE problem (time index t is omitted to simplify the notation) [82]:

$$J_{BDD}(\hat{\boldsymbol{x}}) = \sum_{i=1}^{m} \frac{[z_i - h_i(\hat{\boldsymbol{x}})]^2}{\Omega_{ii}} = \sum_{i=1}^{m} \frac{r_i^2}{\Omega_{ii}} \tag{3.6}$$

  where z_i is i-th measurement, h_i is i-th equation from set $\boldsymbol{h}$ of measurement equations, and Ω_{ii} is i-th diagonal element of $\boldsymbol{\Omega}$.

- From χ^2 distribution table pick up the suspicion threshold corresponding to probability p and $(m - n)$ degrees of freedom [11].

- If $J_{BDD}(\hat{\boldsymbol{x}}) \geq \chi^2_{(m-n),\rho}$ holds, BD exists with confidence probability level ρ and $(m-n)$ degrees of freedom; otherwise, there is no BD.

- If BD is detected, calculate normalized residual for each measurement, r_i^{Norm}, as:

$$r_i^{Norm} = \frac{|z_i - h_i(\hat{\boldsymbol{x}})|}{\sqrt{\Omega_{ii}}} = \frac{|r_i|}{\sqrt{\Omega_{ii}}} \tag{3.7}$$

- If r_i^{Norm} is the largest normalized residual and $r_i^{Norm} > \tau$, where τ is a threshold, then i-th measurement will be suspected as BD. For normalized residuals following Standard Gaussian distribution, threshold τ can be selected as $\tau = 3$. Statistical properties of normalized residuals are further discussed in Appendix A.

3.7 Sudden load/generation change

One of the events that might change the power system state abruptly is SLC or SGC. This might happen due to failure of different power system components, such as circuit breakers or generation units. Considering constant increase in penetration from renewable energy sources into the power system, their intermittent nature might be another reason for sudden state change. It is to be noted that χ^2-test carried out over measurements' residuals obtained via WLS SSE is unable to detect SLC/SGC. On the other hand, application of FASE can be helpful for detection of these events due to advantages that state transition model brings. For the sake of brevity, in this research we have focused on SLC and modeled it as a load shedding at different buses but we point out that similar considerations can be applied to the case of SGC as well. More details about SLC modeling is provided in Section 5.1.7.

3.8 False data injection attack

With the evolvent of the power systems and application of various communication mediums, the possibility for cyber-attacks has increased. Considering the power system as cyber-physical system, its data must satisfy three main principles of in-

formation security, so called availability, integrity and confidentiality. Two well-known attacks in the power system are FDIA and denial of service (DoS), which threaten integrity and availability of the data, respectively [23]. Although the targeted medium for both types of attack is the communication medium, FDIA can lead to critical issues to the secure and economic operations in the power system if it evades being detected by conventional BDD. FDIA can mislead the system operator that the system operates in a normal and secure state, while in reality it is not. Also, the operator may be persuaded to take expensive and unnecessary actions like load shedding or rescheduling generation units. Conventional BDD, such as χ^2-test, utilizes measurement residuals, while stealthy FDIA endeavors to keep residuals unchanged. This might lead to conventional BDD failure against stealthy FDIA.

Assuming the adversarial has perfect knowledge of the system (topology and parameters) and receives the same data as the system operator, it can be capable of manipulating the measurements in a way that BDD will be bypassed. The attack vector $\boldsymbol{a}$ is of the same size as the measurement vector but with non-zero elements (a_i) corresponding only to measurements under the attack. So, under the attack, the i-th measurement will have the following model:

$$z_i^a = \begin{cases} z_i + a_i & \text{if } i\text{-th measurement is attacked} \\ z_i & \text{otherwise} \end{cases} \tag{3.8}$$

From practical point of view, (3.8) indicates that the adversarial needs to have access to measurements and the ability to manipulate their values to the values determined by the attack vector.

Without losing generality, let us assume that the adversarial has obtained the same measurements $\boldsymbol{z}$ and state estimates $\hat{\boldsymbol{x}}$ as the system operator. The FDIA can bypass the BDD if $a_i = h_i(\boldsymbol{x}^a) - h_i(\hat{\boldsymbol{x}})$, where $\boldsymbol{x}^a = \hat{\boldsymbol{x}} + \boldsymbol{c}$ and $\boldsymbol{c}$ represents the change adversarial would like to make to the system states. Vector $\boldsymbol{c}$ has non-zero elements corresponding to state variables that are intended to be changed by the adversarial, while all other elements are equal to 0. Under perfect FDIA described

above, vector of measurement residuals will be [83]:

$$\begin{aligned} \boldsymbol{r}^a &= \boldsymbol{z}^a - \boldsymbol{h}(\boldsymbol{x}^a) \\ &= \boldsymbol{z} + \boldsymbol{a} - \boldsymbol{h}(\boldsymbol{x}^a) \\ &= \boldsymbol{z} + \boldsymbol{a} - \boldsymbol{h}(\boldsymbol{x}^a) + \boldsymbol{h}(\hat{\boldsymbol{x}}) - \boldsymbol{h}(\hat{\boldsymbol{x}}) \\ &= \boldsymbol{z} - \boldsymbol{h}(\hat{\boldsymbol{x}}) + \boldsymbol{a} - [\boldsymbol{h}(\boldsymbol{x}^a) - \boldsymbol{h}(\hat{\boldsymbol{x}})] \\ &= \boldsymbol{z} - \boldsymbol{h}(\hat{\boldsymbol{x}}) = \boldsymbol{r} \end{aligned} \tag{3.9}$$

The above equation proves the feasibility of stealthy attack in case the adversarial has enough information regarding the network measurements and parameters to build attack vector $\boldsymbol{a}$ in a way that $\boldsymbol{a} = \boldsymbol{h}(\hat{\boldsymbol{x}} + \boldsymbol{c}) - \boldsymbol{h}(\hat{\boldsymbol{x}})$. It is worth noting that in this research the conventional BDD is assumed to be χ^2 method, as explained above. More details of modeling FDIA is presented in Section 5.1.7.

3.9 Summary

In this chapter, the mathematical modeling of the state estimation problem together with the model of network component were presented. Moreover, schematics of the IEEE test systems and the concept of sudden load/generation change and false data injection attacks were discussed.

Chapter 4

State estimation algorithms

In this research work, three different types of state estimation (SE) techniques, i.e., centralized, distributed, and quasi-steady are studied. Centralized SE is the most common type of SE. In centralized SE, all of the measurements are collected and processed by a central computer. Centralized SE is simple to implement and provides the most accurate state estimates. However, centralized SE can be computationally expensive for large power systems. Distributed SE is a newer type of SE that is designed to address the computational challenges of centralized SE. In distributed SE, the measurements are processed by multiple computers that are distributed throughout the power system. This can significantly reduce the computational burden of SE for large power systems. Forecasting-aided SE is a type of SE that uses forecasting techniques to improve the accuracy of state estimates. Forecasting-aided SE can be used to predict the future state of the power system based on past measurements. This can be useful for detecting and identifying potential problems in the power system before they occur.

This chapter will discuss the advantages and disadvantages of each type of SE, as well as the different algorithms that can be used to implement each type of SE.

4.1 Centralized

Centralized state estimation is the most common type of state estimation used in power systems. In centralized SE, all of the measurements from various sensors

throughout the power system are collected and processed by a central computer (i.e. central data center), as shown in Fig. 4-1.

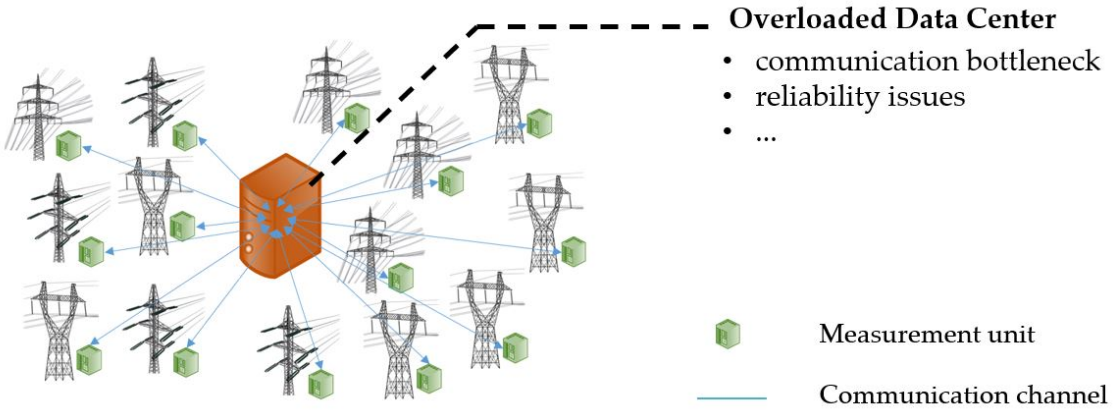

Figure 4-1: Scheme of a centralized state estimator

The central computer uses the measurements to solve a set of nonlinear equations that describe the electrical relationships between the buses in the system. The solution of these equations provides an estimate of the state of the power system, which is typically defined as the set of voltage and current phasors at all buses in the system.

Centralized SE has a number of advantages, including:

- Simplicity: Centralized SE is relatively simple to implement.
- Accuracy: Centralized SE provides the most accurate state estimates.

However, centralized SE also has some disadvantages, including:

- Computational cost: Centralized SE can be computationally expensive for large power systems.
- Communication requirements: Centralized SE requires a reliable communication infrastructure to transmit the measurements from the sensors to the central computer.
- Vulnerability to cyberattacks: Centralized SE is vulnerable to cyberattacks. An attacker could compromise the central computer to inject false data, which

could lead to inaccurate state estimates and disruptions to power system operation.

Despite its disadvantages, centralized SE is the most widely used type of SE in power systems today. This is because centralized SE provides the most accurate state estimates and is relatively simple to implement. One of the well-known methods of centralized SE is the maximum likelihood method discussed below.

4.1.1 Maximum likelihood method

Maximum likelihood estimation (MLE) is a method of estimating the parameters of a statistical model by finding the values of the parameters that maximize the likelihood of the observed data. In other words, MLE finds the set of parameters that makes the data most likely to have occurred. This is done by calculating the likelihood function, which is the probability of the observed data given the model parameters. The parameters are then varied to find the values that maximize the likelihood function

The goal of state estimation is to figure out the most likely state of the system based on what we can measure. One way to do this is by using MLE.

Measurement errors are often assumed to be distributed normally, which means that they follow a bell-shaped curve. The two main parameters of a normal distribution are the mean (μ) and the variance (σ^2).

The normal probability density function (pdf) for a random variable z, with mean (i.e., expected value of z) μ, and standard deviation σ is defined as [11]:

$$f(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\{\frac{z-\mu}{\sigma}\}^2} \tag{4.1}$$

There are commonly accepted statistical assumptions about measurement errors, which are not always valid:

- Errors are distributed according to a normal distribution.
- The expected value of errors is zero.
- Errors are independent.

The third assumption, which is based on considering errors independence, implies that the joint pdf of a set of m measurements can be obtained by simply taking the product of individual pdfs corresponding to each measurement. The resulting product function $f_m(z)$ given by:

$$f_m(z) = f(z_1)f(z_2)...f(z_m) \tag{4.2}$$

that is called the Likelihood Function for the set of m measurements.

For the sake of simplicity, logarithm of likelihood function will be used. So, we have:

$$L = \sum_{i=1}^{m} \log f(z_i) = -\sum_{i=1}^{m} \log \sigma_i - \frac{m}{2}\log 2\pi - \frac{1}{2}\sum_{i=1}^{m}(\frac{z_i - \mu_i}{\sigma_i})^2 \tag{4.3}$$

This is an optimization problem, which can be formulated as:

$$max\ L\ \equiv max\{-\frac{1}{2}\sum_{i=1}^{m}(\frac{z_i - \mu_i}{\sigma_i})^2\} \equiv min \sum_{i=1}^{m}(\frac{z_i - \mu_i}{\sigma_i})^2 \tag{4.4}$$

let's define $W_{ii} = \sigma_i^{-2}$ (σ_i^2 is covariance of measurement unit i and W is inverse covariance matrix).

$$\begin{gathered} min\ \sum_{i=1}^{m}(\frac{z_i - \mu_i}{\sigma_i})^2 \\ \text{s.t. } z_i = h_i(x) + r_i\ ,\ i = 1, ..., m \end{gathered} \tag{4.5}$$

so that, r_i is call residual of measurements. And if we rewrite the above equation in matrix format we will have:

$$\begin{gathered} min\ J(x) \\ J(x) = [z - h(x)]^T W [z - h(x)] \end{gathered} \tag{4.6}$$

In order to solve this optimization problem, the solution must pursue first order optimality condition, which:

$$\begin{aligned} &\frac{\partial J}{\partial x} = 0 \quad \rightarrow \quad [-\frac{\partial h}{\partial x}]^T W[z - h(x)] = 0 \\ &H(x) = \frac{\partial h}{\partial x} \quad \text{is Jacobian matrix of } h(x) \end{aligned} \tag{4.7}$$

From now on, we will have two types of view to the problem. One is DC SE, in which the $h(x)$ is a linear function (that means measurements have linear relation with state variables) and we can do the matrix calculation directly. The other one is AC SE, in which $h(x)$ is nonlinear and one must use methods such as Gauss-Newton to solve the problem. Initially, AC SE will be discussed and afterward, DC SE which is a simpler version of AC SE will be presented.

4.1.2 AC state estimation

Considering $g(x) = \frac{\partial J}{\partial x}$ and expanding $g(x)$ around the state vector x^k into its Taylor series while neglecting higher order terms, results in an iterative solution to the problem known as Gauss-Newton method:

$$g(x) = g(x^k) + G(x^k)(x - x^k) + ... = 0 \tag{4.8}$$

$$x^{k+1} = x^k - [G(x^k)]^{-1} g(x^k) \tag{4.9}$$

$$G(x^k) = \frac{\partial g(x^k)}{\partial x} = H^T(x^k) W H(x^k) \tag{4.10}$$

$$g(x^k) = -H^T(x^k) W(z - h(x^k) \tag{4.11}$$

where k is the iteration index, x^k is the solution vector at iteration k, and $G(x)$ is called the gain matrix.

The Weighted Least Squares (WLS) method is a widely used method for power system state estimation. It is a statistical method that minimizes the weighted sum of the squared residuals of the measurement equations. The residuals are the differences between the measured values and the estimated values. The weights are

used to give more importance to certain measurements than others.

Here are the steps to perform the WLS method for power system state estimation:

- Initialize the state vector to an initial guess (flat start).
- Calculate the gain matrix, Jacobian matrix, and measurement vector.
- Calculate the weighted sum of the squared residuals.
- Update the state vector.
- Repeat steps until the state vector converges.

Measurements in power system state estimation can be of different types, but the most common are line power flows, bus power injections, and bus voltage magnitudes. These measurements can be expressed in terms of the state variables using either rectangular or polar coordinates.

When using polar coordinates for a system with N buses, the state vector will have $2N - 1$ elements; N bus voltage magnitudes, and $N - 1$ phase angles, where the phase angle of one reference bus is set to an arbitrary value, such as 0. If bus 1 is chosen as the reference, the state vector x will have the following form:

$$x^T = [\theta_2 \ \theta_3 \ \ldots \ \theta_N \ V_1 \ V_2 \ \ldots \ V_N] \tag{4.12}$$

Assuming the general two-port π model for network branches, real and reactive power injection (P_i and Q_i) at bus i are:

$$P_i = V_i \sum_{j=1}^{N} V_j (G_{ij} \cos(\theta_{ij}) + B_{ij} \sin(\theta_{ij})) \tag{4.13}$$

$$Q_i = V_i \sum_{j=1}^{N} V_j (G_{ij} \sin(\theta_{ij}) - B_{ij} \cos(\theta_{ij})) \tag{4.14}$$

Real and reactive power flow from bus i to bus j:

$$P_{ij} = V_i^2 (g_{si} + g_{ij}) - V_i V_j (g_{ij} \cos(\theta_{ij}) + b_{ij} \sin(\theta_{ij})) \tag{4.15}$$

$$Q_{ij} = -V_i^2(b_{si} + b_{ij}) - V_i V_j(g_{ij}\sin(\theta_{ij}) - b_{ij}\cos(\theta_{ij})) \tag{4.16}$$

where V_i and θ_i is the voltage magnitude and phase angle at bus i; $\theta_{ij} = \theta_i - \theta_j$; $G_{ij} + jB_{ij}$ is the ijth element of the complex bus admittance matrix; $g_{ij} + jb_{ij}$ is the admittance of the series branch connecting buses i and j; $g_{si} + jb_{si}$ is the admittance of the shunt branch connected at bus i;

The structure of the Jacobian matrix, H, will be as follows:

$$H = \begin{bmatrix} \frac{\partial P_i}{\partial \theta} & \frac{\partial P_i}{\partial V} \\ \frac{\partial P_{ij}}{\partial \theta} & \frac{\partial P_{ij}}{\partial V} \\ \frac{\partial Q_i}{\partial \theta} & \frac{\partial Q_i}{\partial V} \\ \frac{\partial Q_{ij}}{\partial \theta} & \frac{\partial Q_{ij}}{\partial V} \\ \frac{\partial V_i}{\partial \theta} & \frac{\partial V_i}{\partial V} \end{bmatrix} \tag{4.17}$$

The expressions for real and reactive power injections can be obtained via the following equations:

$$\frac{\partial P_i}{\partial \theta_i} = \sum_{j=1}^{N} V_i V_j(-G_{ij}\sin(\theta_{ij}) + B_{ij}\cos(\theta_{ij})) - V_i^2 B_{ii} \tag{4.18}$$

$$\frac{\partial P_i}{\partial \theta_j} = V_i V_j(G_{ij}\sin(\theta_{ij}) - B_{ij}\cos(\theta_{ij}) \tag{4.19}$$

$$\frac{\partial P_i}{\partial V_i} = \sum_{j=1}^{N} V_j(G_{ij}\cos(\theta_{ij}) + B_{ij}\sin(\theta_{ij})) + V_i G_{ii} \tag{4.20}$$

$$\frac{\partial P_i}{\partial V_j} = V_i(G_{ij}\cos(\theta_{ij}) + B_{ij}\sin(\theta_{ij}) \tag{4.21}$$

$$\frac{\partial Q_i}{\partial \theta_i} = \sum_{j=1}^{N} V_i V_j(G_{ij}\cos(\theta_{ij}) + B_{ij}\sin(\theta_{ij})) - V_i^2 G_{ii} \tag{4.22}$$

$$\frac{\partial Q_i}{\partial \theta_j} = V_i V_j(-G_{ij}\cos(\theta_{ij}) - B_{ij}\sin(\theta_{ij}) \tag{4.23}$$

$$\frac{\partial Q_i}{\partial V_i} = \sum_{j=1}^{N} V_j(G_{ij}\sin(\theta_{ij}) - B_{ij}\cos(\theta_{ij})) - V_i B_{ii} \tag{4.24}$$

$$\frac{\partial Q_i}{\partial V_j} = V_i(G_{ij}\sin(\theta_{ij}) - B_{ij}\cos(\theta_{ij}) \tag{4.25}$$

And, the elements for power flow can be calculated utilizing the following equations:

$$\frac{\partial P_{ij}}{\partial \theta_i} = V_i V_j(g_{ij}\sin(\theta_{ij}) - b_{ij}\cos(\theta_{ij})) \tag{4.26}$$

$$\frac{\partial P_{ij}}{\partial \theta_j} = -V_i V_j(g_{ij}\sin(\theta_{ij}) - b_{ij}\cos(\theta_{ij}) \tag{4.27}$$

$$\frac{\partial P_{ij}}{\partial V_i} = -V_j(g_{ij}\cos(\theta_{ij}) + b_{ij}\sin(\theta_{ij})) + V_i(g_{ij} + g_{si}) \tag{4.28}$$

$$\frac{\partial P_{ij}}{\partial V_j} = -V_i(g_{ij}\cos(\theta_{ij}) + b_{ij}\sin(\theta_{ij}) \tag{4.29}$$

$$\frac{\partial Q_{ij}}{\partial \theta_i} = -V_i V_j(g_{ij}\cos(\theta_{ij}) + b_{ij}\sin(\theta_{ij})) \tag{4.30}$$

$$\frac{\partial Q_{ij}}{\partial \theta_j} = V_i V_j(g_{ij}\cos(\theta_{ij}) + b_{ij}\sin(\theta_{ij}) \tag{4.31}$$

$$\frac{\partial Q_{ij}}{\partial V_i} = -V_j(g_{ij}\sin(\theta_{ij}) - b_{ij}\cos(\theta_{ij})) - 2V_i(b_{ij} + b_{si}) \tag{4.32}$$

$$\frac{\partial Q_{ij}}{\partial V_j} = -V_i(g_{ij}\sin(\theta_{ij}) - b_{ij}\cos(\theta_{ij}) \tag{4.33}$$

Finally, the elements related to voltage magnitude measurement can be stated as follows:

$$\frac{\partial V_i}{\partial V_i} = 1, \frac{\partial V_i}{\partial V_j} = 0, \frac{\partial V_i}{\partial \theta_i} = 0, \frac{\partial V_i}{\partial \theta_j} = 0 \tag{4.34}$$

4.1.3 DC state estimation

DC state estimation is a simplified method of state estimation that assumes that the power system is operating in a steady state and that all AC quantities can be approximated by their DC equivalents. This allows for the use of linear equations to model the power system, which makes the state estimation problem much easier to solve.

The DC approximation of real power flow is obtained by assuming that the voltage magnitudes at all buses are equal to 1.0 per unit and that all shunt elements and branch resistances are negligible. Under these assumptions, the real power flow from bus k to m can be approximated by the first-order Taylor expansion around $\theta = 0$, as shown in the following equation:

$$P_{ij} = \frac{\theta_i - \theta_j}{x_{ij}} + e \tag{4.35}$$

where x_{ij} is the reactance of the branch $i - j$; θ_i is the phase angle of the ith bus voltage; and e is the measurement error. Similarly, the power injection measurement at bus i can be stated as the sum of power flows end to (or initiated from) bus i.

Based on least square method, which is explained in the previous section, and considering DC load flow (where $h(x)$ is a linear function of state variables ($h(x) = Hx$), all voltage magnitudes are one per-unit and only vector of voltage phase angles, θ, is the variable. For DC state estimation the relation between measurement and the state variables is linear which means the H matrix is not related to the state variables, but only to the network parameters. Solving (4.7) taking into account the linearized H matrix will give out θ (the main equation would have a form like $A\theta = y$ and finally $\theta = A^{-1}y$):

$$(H^T W H)\theta = H^T W z \tag{4.36}$$
$$\theta^T = [\theta_2\ \theta_3\ \ldots\ \theta_N]$$

4.2 Distributed state estimation

Distributed state estimation is a process of estimating the state of a system using measurements from multiple sensors or agents that are distributed throughout the system. Distributed SE is a powerful tool for estimating the state of complex systems, such as power systems, sensor networks, and multi-robot systems.

Distributed SE offers several advantages over traditional centralized state estimation methods. First, distributed SE is more scalable to large systems. Second, distributed SE is more robust to sensor failures and communication outages. Third, distributed SE can be used to estimate the state of systems in real time. Fig. 4-2 demonstrates a general scheme of distributed SE.

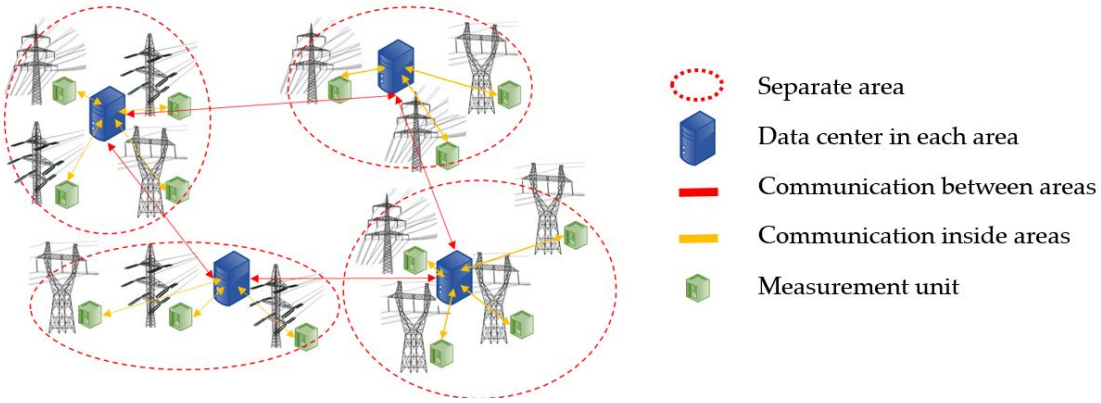

Figure 4-2: A general scheme of distributed state estimation

It is to be noted that to have a solution for (4.36), $(H^T W H)^{-1}$ should be invertible. In other words, if the set of measurements are sufficient and well distributed, the network would be observable, which requires matrix H to be full rank (that has been considered in our simulation) and that would lead to non-singularity of the $(H^T W H)^{-1}$ matrix. In order to calculate (4.36) we need to access all data in the system by a single (or centralized) control unit. As mentioned before, issues like communication bottleneck, data privacy and cybersecurity, are the main reasons which leads power system to utilize decentralized approaches.

4.2.1 Matrix splitting

In order to obtain SE problem's solution in a distributed manner, one can use matrix splitting method and after doing a certain number of iterations the answer converges to the centralized solution [3]. The main equation of matrix splitting for a problem of $Ax = y$ is:

$$x^{t+1} = M^{-1}Nx^t + M^{-1}y \tag{4.37}$$

that A is written as the sum of an invertible (or diagonal) matrix M, and a matrix N; so that $M = D + E'_{ii}$ and $N = E - E'_{ii}$. Note that, D contains diagonal arrays and E contains off-diagonal arrays of matrix A. And E'_{ii} is a diagonal matrix which is defined as follows:

$$E'_{ii} = \alpha \sum_{j=1}^{n} |E_{ij}| \tag{4.38}$$

that we have assumed $\alpha = 1$ for simplicity. It is to be noted that, (4.37) converges if the spectral radius of $M^{-1}N$ matrix be less than 1 ($\rho(M^{-1}N) < 1$). Using (4.37) iteratively, leads to convergence to the system $Ax = y$ final solution, i.e. x^*.

4.2.2 Gossip based

Here another approach to solve (4.36) in a distributed manner is presented, which is discussed in [84] that the authors have considered measurement units in an asynchronous manner (i.e. gossip communication protocol).

Considering the DC approximation, the SE problem in the least squares setting can be formulated by (4.36). Based on what has been stated in (4.36), this problem has a closed-form solution. Let's assume, $L = H^T W H$ and $u = H^T W z$. One way to compute this solution x^* is through the gradient based iterative algorithm given by:

$$x(k+1) = (I - \tau L)x(k) + \tau u \tag{4.39}$$

And the parameter τ is selected from the interval $(0\ ,\ 2||L||^{-1})$; such a τ guarantees the matrix $I - \tau L$ to be Schur stable (i.e. the iterative method converges). At each iteration, a set of two neighboring areas, randomly (based on uniform probability distribution) will be selected to update the common variables.

4.2.3 Decomposition method

In this part the method provided in [9] is discussed. This method applies explicitly power flow and power injection equations to solve multi-area DC SE problem:

$$min\ H_k(x_k) + \sum_{l \in \Omega_k} H_{kl}(x_k, \tilde{x}_l)$$

$$H_k(x_k) = \sum_{i \in \Omega_k^P} \omega_{k,i}^P (P_{k,i}^m - P_{k,i})^2 + \sum_{(i,j) \in \Omega_k^{PF}} \omega_{k,ij}^{PF} (P_{k,ij}^m - P_{k,ij})^2$$

$$H_{kl}(x_k, \tilde{x}_l) = \sum_{i \in \Omega_{kl}^P} \omega_{kl,i}^P (P_{kl,i}^m - P_{kl,i})^2 + \sum_{(i,j) \in \Omega_{kl}^{PF}} \omega_{kl,ij}^{PF} (P_{kl,ij}^m - P_{kl,ij})^2 \tag{4.40}$$

$$+ \sum_{i \in \Omega_{kl}} \omega_{l,i}^x (\tilde{x}_{l,i} - x_{l,i})^2$$

where H_{kl} is weighted measurement error function for area k involving state variables of area k and l, H_k is weighted measurement error function for area k involving only state variables of area k, and Ω_k is the set containing indices for all neighboring areas of area k, ω is weighting factor, $P_{(.),i}^{(m)}$ is active power injection measurement at bus i, $P_{(.),ij}^{(m)}$ is active power flow measurement in between bus i and j; $P_{(.),i}$ and $P_{(.),ij}$ are the physical equation related to power injection and power flow, respectively.

In order to solve (4.40), MATLAB solver (Sequential quadratic programming (SQP)) via MATLAB *R2018b* has been applied.

4.2.4 ADMM

In [85] a new method has been developed for solving distributed SE, which is based on ADMM [86]. As claimed by the authors, ADMM increases existing SE solvers performance and convergence of the method to its centralized counterpart is guaranteed even if we don't have local observability. ADMM can also be considered in the same category as decomposition methods, but due to multiple applications of this method recently, we have decided to consider it separately.

In general, the distributed SE problem can be formulated as:

$$\begin{aligned} &\min_{x_k} \sum_{k=1}^{K} H_k(x_k) \\ &\mathrm{x}_k[l] = x_l[k], \forall l \in N_k, \forall k \in K \end{aligned} \tag{4.41}$$

where N_k is the set of areas sharing states with area k and $x_{k,l}$ is auxiliary variable introduced per pair of interacting areas k, l.

The constraint forces neighboring areas to consent on their shared variables. Augmented Lagrangian function is as follows:

$$\begin{aligned} &L(\{x_k\}, \{x_{kl}\}; \{v_{kl}\}) \\ &:= \sum_{k=1}^{K} [H_k(x_k) + \sum_{l \in N_k} (v_{k,l}^T (x_{k[l]} - x_{kl}) + \frac{c}{2} ||x_{k[l]} - x_{kl}||_2^2)] \end{aligned} \tag{4.42}$$

where $v_{k,l}$ is Lagrangian multiplier and $c > 0$.

$$\begin{aligned} \{\mathrm{x}_k^{t+1}\} &:= \ arg\ min\ L(\{x_k\}, \{x_{kl}^t\}; \{v_{kl}^t\}) \\ \{\mathrm{x}_{kl}^{t+1}\} &:= \ arg\ min\ L(\{x_k^{t+1}\}, \{x_{kl}\}; \{v_{kl}^t\}) \\ v_{k,l}^{t+1} &:= v_{k,l}^t + c(x_{k[l]}^{t+1} - x_{kl}^{t+1}),\ \forall k \end{aligned} \tag{4.43}$$

4.3 Forecasting aided state estimation

Taking into account the slow enough changes in the system operating point (exclusively due to slow and smooth load/renewable generation changes) [81], state transition model can be described by linear stochastic equation [26, 87]:

$$\boldsymbol{x}_t = \boldsymbol{A}_{t-1}\boldsymbol{x}_{t-1} + \boldsymbol{g}_{t-1} + \boldsymbol{\omega}_{t-1} \tag{4.44}$$

where $\boldsymbol{x}$ is the state vector composed of $n = 2N - 1$ elements (bus voltages and phase angles at all buses except phase angle at the slack bus), t is sampling time, $\boldsymbol{A}$ is state transition matrix, $\boldsymbol{g}$ is trend vector, $\boldsymbol{\omega}$ is process noise assumed to have Gaussian distribution with zero mean and covariance matrix $\boldsymbol{Q}$, and N is total number of buses. The widely used approach for updating matrix $\boldsymbol{A}$ and vector $\boldsymbol{g}$ is

Holt's exponential smoothing regression [37].

Set of m measurements considered in this paper is composed of active and reactive power flows, active and reactive power injections, and voltage magnitudes. Relation between these measurements and the states at time t can be expressed as follows:

$$\boldsymbol{z}_t = \boldsymbol{h}(\boldsymbol{x}_t) + \boldsymbol{e}_t \tag{4.45}$$

where $\boldsymbol{z}$ is measurement vector, $\boldsymbol{h}$ is set of nonlinear equations, and $\boldsymbol{e}$ represents vector of measurement noise assumed to be Gaussian distributed with zero mean and covariance matrix $\boldsymbol{R}$.

4.3.1 Extended Kalman filter based Forecasting Aided State Estimation

FASE is a special application of dynamic state estimation concept in which the dynamics of the states are negligible. FASE utilizes both state transition and measurement model represented in (4.44) and (4.45), respectively. Kalman filtering has commonly been used as the optimal solution for many data tracking and forecasting tasks [88]. The Kalman filter is an estimator that uses previous state and current snapshot of measurements to estimate the current state.

Considering the power system, the measurement function is nonlinear in nature. If model is nonlinear, an extension of classical Kalman filter, so called EKF, can be utilized through the linearization of the nonlinear model via Taylor series [89]. There are also other extensions of Kalman filter built to deal with nonlinearity of the system, such as UKF [26], Particle Filter [90], iterated EKF [91], Ensamble Kalman filter [92], Second order Kalman filter [93], Cubature Kalman filter [94], to name a few. However, for the sake of simplicity and less computation burden, a first order EKF based estimator has been utilized in this study.

4.3.2 Prediction equations

Consider $\hat{\boldsymbol{x}}_{t-1}$ and $\hat{\boldsymbol{P}}_{t-1}$ are estimated state vector and estimated states' error covariance matrix at time step $t-1$, respectively. Assuming that $\hat{\boldsymbol{x}}_{t-1}$ and $\hat{\boldsymbol{P}}_{t-1}$ are

known, the system state vector can be predicted using following equations [87]:

$$\tilde{\boldsymbol{x}}_t = \boldsymbol{A}_{t-1}\hat{\boldsymbol{x}}_{t-1} + \boldsymbol{g}_{t-1} \tag{4.46}$$

$$\tilde{\boldsymbol{P}}_t = \boldsymbol{A}_{t-1}\hat{\boldsymbol{P}}_{t-1}\boldsymbol{A}_{t-1}^T + \boldsymbol{Q}_{t-1} \tag{4.47}$$

where $\tilde{\boldsymbol{x}}_t$ is predicted state vector and $\tilde{\boldsymbol{P}}_t$ indicates the state prediction error covariance matrix at time step t.

4.3.3 Filtering equations

After predicting the system state using (4.46) and (4.47), and receiving a new set of measurements $\boldsymbol{z}_t$ at time instant t, estimated state $\hat{\boldsymbol{x}}_t$ and its covariance matrix $\hat{\boldsymbol{P}}_t$ can be obtained as [87]:

$$\boldsymbol{\nu}_t = \boldsymbol{z}_t - \boldsymbol{h}(\tilde{\boldsymbol{x}}_t) \tag{4.48}$$

$$\boldsymbol{M}_t = \boldsymbol{H}_t\tilde{\boldsymbol{P}}_t\boldsymbol{H}_t^T + \boldsymbol{R}_t \tag{4.49}$$

$$\boldsymbol{K}_t = \tilde{\boldsymbol{P}}_t\boldsymbol{H}_t^T\boldsymbol{M}_t^{-1} \tag{4.50}$$

$$\hat{\boldsymbol{x}}_t = \tilde{\boldsymbol{x}}_t + \boldsymbol{K}_t\boldsymbol{\nu}_t \tag{4.51}$$

$$\hat{\boldsymbol{P}}_t = \tilde{\boldsymbol{P}}_t - \boldsymbol{K}_t\boldsymbol{M}_t\boldsymbol{K}_t^T \tag{4.52}$$

where $\boldsymbol{H}$ represents Jacobian of $\boldsymbol{h}$ evaluated at $\tilde{\boldsymbol{x}}$; $\boldsymbol{\nu}$, $\boldsymbol{M}$ and $\boldsymbol{K}$ indicate innovation vector, innovation covariance matrix and Kalman gain, respectively.

4.4 Summary

This chapter provided mathematical formulation of three different state estimation (SE) techniques: centralized, distributed, and forecasting-aided. Centralized SE was the most common technique. Distributed SE addresses the challenges of centralized SE by distributing the computational load across multiple computers. Forecasting-aided SE improved the accuracy of SE by predicting future system states. This chapter discussed the advantages and disadvantages of each technique, as well as the different algorithms that could be used to implement each one.

Chapter 5

Implementation and simulation

In this chapter, we delve into the outcomes of extensive research and simulations, shedding light on the significance of our contributions. Our investigations encompass two key areas: the detection and identification of anomalies in power system SE and the pursuit of optimal area partitioning strategies within the context of distributed SE, considering a modified convergence criterion.

In this chapter, we present the results of our research, which includes the development and implementation of a novel algorithm for anomaly detection and classification. Leveraging the power of machine learning, this algorithm offers the potential to improve the way we detect and identify critical events in the power system, ultimately enhancing its reliability and resilience.

In parallel, we explore the distributed SE, where the challenges of data communication, computation, and convergence are prevalent. Our investigations center on the concept of optimal area partitioning, a strategy that has the potential to significantly reduce communication overhead and expedite the convergence of distributed SE methods. We consider the implications of applying a modified convergence criterion to evaluate the performance of these methods as well.

5.1 Detection, classification, and identification

State estimation (SE) is the core component of the energy management system (EMS) for power grids. SE provides the most likely values of the voltage magni-

tudes and phase angles for all buses in the power system. The accuracy of these values is essential for achieving optimal and secure operation of the system [7]. Fig. 5-1 demonstrates the connection between physical, communication and energy management systems, emphasizing the role of anomaly detection, classification and identification units within the state estimator.

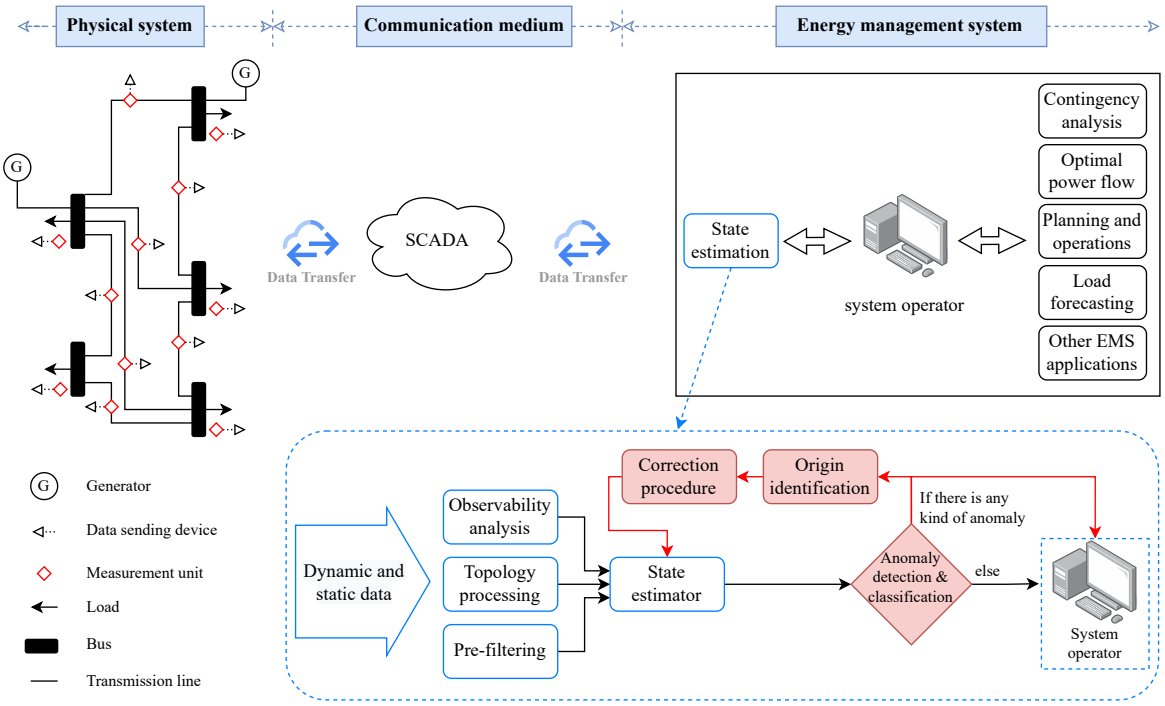

Figure 5-1: State estimator's role in the power system and position of anomaly detection, classification and identification unit within the State estimator

This section describes the workflow of the proposed method to detect, classify, and identify anomalies. As it is shown in Fig. 5-2, the workflow combines analytical and ML approaches which are explained in the following subsections in detail. Input data, i.e. observed SCADA measurements and statuses of switching devices, is being delivered to the EMS and after that SE is performed using WLS and EKF. Measurement residuals obtained at WLS output are used to carry out χ^2-test in order to check for BD. If some measurements are corrupted with BD, χ^2-test will rise a flag indicating BD presence. If SLC or FDIA occurs, χ^2-test will not detect anomaly presence and the process will continue to FASE-WLS based stage. In the FASE-WLS based stage, the estimated states that are obtained by WLS and EKF are utilized to form an anomaly detection index (5.1). In case the index value is

equal or higher than the specific threshold, anomaly is detected; otherwise, system is considered to be in the normal operation mode. Although the index is capable to detect presence of those anomalies for which χ^2-test stays blind, it is still not capable of classifying the anomaly type due to similar impacts which SLC and FDIA have on WLS and EKF estimates. To classify detected anomaly as SLC or FDIA, workflow leverages the ML based stage. After classification of the anomaly, the next step in the ML based stage is to identify at which bus(es) SLC has happened, or which state(s) has been affected by FDIA.

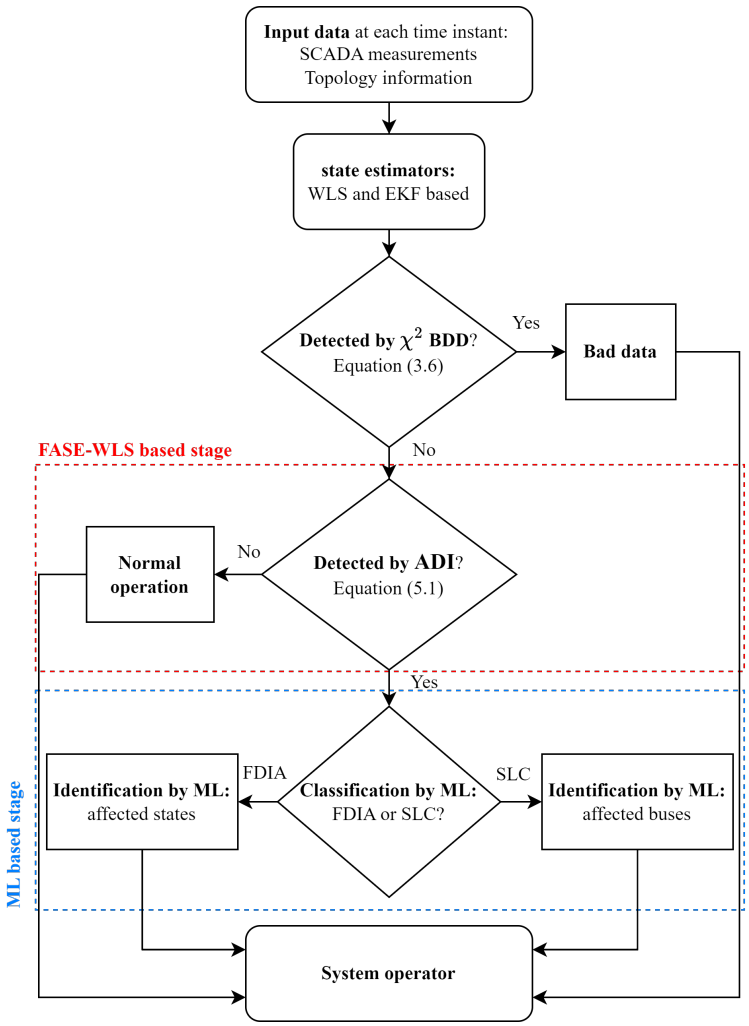

Figure 5-2: Flowchart demonstration of the proposed algorithm for anomaly detection and classification

5.1.1 FASE-WLS based approach

In this research work, in order to detect SLC or FDIA presence, an anomaly detection index (ADI) that combines WLS and EKF estimates is employed [5]:

$$ADI_i = \frac{|\hat{x}_i^{WLS} - \hat{x}_i^{EKF}|}{\sqrt{\hat{P}_{ii}}} \quad i = 1, 2, ..., n \tag{5.1}$$

where $\hat{x}_i^{WLS}$ is i-th state variable estimated by WLS; $\hat{x}_i^{EKF}$ is i-th state variable estimated by EKF; $\hat{P}_{ii}$ is the i-th diagonal element of EKF estimated states' error covariance matrix, $\hat{\boldsymbol{P}}$. If $\max_i\{ADI_i\} \geq \gamma$, SLC or FDIA presence is detected. Here, γ represents detection threshold that has to be selected to clearly discriminate between normal operation and anomalies like SLC and FDIA. BD is less relevant for the threshold setting because, when it occurs, it is expected to be detected by χ^2-test. In this thesis, EKF has been selected as a Kalman filter extension for nonlinear systems. However, the anomaly detection index (5.1) can be equally utilized for anomaly detection if any other type of Kalman filter has been used.

5.1.2 Supervised machine learning algorithms

As previously mentioned SLC and FDIA would have similar impact on WLS and EKF estimates. This fact makes it impossible to classify whether the anomaly is SLC or FDIA using ADI or similar analytical methods. That makes ML algorithms an appropriate choice for SLC and FDIA classification. Because ML algorithms have low dependency on the system model. To classify SLC or FDIA, as well as to determine the buses (or states) that have been affected by SLC (or FDIA), the following supervised ML algorithms were employed and compared:

- Logistic Regression (LR),
- K-Near Neighborhood (KNN),
- Random Forest (RF),
- Extreme Gradient Boosting (XGB).

Classification is a ML supervised concept which categorizes a set of data into classes. Classification algorithms determine the class of new samples based on past samples during training. Depending on the number of classes, the problem can be considered as a binary or multi-class classification. In our case, the classification of FDIA or SLC is a binary classification task, while determination of the state or bus that has been affected by anomaly, represents a multi-class classification problem.

Proposed supervised algorithms work with labeled data. The number of labeled samples used in this thesis depends on the task which we consider. Each collected sample is represented by $\boldsymbol{x}$ and y coordinates $\{(\boldsymbol{x}_i, y_i), i = 1, ...m_{total}\}$, where $\boldsymbol{x}$ and y describes input and output of the models, respectively; m_{total} is total number of samples. Each input is a n_x-dimensional feature vector $\boldsymbol{x}_i \in R^{n_x}$. However, each output is a class labeled that represent FDIA or SLC $[y_i \in \{0, 1\}]$, or states/buses of considered IEEE 14-bus system.

To train and validate models, data set samples were splitted in training $\{(\boldsymbol{x}_i, y_i), i = 1, ...m_{train}\}$ and testing $\{(\boldsymbol{x}_i, y_i), i = 1, ...m_{test}\}$ subsets; m_{train} and m_{test} are training and testing number of samples, respectively. Splitting was conducted in a stratified fashion, containing approximately the same percentage of samples of each labeled class. 80% of data samples were used for training and 20% data samples were tested. The controllable hyperparameters are tuned using sequential optimization with gradient boosting [95] as a surrogate probability model of the objective function.

LR

It is a statistical *linear* classifier using a logistic function to frame a binary output model [96]. The cost function $J(\boldsymbol{\theta})$ that is a function of model parameters $\boldsymbol{\theta}$ is given by:

$$J(\boldsymbol{\theta}) = \frac{1}{m}\sum_{i=1}^{m} l(h_{LR}(\boldsymbol{x}_i), y_i) + \lambda \sum_{j=1}^{n_x} \boldsymbol{\theta}_j^2 \tag{5.2}$$

$$l(h_{LR}(\boldsymbol{x}_i), y_i) = \begin{cases} -log(h_{LR}(\boldsymbol{x}_i)) & y_i = 1 \\ -log(1 - h_{LR}(\boldsymbol{x}_i)) & y_i \neq 0 \end{cases} \tag{5.3}$$

where λ is a regularization strength hyperparameter and h_{LR} represents a hypothesis function. $\lambda \sum_{j=1}^{n_x} \boldsymbol{\theta}_j^2$ represents regularization term in order to avoid overfitting and penalise large values of the model parameters.

One-vs-rest method is used to train h_{LR}^c in the multi-class case. For an unseen input x_t, $h_{LR}^c(x_t)$ determines the probability $p(\cdot)$ for each class label c using the sigmoid function as follows:

$$h_{LR}^c(\boldsymbol{x}_t) = p(y = c|\boldsymbol{x}_t, \boldsymbol{\theta}) = \frac{1}{1 + e^{\boldsymbol{\theta}^T \boldsymbol{x}_t}} \tag{5.4}$$

The final predicted outcome is find as:

$$\max_c \; h_{LR}^c(\boldsymbol{x}_t) \tag{5.5}$$

KNN

It is an instance-based learning algorithm in which the labeled training data of different classes are simply stored and used to make new predictions [97]. Learning is based on the k-nearest neighbors of each unseen samples, assuming that the most of nearby samples belong to a predetermined class. Accordingly, the unseen sample will also be attributed to this class. In order to find the k-nearest neighbors, the standard Euclidean metric was proposed as a measure of the distance. It is defined as the 2-norm of the vector between $\boldsymbol{x}_t$ and training points in an n_x-dimensional space, which can be written as follows:

$$d_i(\boldsymbol{x}_t, \boldsymbol{x}_i) = \left\| \boldsymbol{x}_t^T - \boldsymbol{x}_i^T \right\|_2 = \sqrt{\sum_{j=1}^{n_x} (\boldsymbol{x}_{t,j} - \boldsymbol{x}_{i,j})^2} \tag{5.6}$$

$$d_i(\boldsymbol{x}_t, \boldsymbol{x}_i) \geq 0, \quad i = 1, ...m_{train}$$

RF

It is an non-linear ensemble classifier based on multiple decision trees, using a randomly selected subset of training samples and variables [98]. RF is designed using the bootstrap aggregating, also known as the bagging technique based on the combi-

nation of weak decision tree algorithms in parallel that will improve the effectiveness of the prediction [99].

The RF architecture represents a collection of t_{RF} randomized classification trees with tree-structured classifier h_{RF}. The predicted value for the given $\boldsymbol{x}_t$ and j-th tree in the forest is expressed:

$$h_{RF}(\boldsymbol{x}_t, \boldsymbol{\eta}_j), \ j = 1, ..., t_{RF} \tag{5.7}$$

where $\boldsymbol{\eta}_j$ are the independent identically distributed random vectors.

h_{RF} trees are created in parallel, independent of one another, using the bootstrapped data sets for growing the trees. Random subset of variables at each step of tree growth is used to split the node. The quality of the splitting is measured by applying the gini impurity criterion G:

$$\mathcal{G} = \sum_{i=1}^{c} p(i) \cdot (1 - p(i)) \tag{5.8}$$

where $p(i)$ is the probability of choosing the data point with class i at a given node.

The final RF prediction $\hat{y}$ is the majority vote over collection of trees:

$$\hat{y} = \ majority \ vote \ \{h_{RF}^{j}(\cdot)\}_{j=1}^{t_{RF}} \tag{5.9}$$

XGB

It is designed using boosting technique and attempts to build a robust model from the number of weak tree classifiers in series [100]. XGB is a specific implementation of the Gradient Boosting method [101], which uses more accurate approximations by employing second-order gradients and advanced regularization.

For a given number of t_{XGB} sequentially connected decision tree models h_{XGB}, the final XGB prediction of $\boldsymbol{x}_t$ is computed as:

$$\hat{y} = \sum_{j=1}^{t_{XGB}} h_{XGB}^{j}(\boldsymbol{x}_t) \tag{5.10}$$

where $h_{XGB}^{j}(x_t) = w_q^j(x_t)$. w_q^j is the score of the corresponding leaf q in the j-th

tree.

To optimize the functions used in the model, XGB minimizes the following regularized loss function $\mathcal{L}$:

$$\mathcal{L}(y, \hat{y}) = \sum_{i=1}^{m_{train}} l(y_i, \hat{y}_i) + \sum_{j=1}^{t_{XGB}} \Omega(h^j_{XGB}) \tag{5.11}$$

where $l(\cdot)$ is a logistic loss function given with (5.3). The second term Ω of the loss function is the regularization term which penalizes the complexity of the model:

$$\Omega(h_{XGB}) = \gamma T + \frac{1}{2}\lambda \sum_{q=1}^{T} w_q^2 \tag{5.12}$$

where γ is the pseudo-regularization hyperparameter and T is the number of leaves in the tree. It is to be noted that the hyperparameters are tuned using sequential optimization.

Since XGB ensemble model includes functions as parameters and cannot be optimized using conventional optimization methods, the training of the model is executed in manner of t steps:

$$\mathcal{L}(t) = \sum_{i=1}^{m_{train}} l(y_i, \hat{y}_i^{t-1} + h^t_{XGB}(x_i)) + \Omega(h^t_{XGB}) \tag{5.13}$$

where h^t_{XGB} is greedily added to improve model.

5.1.3 Evaluation metric

To evaluate the accuracy of the FDIA or SLC classification, and identification of states or buses that have been affected by these anomalies, the (*macro*) F1-score metric is used [102,103]. F1-score is a *harmonic* mean of the *precision* Pr and *recall* Re. For a particular predicted output vector $\hat{y}$ and ground-truth y, the percentage of F1-score is computed as follows:

$$F1 = 2 \cdot \frac{Pr \cdot Re}{Pr + Re} \times 100 \tag{5.14}$$

The *precision* represents the number of *True Positives (TP)* over *TP* plus the number of *False Positives (FP)*:

$$Pr = \frac{TP}{TP + FP} \tag{5.15}$$

The *recall* is the *TP* over the *TP* plus the number of *False Negatives (FN)*:

$$Re = \frac{TP}{TP + FN} \tag{5.16}$$

As an example for a binary classification task, *TP* is the number of correctly identified FDIAs. *FP* is the number of SLCs identified as FDIAs, and *FN* is the number of FDIAs identified as SLCs. The similar logic is applicable for SLC, multi-bus SLC and multi-state FDIA.

The macro F1-score is calculated as the arithmetic mean over the F1-scores of each class:

$$macro\ F1 = \frac{1}{c}\sum_{i=1}^{c} F1_i. \tag{5.17}$$

5.1.4 Maximum Relevance – Minimum Redundancy

Increase in the size of the system will lead to increase in the number of features which will consequently increase optimization complexity of the ML algorithm. Maximum Relevance Minimum Redundancy (MRMR) is a feature selection algorithm for finding the minimal-optimal subset of features [104, 105]. Minimal-optimal methods select a small set of features that have the maximum possible predictive power by eliminating irrelevant features. Accordingly, the model optimization complexity is reduced.

The purpose of this method is to reduce the number of input features that linearly increases with the system size:

$$n_x = \alpha N - \beta \tag{5.18}$$

where $\alpha = 16$ that is sum of features for each bus and $\beta = 10$ is sum of features related to slack bus. Accordingly, by employing MRMR we are able to select just a few main features to achieve the high enough level of accuracy. However, the number

of optimal features is specific to each test system and can not be generalized.

MRMR works iteratively. At each iteration i, it identifies the best feature f_b that has maximum relevance with respect to the target variable and minimum redundancy with respect to the features that have been selected at previous iterations. Since nonlinear dependency exists between input and target variables, to compute maximum relevance of the feature f_b, we use mutual information method [106]. In contrast, for computing minimum redundancy of the feature f_b, we employ Spearman's rank based correlation [107] because of handling non-normality data.

The score for each feature f_b at each iteration i is computed as follows [105]:

$$score_i(f_b) = \frac{relevance(f_b|target)}{redundancy(f_b|f_{b_{selected\ until\ i-1}})} \tag{5.19}$$

where the best feature f_b at iteration i is the one having the highest score.

In this thesis, MRMR has been applied to select a few main features that will result in reducing the optimization complexity during training of the ML algorithms. As a parameter of optimization complexity, training time has been considered. The reduced training times are also presented in the next section for the utilized supervised ML algorithms.

5.1.5 Solution for topology changes

As mentioned before, the transmission system faces topology changes (i.e. change in the network configuration). This will require ML algorithm to be retrained. Retraining ML algorithm can be time consuming and inefficient depending on the size of the system. To eliminate the need for retraining the ML algorithm after the change in network topology, features related to the branches are excluded and only features associated with the buses are applied for ML algorithm training. These features are the ones associated only with the buses such as: a) nodal measurements and normalized measurement innovations of voltage magnitude and active/reactive power injection; b) estimates and predictions of voltage magnitude, phase angle and active/reactive power injection. This solution is only applicable for networks for which the number of buses does not change with topology changes. Therefore, it

can be applied on meshed networks but not on radial networks. Five topologies are used to train and afterwards examine the accuracy of anomaly classification and identification of its origin. These five topologies include original topology of the IEEE 14 bus system. Fig. 5-3 presents the original topology of the IEEE 14 bus system [1]. Four new topologies obtained by disconnecting an existing branch from the original topology and making new connection. These new topologies are specified in Table 5.1.

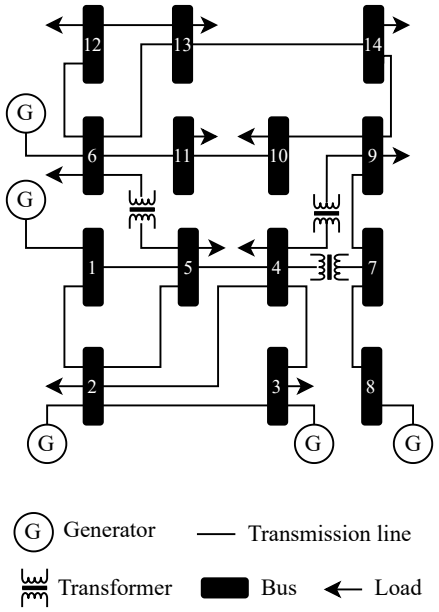

Figure 5-3: Single line diagram of IEEE 14 bus system [1]

Table 5.1: Connection and disconnection of branches for topology changes

Topology number	Disconnected branch From bus – To bus	Connected branch From bus – To bus
1	5 – 6	1 – 6
2	6 – 13	6 – 14
3	4 – 9	4 – 10
4	2 – 4	3 – 5

The obtained data considering these togologies are utilized for training and testing of the ML algorithms and the results are presented in the next section.

5.1.6 Complexity and scalability of ML algorithms

In this thesis, XGB and RF algorithms are trained offline. The relation between the number of features and the number of system buses is given by (5.18). The training time complexity of both algorithms depends linearly on the number of features, n_x, the number of samples, m_{total}, and the number of trees, n_t. According to [100] and [108], the training time complexity can be expressed as:

$$O(n_t m_{total} n_x log(n_x)) \tag{5.20}$$

During the real-time prediction, the time complexity of RF varies. In the worst case, it is $O(n_t n_x)$ [108]; however, in general, it is $O(n_t n_d)$, where n_d is the maximum depth of the trees. This complexity arises from traversing each tree and reaching the corresponding leaf nodes. On the other hand, XGB has a prediction time complexity of $O(n_t n_d)$, as the maximum depth of the trees is typically limited [108]. So, in the context of real-time prediction, the prediction time of both algorithms depends on n_d and n_t, which are specified by hyperparameter optimization. Besides the results of hyperparameter optimization, the computational resources also impact the prediction time of the algorithms. In our case, with Intel Corei7-5500U CPU @ 2.40GHz and 8GB of RAM, the time for real-time prediction was in milliseconds for the IEEE 14 bus test system.

However, it is important to note that the accuracy of the algorithms depends on how well the most relevant features are selected. Including irrelevant features can compromise the algorithm's accuracy and may require additional samples for training. Additionally, increasing the number of samples further increases the time complexity of decision tree-based algorithms during the training stage.

5.1.7 Simulation results

In this section, various possible scenarios of single/multi-bus SLC and single/multi-bus FDIA occurrence are considered for analysing the accuracy of the proposed methodology.

The consecutive optimal power flows were run to get the true values of states

and measurements over the time interval with 100 time samples. A noise having Gaussian distribution with zero mean and 0.001 standard deviation is added to true measurements to get the observed measurements. The higher value of standard deviation of measurement noise, might affect the detection algorithm precision. In our case, the effect of change in standard deviation, up to 5 times of its initial value, is negligible. To model a BD, corresponding measurement is corrupted with a random error which does not fall under the predefined Gaussian distribution. For modeling SLC, specified amount of the load is curtailed at the desired time instant during the execution of the consecutive optimal power flow. In the case of FDIA, the observed measurements are modified according to the attack vector. SE is carried out under normal and abnormal operation, and afterwards the proposed algorithm for anomaly detection, classification and identification of its origin is executed.

The overall data set contains numerous SLC and FDIA scenarios. For SLC, following scenarios have been considered:

- Single-bus SLC is simulated for different buses.
- For each bus, single-bus SLC is simulated numerous times; every next time, different portion of the load has been curtailed from the corresponding bus.
- Multi-bus SLC is simulated for different combinations of buses.
- For each combination of buses, multi-bus SLC is simulated numerous times; every next time, different portions of loads at corresponding buses are curtailed.
- All the above mentioned events are simulated considering different topologies.

For FDIA, following scenarios have been considered:

- Single-state FDIA is simulated for different state variables in the system.
- For each state variable under the attack, single-state FDIA is simulated numerous times; every next time, the corresponding element in the attack vector has a different value.

- Multi-state FDIA is simulated for different combinations of state variables.
- For each combination of state variables under the attack, multi-state FDIA is simulated numerous times; every next time, corresponding elements in the attack vector have different values.
- Again, all the above mentioned events are simulated considering different topologies.

The number of the state variables that can be affected by the FDIA depends on how many measurements can be accessed by the adversarial. Although it is feasible for the adversarial to access all the measurements, it is more realistic that it can access and manipulate the measurements within a local area [109]. In this thesis, it has been assumed that the adversarial is capable of manipulating state variables associated with the maximum 4 buses simultaneously. Moreover, FDIA can target both voltage magnitude and phase angle (both in single or multi-state attack scenarios), while in the reported literature researchers mostly have focused on FDIA on voltage magnitudes.

As mentioned in previous sections and shown in Fig. 5-2, the second stage of the classification process is based on ML algorithm. ML algorithm is trained offline and then executed in real-time. Supervised ML algorithms' performance are compared and the results are presented. Presented results are given for the case of SLC and FDIA classification and identification of their origin. In case BD occurs, χ^2-test will detect the anomaly and recognize it as BD, while LNR will be sufficient to identify the measurements corrupted with the BD. The confidence probability level is set to be $\rho = 99\%$ in this thesis.

ML algorithms are developed in Python using scikit-learn and scikit-optimize libraries [4]. Fig. 5-6 to Fig. 5-12 illustrate the performance of the supervised ML algorithms for anomaly classification and identification of its origin. Each figure demonstrates *macro* F1-score and training time without utilizing MRMR method (specified as "WO MRMR" in the figures). Additionally, the results for the case considering the features selected by MRMR method are presented for each ML algorithm (specified as "MRMR" in the figures). Total number of features is 214

and the number of features selected by MRMR method is given for each case study.

5.1.8 Detection of SLC and FDI

An example of anomaly detection by the proposed algorithm is illustrated here. Fig. 5-4 illustrates detection indices value when the system is under normal operation, while Fig. 5-5 demonstrates the case when BD and single bus/state SLC/FDIA happen in the system.

In Fig. 5-4, load at every bus is assumed to linearly decrease during the simulation period from 100% to 95% of its nominal value, making slow changes in the system state. When system is under normal operation, value of the index should be below the specified thresholds. In order to increase (or decrease) the sensitivity for anomaly detection, the threshold can be lowered (or raised). However, if the threshold is set too low, this may increase the number of false alarms. On the other hand, if the threshold is set too high, anomaly presence might not be detected. To properly select the threshold for ADI, γ, it is necessary to run simulations under normal and abnormal operation conditions for each particular test system. In the case of IEEE 14 bus system, extensive simulations have shown that $\gamma = 6$ can clearly distinguish between normal operation and anomalies like SLC and FDIA.

Fig. 5-5 shows the ADI and χ^2-test values when the system is affected by anomaly. BD, SLC and FDIA are happening at bus 14 but at different time instants (not simultaneously). BD_1 is related to BD when power injection measurement value at bus 14 contains 5% error at $t = 5$, and BD_2 at $t = 10$ refers to the situation when this error has been removed. SLC_1 represents 20% load shedding at bus 14 which happens at $t = 26$. SLC_2 corresponds to the situation when the load at bus 14 is restored at $t = 46$. FDIA tends to increase the voltage magnitude at bus 14 for 0.05 p.u. starting from $t = 71$ and persist till the end of simulation. ADI also might be applicable for higher error values of BD, but χ^2-test is resilient enough for this anomaly. It is obvious that ADI is highly capable of detecting SLC and FDIA anomalies, while they bypass the conventional BDD. Yet, this method is not able to classify the occurred anomaly as SLC or FDIA. This is the main concern within this research and application of ML algorithm is proposed to address it.

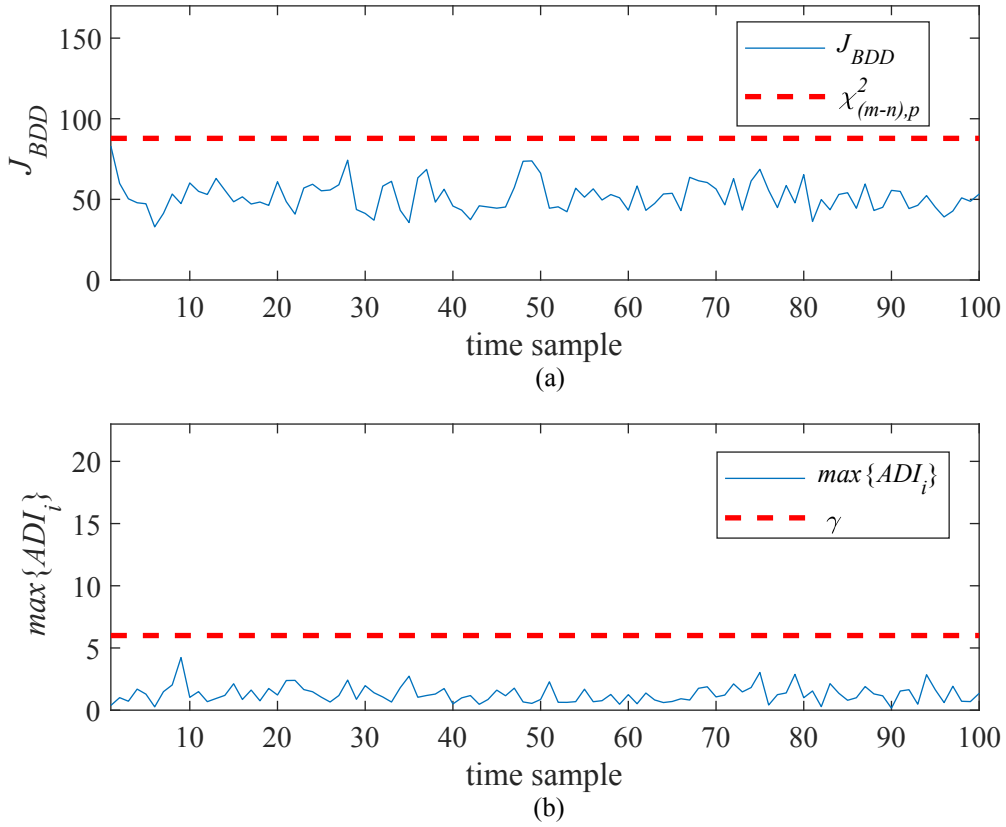

Figure 5-4: Detection index value when the system is under normal operation: (a) χ^2-test (b) ADI

5.1.9 Classification and identification of single bus/state SLC/FDIA

In this case study, the performance of the ML algorithms for classification of single bus SLC and single state FDIA are presented. After classification of the anomaly, the ML algorithms are utilized to identify the origin of the SLC or FDIA. As stated before, these anomalies are happening at different time samples (not simultaneously).

As indicated in the Fig. 5-6, if all features are utilized, the classification accuracy of each method is higher than 98%, which can be considered as acceptable. Due to the fact that some of the features might be redundant (superfluous) or less relevant for training of the ML algorithms, MRMR has been applied to select the most relevant features. In the case of single bus/state SLC/FDIA classification using all topologies to gather training and testing data, the number of features selected by MRMR is 70. This has helped to reduce the training time of the ML algorithms.

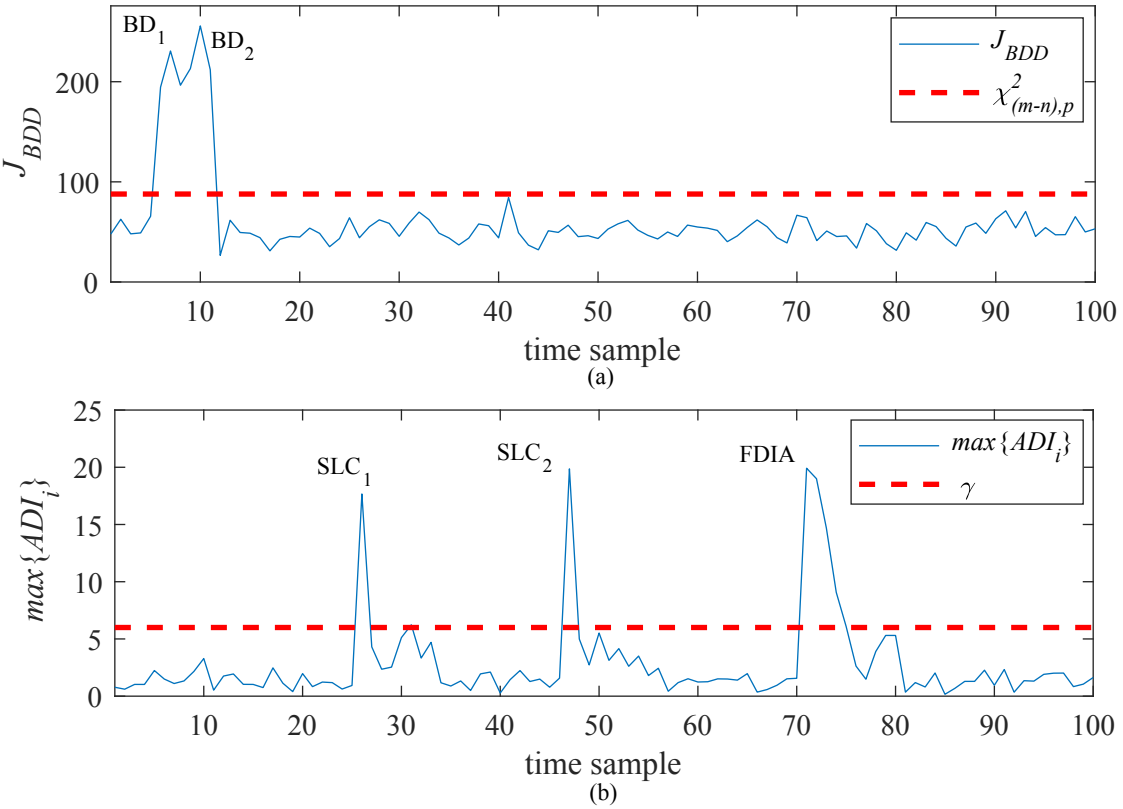

Figure 5-5: Detection index value when the system is affected by anomaly: (a) χ^2-test (b) ADI

Although the accuracy of LR and KNN methods slightly decreases, in the case of RF and XGB methods the accuracy remains the same.

The results given in Fig. 5-6 correspond to the case when both training and testing data set contains the data obtained under 5 different topologies. This means that ML algorithms are tested using the same network topologies for which they have been trained. To check how ML algorithms perform against untrained network topologies, 3 out of 5 topologies have been used in the training phase of the ML algorithms, while in the testing phase ML algorithms are tested using the data corresponding to the other 2 topologies. The results are shown in Fig. 5-7.

It is clear that all examined ML algorithms show an acceptable accuracy if they encounter the data corresponding to the topologies which have not been used in their training phase. This means that SLC and FDIA classification can be achieved without retraining the ML algorithm once network topology changes. This is because the proposed methodology excludes the features associated with the branches and utilizes only the features associated with the buses. If MRMR is used to optimize

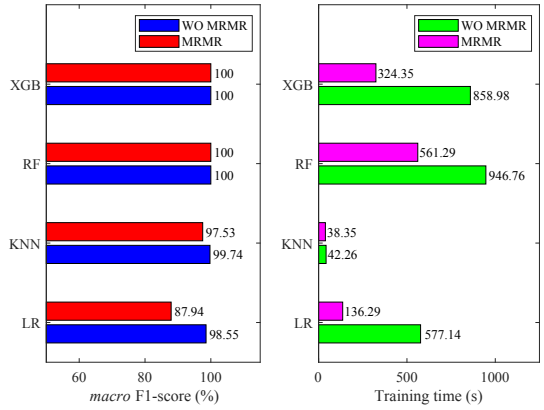

Figure 5-6: Single bus/state SLC/FDIA classification using all topologies to gather training and testing data

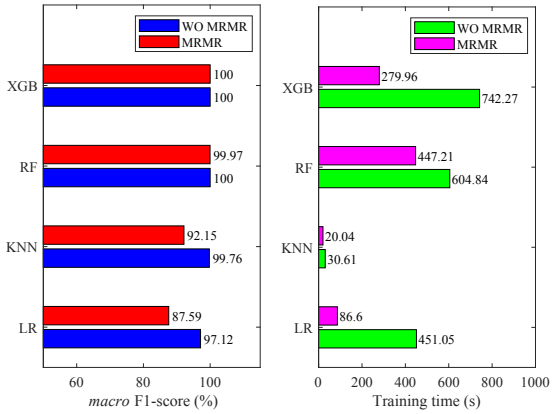

Figure 5-7: Single bus/state SLC/FDIA classification using untrained topologies to gather testing data

the number of features, it will turn out that 80 features would be sufficient. Based on the classification accuracy the algorithms can be sorted in the following descending order: XGB, RF, KNN and LR.

After classification of the anomaly, the bus associated with load experiencing a sudden change, or state variable targeted by FDIA have to be identified. The results for identification of SLC and FDIA origin are presented in Fig. 5-8 and Fig. 5-9, respectively.

Based on the demonstrated results, it is clear that ML algorithms are successful in identifying the bus (or the state variable) which is affected by SLC (or FDIA). Furthermore, MRMR algorithm provides the optimal number of features for identification of anomaly's origin, which in the case of single bus SLC and single state FDIA is 40 and 15, respectively. As in the case of classification, RF and XGB also

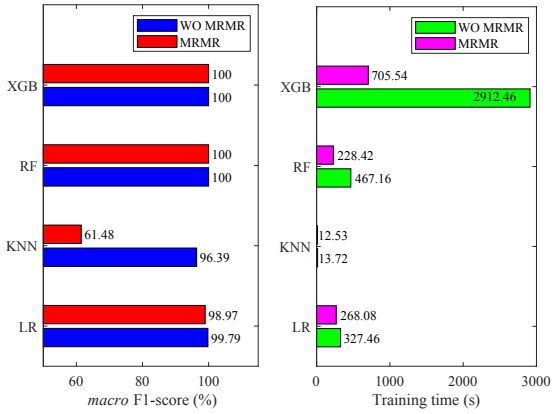

Figure 5-8: Identification of single bus SLC

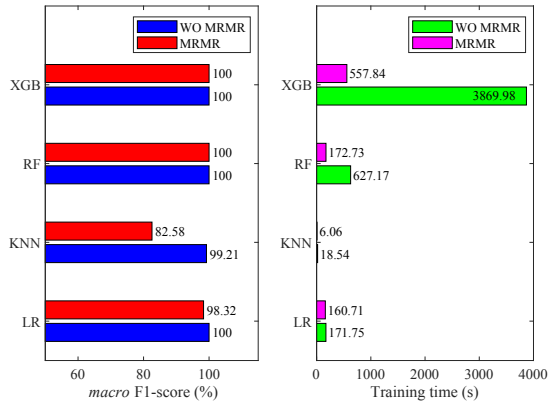

Figure 5-9: Identification of single state FDIA

provide the best identification accuracy, while using MRMR decreases the training time for these two algorithms significantly.

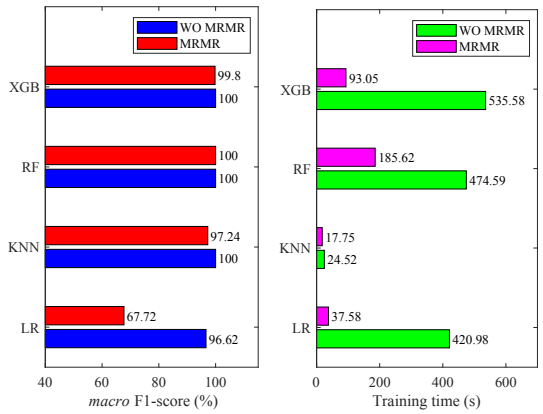

Figure 5-10: Multi-bus/state SLC/FDIA classification using all topologies to gather training and testing data

5.1.10 Classification and identification of multi bus/state SLC/F-DIA

In this case, classification of multi-bus SLC (i.e., SLC is happening at different buses simultaneously) and multi-state FDIA (i.e., multiple states have been targeted by FDIA) is analyzed, as well as identification of the origin of these two kinds of anomalies. The results for classification of multi-bus SLC and multi-state FDIA are presented in Fig. 5-10. The results for identification of the buses associated with loads experiencing a sudden change and the results for state variables targeted by FDIA are demonstrated in Fig. 5-11 and Fig. 5-12, respectively. The number of features selected by MRMR algorithm for classification of anomaly is 30. The number of selected features for identification of multi-bus SLC is 150, while this number is 70 for identification of multi-state FDIA.

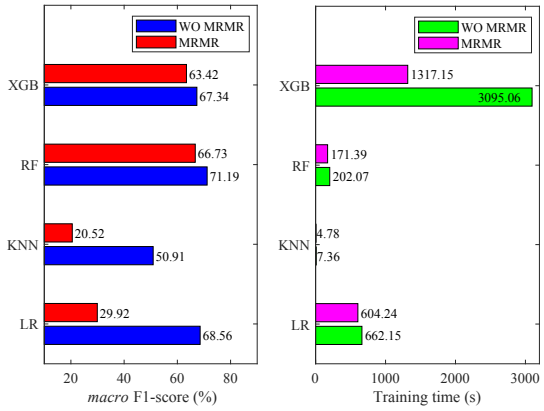

Figure 5-11: Identification of multi-bus SLC

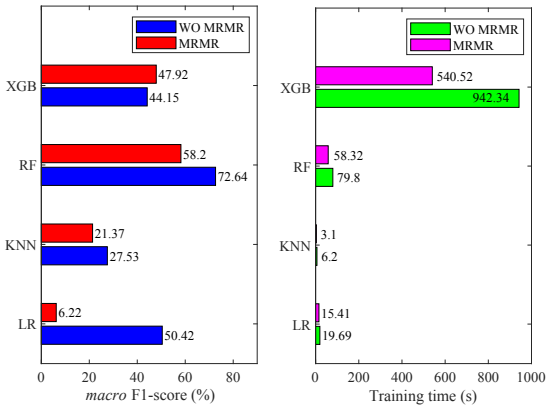

Figure 5-12: Identification of multi-state FDIA

Based on the results presented in Fig. 5-10, it can be concluded that the ML algorithms have a satisfying accuracy for classification of multi-bus SLC and multi-state FDIA. As before, MRMR algorithm helps to reduce the training time. Considering only those features selected by MRMR, RF provides the best classification accuracy followed by XGB, KNN and LR.

The accuracy of the ML algorithms is highly related to the amount of data available for their training. Based on the results, it can be seen that the amount of data is quite sufficient for classification of multi-bus/state SLC/FDIA. However, for accurate identification of the anomaly's origin, an increased amount of data is required due to the fact that the number of possible combinations of their origin (both for multi-bus SLC and multi-state FDIA) is very huge.

5.2 Optimal partitioning

5.2.1 Convergence criterion

One of the trivial ways to stop an algorithm is to set specific number of iterations and hand out the solution when the iterations finish. Obviously, this way can not give a satisfactory result to most problems, specially SE which plays a vital role in power system management. In addition, the problem is not centralized anymore, which hands out the fact that we need to develop and implement a simple yet effective distributed method to deal with it. The following algorithm shows the general approach to optimal partitioned distributed SE with proposed convergence criterion.

5.2.2 Power system partitioning

It is possible to represent the entire power system using an undirected weighted graph and the connectivity between vertices (buses) of this graph (the power system) can

Algorithm 1 distributed SE with convergence criterion

- Optimal partitioning of the system

- Initialization of the distributed SE parameters

- Define area number (AN), state number (SN), each area's measurements and needed data

- Specify convergence criterion parameter (ϵ)

- Do the first iteration and then transmit the needed data between each area

while $||x^t - x_{i,k}^{t-1}|| > \epsilon$ **do**
  Doing local computation
  for $k = 1$ to AN **do**
    for $i = 1$ to SN **do**
      if $||x_{i,k}^t - x_{i,k}^{t-1}|| < \epsilon$ **then**
        $x_{i,k}$'s in the next steps will be equal to $x_{i,k}^{t-1}$
        No need to transfer this data anymore
      else
        keep on sending the needed data
      end if
    end for
  end for
end while

be represented by the following connection matrix (C_L):

$$C_L = \begin{bmatrix} c_{1,1} & \dots & c_{1,M} \\ \vdots & \ddots & \vdots \\ c_{M,1} & \dots & c_{M,M} \end{bmatrix} \tag{5.21}$$

$$s.t.$$
$$c_{i,j} = c_{j,i}, \quad \{i,j\} = 1,2,...,M$$
$$c_{i,i} = 0$$

where M is number of buses and the availability of a physical connection between nodes i and j. So, if there is a connection between nodes i and j, the value of $c_{i,j}$

will be assigned "1", else it would be "0".

We need to define a weight matrix (W_L) with value ($w_{i,j}$) for each element corresponding to connection matrix that introduced in (5.21) such as:

$$W_L = \begin{cases} w_{i,j}, & \text{if } c_{i,j} = 1 \text{ and } i \neq j \\ 0, & otherwise \end{cases} \tag{5.22}$$

$$\{i, j\} = 1, 2, ..., M$$

Based on what has been mentioned in (5.21) and (5.22) the total cost (TC_L) for cutting the connection between buses i and j, can be obtained using $TC_L(i, j) = c_{i,j} w_{i,j}$. Finally, if we want to divide a power system with M bus to K areas, we can formulate the objective function (J_k) for each area (or partition) of the system as follows:

$$min\ J_k = min \sum_{i=1}^{M} \sum_{\substack{j=1 \\ j \notin \phi_k}}^{M} TC_L(i, j) \tag{5.23}$$

$$s.t.$$

$$n(\phi_k) > b_{lim}$$

$$K \geq 2$$

where k indicates number of area ($k = 1, 2, \ldots, K$); i and j indicate the bus number $i, j = 1, 2, \ldots, M$; ϕ_k is the set of buses in area k; $n(\phi_k)$ and b_{lim} are the number of elements in ϕ_k and minimum number of bus we expect to be in each area, respectively. b_{lim} has been set to the minimum number of buses per area presented by conventional partitioning method available in the literature. It is to be noted that the specified constraints in (5.23) make sure that the number of buses in each area are more than a pre-specified threshold. Additionally, considering $K \geq 2$, avoids having only one area which is same as centralized SE.

5.2.3 Simulation results

In this section, the proposed method's results on a test case, i.e. IEEE 14 bus system, are presented. The system has been divided into four areas. Fig. 5-13 shows the topology of the studied test case.

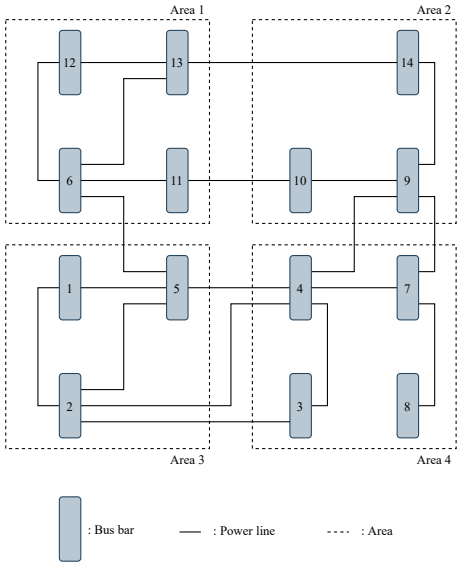

Figure 5-13: Topology of the IEEE 14 bus system

As mentioned in the previous sections, static DC distributed SE has been considered in this study, where the state variables would be only phase angles at each bus. It is to be noted that the measurements are consist of active power flows and injections. The noise covariance for all measurement units has been considered 10^{-4}, and the initial value for state variables are "0". Moreover, bus number "1" has been selected as the slack bus. The simulation has been implemented via MATLAB *R*2018*b* on a computer with Intel(R) Core i5 processor and 8 GB of RAM.

Table 5.2 provides the detailed numerical results of distributed SE for IEEE 14 bus system. It is to be noted that system data and area specification for IEEE 14 bus system is adapted from [110]. The DC centralized state estimation objective value for IEEE 14 bus system is 10.0524.

Also, Table 5.3 presents numerical results for IEEE 118 bus system. For the sake of brevity, the distributed scheme of IEEE 118 bus system is not provided here but it should be mentioned that, the topology of distributed IEEE 118 bus system is adopted from [111]. The objective value for IEEE 118 bus system is 102.7758.

The main reason for considering IEEE 118 bus system is to check scalability of the problem. Additionally, in contrast to the IEEE 14 bus system, in bulk power systems, like IEEE 118 bus, the effects of considering convergence criterion are more visible.

Table 5.2: Numerical results of IEEE 14 Bus system

Methods		Iter	ϵ_1	ϵ_2	CB	OT	OV
Matrix splitting	WOCC	1042	1.26e-3	1.31e-4	8.4546	529.45	10.0565
	WCC	927	4.6e-3	4.16e-4	8.6483	472.15	10.1307
Gossip based	WOCC	2217	2.59e-3	3.15e-4	0.56	1109.06	10.0689
	WCC	1870	1.27e-2	1.22e-3	0.87308	935.873	10.5662
Decomposition	WOCC	45	1.89e-3	3.97e-4	2.89	25.39	10.526
	WCC	42	2.23e-3	5.57e-4	2.77	23.77	10.542
ADMM	WOCC	245	2.39e-2	2.43e-3	0.42828	122.93	12.036
	WCC	213	2.3e-2	2.37e-3	0.38293	106.88	11.8174

Table 5.3: Numerical results of IEEE 118 Bus system

Methods		Iter	ϵ_1	ϵ_2	CB	OT	OV
Matrix splitting	WOCC	65301	2.95	6.3e-2	7319.3546	39969.85	396.283
	WCC	39744	7.19	7.6e-2	3910.7458	23782.75	2194.4171
Gossip based	WOCC	58811	17.0111	0.30187	98.523	29504.02	12130.307
	WCC	37163	17.8056	0.33289	100.0473	18681.55	42593.4768
Decomposition	WOCC	182	2.34e-2	3.51e-3	104.326	195.326	105.287
	WCC	174	3.93e-2	4.86e-3	99.33	186.33	105.374
ADMM	WOCC	1621	1.11	1.6e-2	5.8758	816.37	142.5793
	WCC	998	1.01	1.4e-2	3.9752	502.98	137.014

Results provided in Table 5.2 and 5.3 are separated into two different categories. First one is without modified convergence criterion (WOCC), which considers the centralized objective value and compares it with a threshold value, and the second on is with modified convergence criterion (WCC). Also, the number of iterations (Iter) of different methods, error values compared to centralized solution, computational burden (CB), overall elapsed time (OT) and finally the objective function value (OV) are described for both categories here. Convergence limit ϵ was set to 10^{-6} for all cases. Two different scales were applied for measuring the error of each method's solution compared to the answer obtained using the centralized method. ϵ_1 is the sum of absolute values of difference between centralized and distributed solution (i.e. $\sum |x_{cent} - x_{dist}|$), and ϵ_2 is $max(|x_{cent} - x_{dist}|)$. Computation burden means the time

that has been spent by computer to solve the problem in a distributed manner. As stated in [112], time delay for data transmission in power system can be considered between 0.1 to 0.5 second. So, data transmission delay $tdelay = 0.5$ as the worst case, and the overall time can be calculated using the following equation:

$$OT = (tdelay \times Iter) + CB \tag{5.24}$$

It is to be noted that the time presented here is not per area. One might divide the obtained time to area number (e.g. 4 areas for IEEE 14 and 5 areas for IEEE 118) to calculate the results per area. However this would be wrong, due to non-linear behavior of the solver when the number of variables decreases. Finally, the OV for optimal state variables, which was obtained applying different methods, was evaluated using (3.2).

In order to select the best algorithm amongst the ones which have been presented, features such as scalability, data needed to be transmitted and closer objective value to the centralized solution. Taking into account the mentioned details, the decomposition methods serves the best for the purpose of distributed SE.

After specifying the distributed SE algorithm that has the closest results to centralized SE, the proposed optimal system partitioning has been applied on test systems. For IEEE 14 bus system we have, $M = 14$; $b_{lim} = 3$; $w_{i,j} = 0.01 \ \forall \, i, j$ [111]; And for IEEE 118 bus system we have $M = 118$; $b_{lim} = 11$; $w_{i,j} = 0.01 \ \forall \, i, j$; And MATLAB solver (Sequential quadratic programming (SQP)) has been applied for solving (5.23).

In order to check the security of the system, in Fig. 5-14 and 5-15 a sensitivity analysis on the measurements has been done. For the sake of brevity, the analysis has been done only for IEEE 14 bus system. The value of the measurement unit, has been increase by 10% each separately and the results of all areas have been collected in the one figure. The aim is to identify objective value with bad data and compare it with the chi-square value (chi-square probability distribution function is conventionally used for bad data detection in power system [14]).

The following figures show the results for two case. Fig. 5-14 is related to the

partitioning which is normally used in the literature (e.g. [113]) and the system configuration is as follows: Area 1 = {6 11 12 13}, Area 2 = {14 9 10}, Area 3 = {1 2 5}, Area 4 = {3 4 7 8}. The number of bad data detection is 7 in this case.

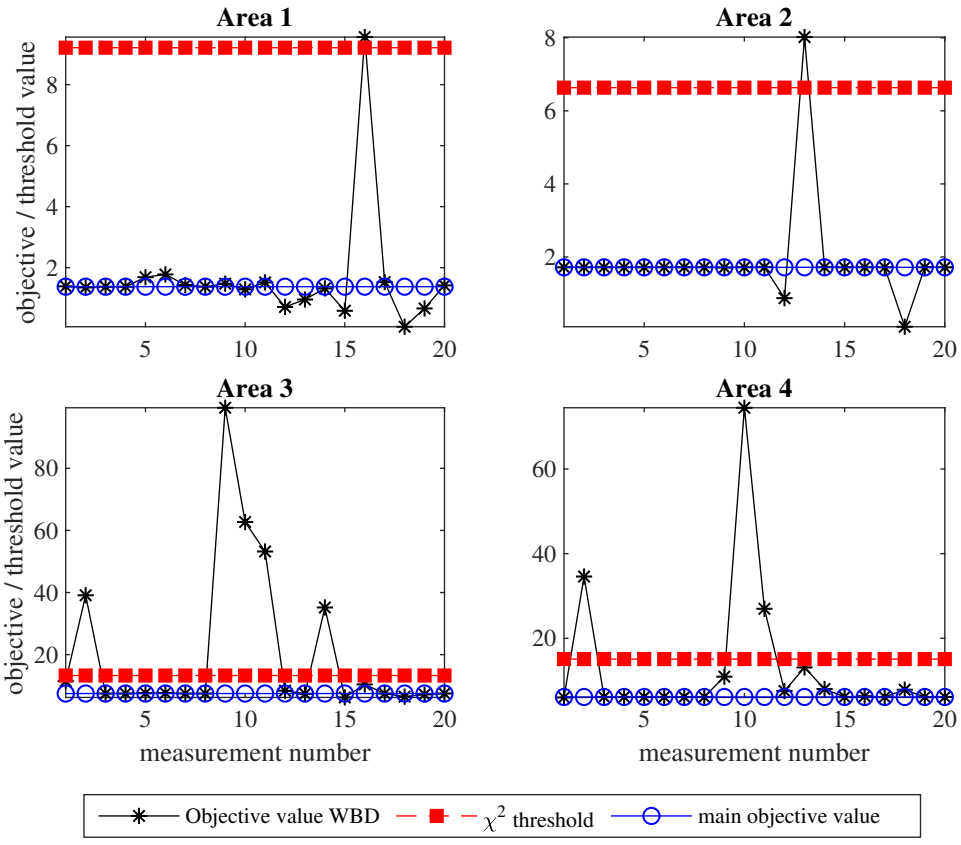

Figure 5-14: Objective value of areas by increasing each measurement value 10% (each at a time) for partitioning case 1

Fig. 5-15 is related to the proposed partitioning, the case when the system configuration is as follows: Area 1 = {6 12 13}, Area 2 = {14 11 10}, Area 3 = {1 2 5 3 4}, Area 4 = {9 7 8}. The number of bad data detection, similar to case 1, is 7 as well.

It is to be noted that, there might be a case, that two areas are going to have residuals more than the chi-square threshold (which means there is a bad data), at the same time. In this case it will be counted as one. Additionally, some of the measurements have zero value, so there is no change in their value, in whole 4 areas. It is clear that the overall security of the system has not changes in both cases has not changed.

Finally, Table 5.4 compares the numerical results for case 1 and 2. Second case,

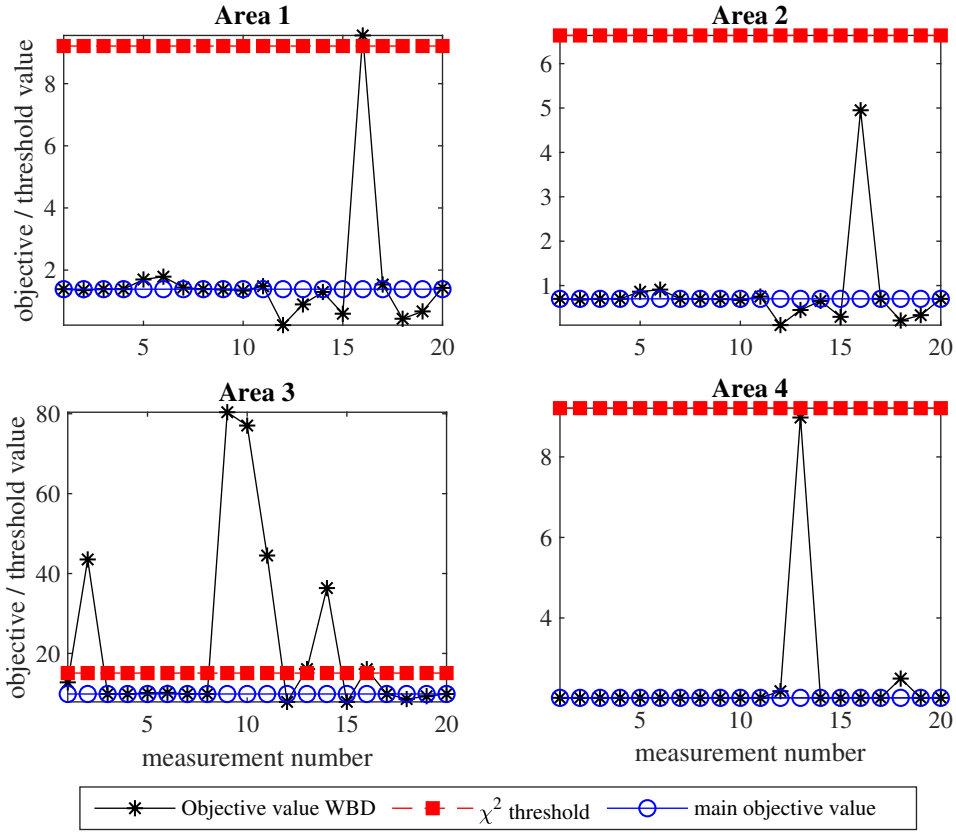

Figure 5-15: Objective value of areas by increasing each measurement value 10% (each at a time) for partitioning case 2

Table 5.4: Numerical results for comparing case 1 and case 2

IEEE 14	Iteration	$\sum J_k$	distributed SE Objective	Error (%)	CB
case 1	42	0.18	10.56	4.85 %	2.77
case 2	38	0.14	10.23	1.76 %	2.1
IEEE 118					
case 1	174	0.38	105.374	2.59 %	99.33
case 2	173	0.36	105.315	2.47 %	86.74

which is related to system optimal partitioning has led to less iteration number that consequently results in less data communication and faster implementation. Then the sum of partitioning objective value for all areas ($\sum J_k$) is presented. The individual partitioning objective has been evaluated for each area using (5.23), to compare between partitioning scheme available in the literature and the proposed one. The obtained $\sum J_k$ result for case 2 it better than case 1. At the same time, due to decreased number of auxiliary variables due optimal partitioning, case 2 has

lower objective value compared to case 1, which is closer to the centralized solution. Then the error percentage that shows the relative error of distributed SE objective compared to centralized SE objective value are presented. Finally, the CB for both cases of the test systems are presented.

5.3 Summary

In this chapter, the results of implementing the proposed methods within the simulation environment were demonstrated. We presented the development and implementation of a novel algorithm for anomaly detection and classification, harnessing the power of machine learning to improve anomaly detection and classification in power systems, thereby enhancing reliability and resilience.

Regarding distributed SE, we explored the concept of optimal area partitioning, seeking to reduce communication overhead and expedite convergence. We also examined the implications of applying a modified convergence criterion for evaluating the performance of distributed SE methods.

Chapter 6

Impact and Applications

In this chapter, we unveil not only the transformative applications of our proposed methods and tools but also the profound impact they may have on the broader research community. This may motivate researchers to explore, adapt, and build upon our work. As we navigate through the concrete applications of our research, we invite readers to advance the available proposed model. The material available in this chapter focuses on two topics. First is anomaly detection, classification, and identification. And, the second is the application of blockchain in distributed power system state estimation.

6.1 ADCIT

Power system state estimation (SE) plays an important role in energy management systems. Its task is to provide accurate estimates of voltage magnitudes and phase angles for all nodes in the system [14]. SE can be subjected to many types of anomalies, among which bad data (BD), sudden load changes (SLC) and false data injection attacks (FDIA) are common. In order to take proper counter measures by the system operator, anomalies must be reliably detected, classified and identified [4]. To this end, the Anomaly Detection, Classification and Identification Tool (ADCIT) is developed. Detection of anomalies takes place in the first ADCIT stage through application of the Matlab source code. In the second ADCIT stage, classification of anomalies and identification of anomalies' origin is done in Python.

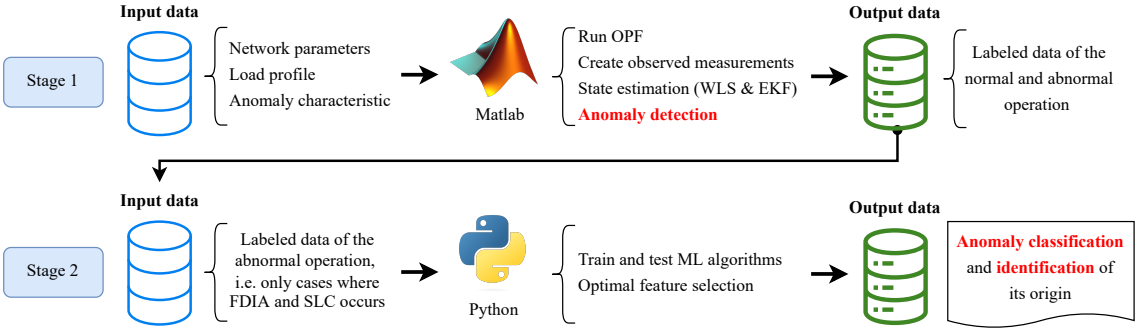

Figure 6-1: General scheme of ADCIT algorithm

Detection of BD is usually done by χ^2-test [11]. However, it is difficult to detect either SLC or FDIA by applying χ^2-test within weighted least squares (WLS) estimator. The ADCIT combines estimated states of WLS and extended Kalman filter (EKF) to set up an anomaly detection index (ADI) capable of detecting both SLC and FDIA; however, ADI cannot discriminate between SLC and FDIA. To classify (or, discriminate) SLC and FDIA correctly, various supervised machine learning (ML) algorithms have been implemented. Moreover, the case when load is abruptly changed at multiple nodes simultaneously (named "multi-bus SLC"), or FDIA is targeting multiple states at the same time (named “multi-state FDIA”) are for the first time considered and ADCIT is capable to correctly discriminate between multi-bus SLC and multi-state FDIA. Furthermore, the features utilized for training the ML algorithm(s) are associated only with the network nodes. This increases the robustness of the algorithms by eliminating the need to retrain the ML algorithm in case of network topology changes. Inside the ADCIT, different ML algorithms are available and they are user-defined. Finally, different types of anomalies can be successfully analysed by the ADCIT.

6.1.1 ADCIT algorithm

The ADCIT algorithm has been implemented in Matlab and Python. Fig. 6-1 demonstrates the general scheme of the ADCIT algorithm. The details regarding each code are presented below.

6.1.2 Matlab: Data preparation and detection

To provide the labelled data for the training of the ML algorithms power system simulations are conducted within Matlab environment. Firstly, raw measurements are generated using the procedure described below. Next, raw measurements are processed by two types of state estimators, namely WLS and EKF, to get the estimated (and, in case of EKF, predicted) electrical quantities. IEEE 14 bus test system [1] has been selected as the benchmark.

MATPOWER, an open-source Matlab extension for solving steady-state power system optimization problems, has been utilized to execute consecutive optimal power flows (OPFs) over the time [114]. Considering that the load at each consumption node is given, the OPF provides nodal voltage magnitudes, active/reactive power flows in branches and active/reactive power injections at generator nodes. These values are used as the true values of measurements. A noise term, having Gaussian distribution with zero mean and 0.01 standard deviation, is added to the true measurements to get the raw measurements.

To simulate a BD case, corresponding raw measurements are corrupted with a random error which does not fall under the predefined Gaussian distribution. For simulating a SLC, a pre-specified amount of load is curtailed at the desired time instant during the execution of the consecutive OPFs. In the case of FDIA, the raw measurements are modified according to the attack vector. To simulate multi-bus SLC or multi-state FDIA, the user can change the setting of parameters *SLC_bus* or *FDIA_state* from a scalar value to a vector. This change has to be made in the *main* m-file. For instance, *SLC_bus* = [5 10 12] means that the SLC is happening at the nodes 5, 10 and 12 simultaneously.

WLS and EKF based state estimations are carried out under normal operating conditions (i.e., quasi steady state) and abnormal operating conditions (BD, SLC or FDIA). Apart from the estimated states, other outputs, such as predicted states, normalized residuals and normalized innovations, are obtained. To detect BD, measurement residuals obtained via WLS state estimation are used to carry out χ^2-test. In case there is no BD, the algorithm will check for SLC or FDIA using ADI [115]. In case ADI value is equal or higher than a specific threshold, anomaly is detected;

otherwise, system is considered to be in the normal operation mode.

It is to be noted that the moment when an anomaly occurs and vanishes can be specified within the code. Additionally, it is also possible to change the test system or network topology; however, it requires further modification of the parameter settings in MATPOWER source file and several m-files.

6.1.3 Python: Classification and identification

Python environment is used for input data pre-processing and application of the ML algorithms for anomaly classification and identification. For the sake of comparison, four supervised ML algorithms, namely Random Forest (RF), Extreme Gradient Boosting (XGB), Logistic Regression (LR) and K-Near Neighbours (KNN) are applied [115].

As mentioned before, to eliminate the need to retrain the ML algorithms when the network topology changes, the features associated with the power lines are excluded and only the features associated with the nodes are utilized. The features associated with the nodes are: a) Nodal measurements and normalized measurement innovations of voltage magnitudes and active/reactive power injections; b) Estimates and predictions of voltage magnitudes, phase angles and active/reactive power injections.

Maximum relevance – minimum redundancy (MRMR) algorithm has been applied for the feature selection [105]. The parallelization function has been included in the MRMR script to utilize multiple cores of the CPU and run tasks in parallel. Accordingly, the most relevant features can be found fast and without additional computational complexity.

For executing the ML algorithms, standard models from the *scikit-learn* library have been applied [116]. All models are trained by tuning appropriate hyperparameters. Hyperparameters are tuned using sequential optimization with gradient boosting as a surrogate probability model of the objective function [95]; the *scikit-optimize* library is used for this purpose.

6.1.4 Illustrative example

An example of the results obtained by ADCIT is demonstrated in Fig. 6-2.

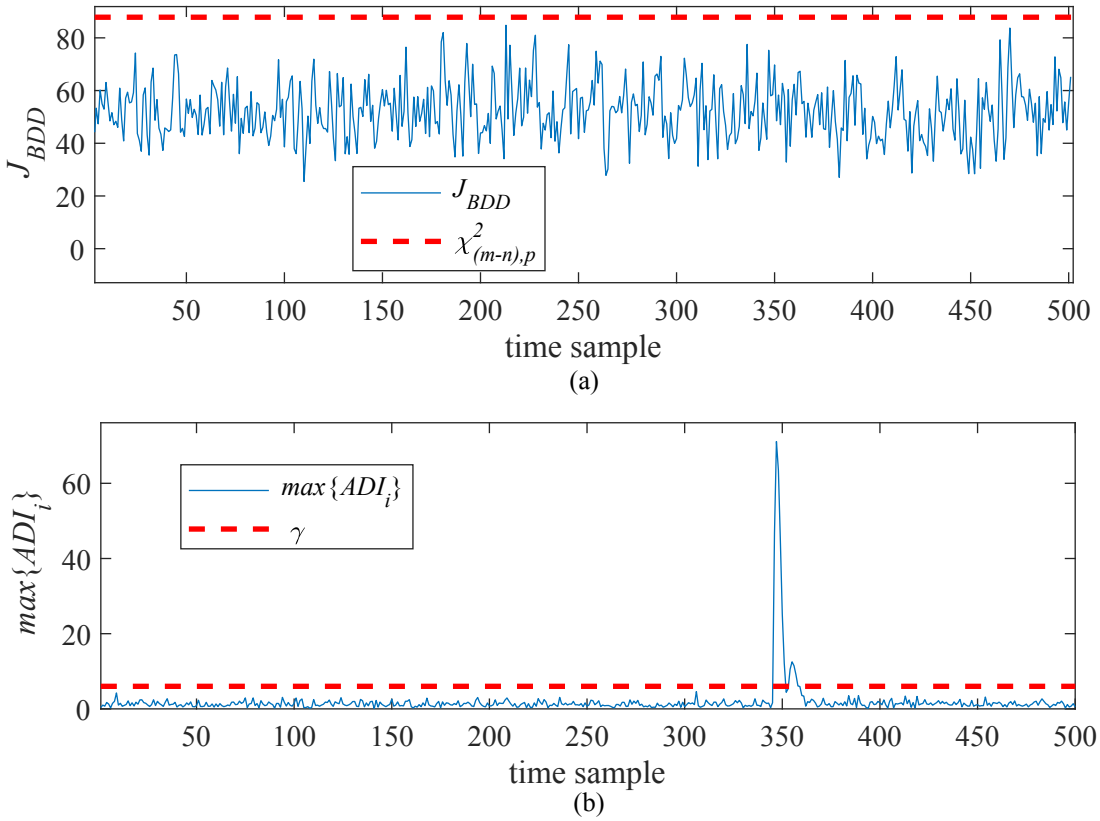

Figure 6-2: Detection tests in the presence of FDIA: (a) χ^2-test (b) Largest ADI test

If $J_{BDD} \geq \chi^2_{(m-n),p}$ holds, then there is a high probability of the existence of a BD. Here, J_{BDD} stands for χ^2-test's objective function; $\chi^2_{(m-n),p}$ corresponds to a value from the χ^2 distribution table with the probability p and $(m-n)$ degrees of freedom; m and n is the number of observed measurements and number of estimated states, respectively.

If $max\{ADI_i\} \geq \gamma$, SLC or FDIA presence is detected. Here, $max\{ADI_i\}$ stands for maximum ADI value, and γ represents the detection threshold that has to be selected to clearly discriminate between normal operation and anomalies [115].

Fig. 6-2 shows detection indices when the system is affected by FDIA. FDIA tends to increase the voltage magnitude at bus 14 for 0.1 p.u., starting from $t = 350$ and persists until the end of the simulation.

It is obvious that the largest ADI test is highly capable of detecting anomaly

presence, while anomaly bypass the χ^2-test. Yet, this test is not able to classify the occurred anomaly according to its type. Therefore, ML algorithms have been utilized for the classification of the anomalies. Table 6.1 summarizes the performances of the ML algorithms in terms of calcification accuracy and training time, for the classification of SLC and FDIA, i.e., to discriminate between SLC and FDIA. [115]. Due to the fact that some of the features might be redundant or less relevant for the training of the ML algorithms, MRMR has been applied to select the most relevant features. In this example, the number of features selected by MRMR is 70 compared to the number of features without MRMR (WO MRMR) which is 214. This has helped to reduce the training time of the ML algorithms. Although the accuracy of LR and KNN algorithm slightly decreases, in the case of RF and XGB algorithm the accuracy remains the same.

Table 6.1: Single bus/state SLC/FDIA classification

ML algorithm	Classification accuracy WO MRMR (%)	Training time (s)	Accuracy using MRMR (%)	Training time using MRMR (s)
LR	98.55	577.14	87.94	136.29
KNN	99.74	42.26	97.53	38.35
RF	100	946.76	100	561.29
XGB	100	858.98	100	324.35

6.1.5 Software impacts

Accurate anomaly detection, classification, and identification are of great importance for power system state estimation. The impacts of the ADCIT are twofold. Firstly, it can be applied as an educational tool. It provides an opportunity for the researchers to observe the adverse effects of the anomalies on the state estimates, and to analyze how the ADCIT enables anomaly detection, classification, and identification in order to avoid these effects [117]. In another word, the researchers can modify/extend the ADCIT to implement their own ideas. This means that the ADCIT can be used as a platform for the future research work.

Yet another impact of the tool is its capability for industrial implementation. Without the requirement of any additional hardware installation, the ADCIT can

be integrated within the energy management systems in power system control rooms [14]. The ADCIT enables a better situational awareness in the presence of the anomalies, which are typical in case of system emergencies [4, 15]. Besides, the specification of the anomaly type (i.e., anomaly classification) by the ADCIT will assist the system operator for proper decision making.

Creating a graphical user interface, considering other types of anomalies such as network parameter errors, and designing suitable countermeasures against anomalies, can be considered as future directions for the development of ADCIT.

6.2 Application of blockchain

BC is a digital ledger of transactions distributed across the network of computer systems, with atomic changes to the database. The integrity and tamper-resistance of the transaction logs are assured because of the cryptographic hash linked among the blocks. BC is usually assumed to be decentralized architecture maintained by individual parties. Each node of the network owns a copy of the BC. Each BC block contains transactions, and every time a new transaction occurs on the BC, it is broadcasted to all nodes and added to a block along with other transactions waiting to get committed in a block. This technology has developed over the last decade and can be categorised as private, public or consortium BC, each further divided by permissioned or permissionless. As shown in Fig. 6-3, every new block N generated at time T contains information from the previous $N-1$ block generated at time T', where $T > T'$.

6.2.1 Consensus algorithm for decentralized ledger

BC is a peer-to-peer network of nodes that functions individually without any central authority. Each node of the network can function individually, i.e., update ledger (creating and adding a block to the BC) and broadcast new block to the other nodes of the network using the gossip protocol [54]. The nodes verify the broadcasted block's validity, and have to either accept or reject the proposed block, thus reaching a consensus. In distributed ledger technology, there exists a funda-

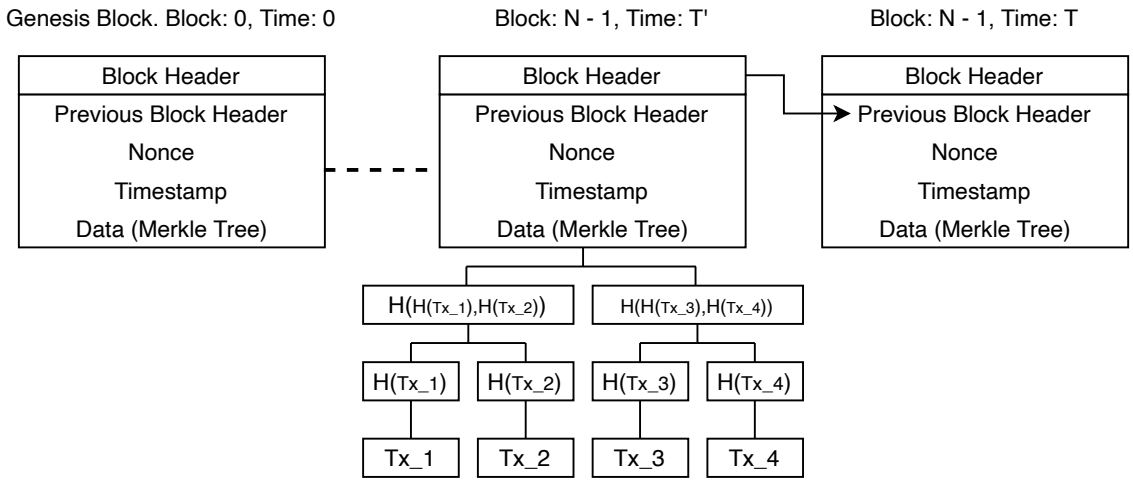

Figure 6-3: Data organization in blockchain

mental problem of reaching consensus. Majority of the BC projects use any of the three most common consensus algorithms, i.e., proof of work (PoW), proof of stake (PoS), and Byzantine fault tolerant. Similar to Bitcoin, Ethereum uses a PoW consensus algorithm. In December 2020, Ethereum 2.0 was launched, which uses PoS consensus. In PoW BCs, block creators (which are called miners) are rewarded with mining rewards along with transaction fees included in the block. This mining reward is the incentives for using computation power and electricity in finding the correct nonce within the target range.

Miners have to perform computation by running a hash of block's content and incrementing a nonce until it produces a value less than the target. Nonce is an integer that starts from 1 and increments until it produces a hash of block's content less than specific target value [58]. Generating a hash on an arbitrary size input is a one-way function that produces a fixed output length [118], i.e., given input, we can generate an output of fixed-length, but not vice versa. The hash function used is cryptographically secure and with brute force there exists a potential solution with complexity of $O(2^n)$ for $H(m) = H(m')$, where H is a SHA function [119] on an input m and m' and $m \neq m'$. This means that for a fixed output length on n, for example, $n = 256$ in the case of SHA256, the probability of success is $k/2^n$, where k is a number of queries [120].

6.2.2 Ethereum Architecture

A computer (node) can be a full node or light node [121] running an instance of the Ethereum BC. A full node stores the entire BC data and can serve any request. It verifies all blocks and states and can propose a new block to append on the ongoing chain. Light node stores only the header of the chain and can verify the validity of the state roots' data in a block header. To interact with the DApp, clients should interact with the BC by running a full node by itself and using ethereum clients, like Geth, OpenEthereum, etc., to interact with the network. Ethereum BC has grown and consumes a significant storage amount and can be difficult to run a full node. Therefore, via a third-party platform like Infura, Alchemy, etc., [122, 123] provides application programming interface (API) to interact with ethereum BC feasible.

Ethereum comprises two main components:

- *Database:* All activities on the network are recorded on the BC in the form of a transaction. Sending cryptocurrency from one address to another is recorded in a transaction with valid signatures and broadcasted to the network where other nodes commit to a block after verification. PoW consensus algorithms make sure that all the nodes in the network have the same BC data as all the valid transaction data. The data are stored in the form of a Merkle Patricia Tree. There are two types of addresses in Ethereum, Externally Owned Account (EOA), controlled by private keys and Contract Address, controlled by contract code. When a smart contract code is compiled and deployed from EOA, a contract address is created, and bytecode is stored in it.

- *Code:* The smart contract is stored on the BC in a contract address in the form of code, known as byte code. The codes in contract addresses execute contract when a transaction is sent from EOA to contract addresses.

For each transaction on the Ethereum BC, there is a fee known as Gas for executing transactions. Once a transaction is added to the block, the transaction fee goes to the miner as a reward for using computational resources. Gas is a unit to measure computation difficulty in Ethereum Virtual Machine (EVM). Gas is charged only when data are modified on the BC, i.e., reading and accessing data are

not chargeable. Once the sender signs a transaction and broadcasts it, the Ethereum protocol debts gas fees in a fraction of ethers from the Ethereum account, lack of required gas amount will not allow the transaction to be execute. If there are no fees, attackers can flood the node's memory pool with bogus transactions, causing distributed DoS attacks. Gas is not fixed for the transaction but it is variable and depends on the computational difficulty of a smart contract. The sender of the transaction pays gas, and the miner who mines a block receives gas. Miner receives all the transaction gas that he includes in the block along with the block generation reward. Miners set the price of gas based on the computational power of the network required to process transactions and smart contract.

Since ether is not stable in value but sees daily change, therefore gas is a relative price converted to ethers based on the load on the network. In a congested network, the gas price will increase for each unit of gas. So there is a gas price, i.e., how many units of ether are transactor willing to pay for one gas unit. Each opcode in Ethereum has a cost. The total cost of the contract is the summation of all the opcodes [124].

The EVM is a virtual stack embedded within each full Ethereum node that allows anyone to execute arbitrary bytecodes and plays a crucial role in the consensus engine of the Ethereum system. It allows anyone to execute arbitrary code in a trustless environment in which the outcome of execution can be guaranteed and is entirely deterministic. When you install and start the Geth, parity or any other client, the EVM is started, and it starts syncing, validating and executing transactions. The EVM is Turing complete, i.e., capable of performing any algorithm.

6.2.3 Data Verification

Before broadcasting the data that contains the formation of the transaction to other nodes in the network, the data should be signed using the private key. A signature is required to prove that the sender of the data are genuine and not an imposter who signed the message without the private key. BC uses asymmetric cryptography based on public key infrastructure. Like a physical signature, digital signatures are used to authenticate electronically a document's contents like pdf, emails, etc. [125].

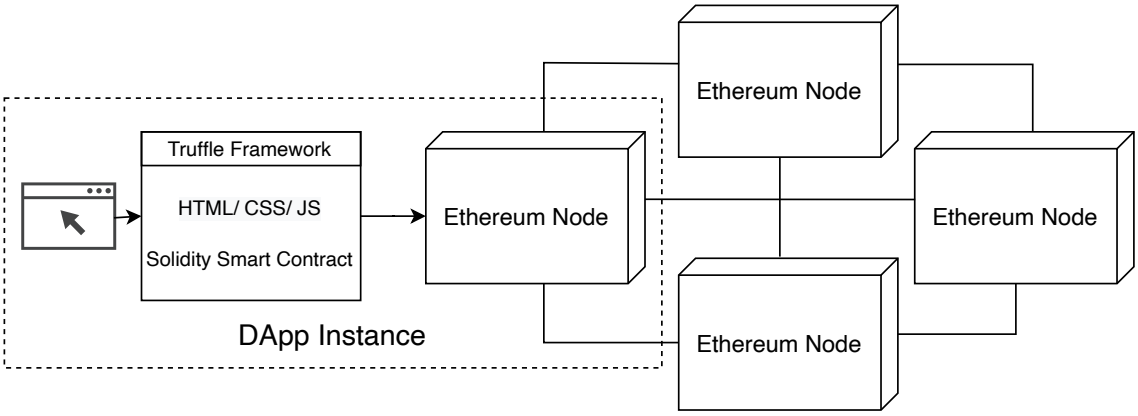

Figure 6-4: Ethereum network structure

In the BC network, each node has its pair of public and private keys, and the public key is shared with all the other nodes. Owning a private key is equivalent to owning or controlling a node associated with its public key and accessing the activities restricted to it.

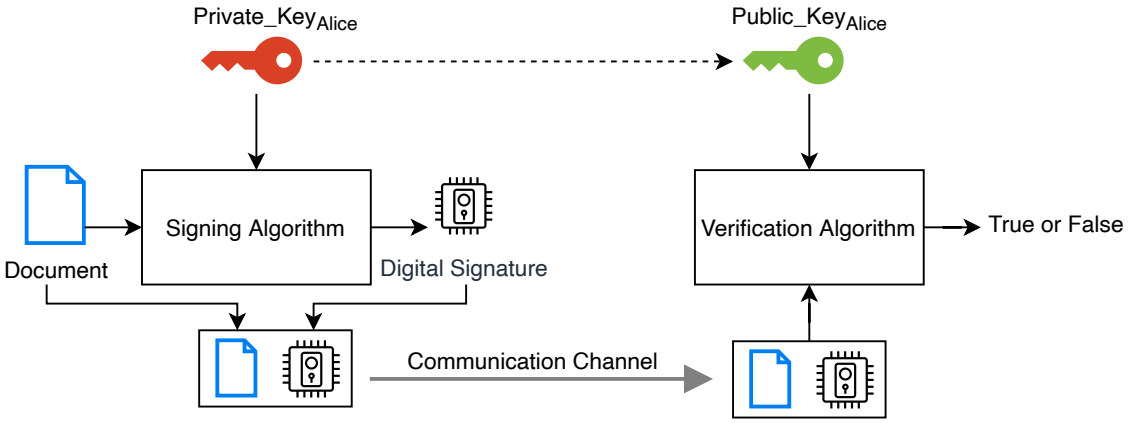

Figure 6-5: Data verification using private and public key

To sign a message (or data), a function is calculated using the private key of the document's sender. The recipient's using the public key of the sender, can verify if the document is correct and not tampered.

6.2.4 Asynchronous data transfer

The combination of renewable energy sources and information and communication technology (ICT) changes the power system's nature from a physical system to a CPPS [15]. Therefore, the physical part consists of a power grid, and the cyber part comprises a control and computation layer. The physical layer consists of physical

elements such as generators, transmission lines, transformers, etc. On the other hand, the cyber layer is responsible for computation, analysis and assessment of the power grid, and includes elements such as sensors, communication medium, control system, etc. [15]. A CPPS encounters different types of cyberattacks, such as DoS and FDI. However, latency attack has the potential to be considered as a new type of attack in the area of the power system, while it is already well-known for wireless network community [126]. The power system undergoes a time delay of several milliseconds, while increasing this latency or time delay maliciously may lead to the power system instability [126]. Although the application of distributed methods and implementation of a BC based communication network may dissolve the issue, still there would be a chance that the latency happens for the system. To study such a case, we have considered a delay in data transfer between areas in a randomized manner. In other words, at some iterations, an area randomly (based on uniform probability distribution) will be selected so as not to update its state variables. The comparison of the distributed SE results with and without delay is presented in section IV.

6.2.5 Problem formulation

Suppose that we have divided the power system into N areas, having z_N measurements composed of power injection, power flow and voltage magnitude. Considering x_k as the state variables related to area k and $\tilde{x}_l$ as the auxiliary variables estimated by area k related to its neighboring area l, one can rewrite (4.6) into the following equation:

$$\min_{x_k} \; f_k(x_k) + \sum_{l \in \Lambda_k} f_{kl}(x_k, \tilde{x}_l), \tag{6.1}$$

where Λ_k indicates the set of all neighboring areas of k^{th} area. It is clear that (6.1) is composed of two statements. The first statement is related to the measurements that the physical equation for calculating them only requires the state variables

inside the area and can be written as follows:

$$\begin{aligned} f_k(x_k)=&\textstyle\sum_{i\in\Lambda_k^v} W_{k,i}^v(v_{k,i}^m-v_{k,i})^2 \\ &+\sum_{i\in\Lambda_k^P} W_{k,i}^P(P_{k,i}^m-P_{k,i}(.))^2+\sum_{i\in\Lambda_k^Q} W_{k,i}^Q(Q_{k,i}^m-Q_{k,i}(.))^2 \\ &+\sum_{(i,j)\in\Lambda_k^{PF}} W_{k,ij}^{PF}(P_{k,ij}^m-P_{k,ij}(.))^2+\sum_{(i,j)\in\Lambda_k^{QF}} W_{k,ij}^{QF}(Q_{k,ij}^m-Q_{k,ij}(.))^2, \end{aligned} \tag{6.2}$$

where i and j are arbitrary buses; Λ_k^v, Λ_k^P, Λ_k^Q, Λ_k^{PF} and Λ_k^{QF} indicate the set of voltage, active power injection, reactive power injection, active power flow and reactive power flow measurements in area k, respectively; $W_{(.)}^{(.)}$ weighting factor for the measurements; $P_{(.)}^m$, $Q_{(.)}^m$ and $v_{(.)}^m$ are the active power injection or power flow, reactive power injection or power flow and voltage observed measurements, respectively; While $P_{(.)}$, $Q_{(.)}$ and $v_{(.)}$ are the physical equations of these measurements. These physical equations governing the power system are provided in appendix.

The second statement of (6.1), is related to the measurements in k that need to receive state values regarding the buses in connection with the neighboring area l. It is to be noted that, for calculation of the physical equations regarding these measurements, we use the auxiliary variables:

$$\begin{aligned} f_{kl}(x_k,\tilde{x}_l)=&\textstyle\sum_{i\in\Lambda_{kl}^P} W_{kl,i}^P(P_{kl,i}^m-P_{kl,i}(.))^2 \\ &+\sum_{i\in\Lambda_{kl}^Q} W_{kl,i}^Q(Q_{kl,i}^m-Q_{kl,i}(.))^2+\sum_{(i,j)\in\Lambda_{kl}^{PF}} W_{kl,ij}^{PF}(P_{kl,ij}^m-P_{kl,ij}(.))^2 \\ &+\sum_{(i,j)\in\Lambda_{kl}^{QF}} W_{kl,ij}^{QF}(Q_{kl,ij}^m-Q_{kl,ij}(.))^2+\sum_{i\in\Lambda_{kl}} W_{k,i}^{\tilde{v}}(v_{l,i}-\tilde{v}_{l,i})^2 \\ &+\sum_{i\in\Lambda_{kl}} W_{k,i}^{\tilde{\theta}}(\theta_{l,i}-\tilde{\theta}_{l,i})^2, \end{aligned} \tag{6.3}$$

where $\tilde{\theta}_{(.)}$ and $\tilde{v}_{(.)}$ are the auxiliary variables. It is worth noting that the last two statements in (6.3) are utilized to provide a consensus for this minimization function.

6.2.6 Proposed Blockchain Solution

Building distributed SE's data transmission architecture based on BC provides a security feature of the technology to transfer data among system areas. BC inte-

gration can ensure honesty in the system as the transaction's sender can only sign each transaction.

A PoC is developed on the Ethereum test network and deployed using Truffle framework and Ganache. Ethereum provides tools to build smart contracts and decentralized applications without any downtime or any third-party interference. Truffle Suite is a BCs development environment, testing framework, and asset pipeline using the EVM. Ganache [127] is a personal BC for Ethereum and Corda based distributed application development. Utilizing Ganache and Truffle, the entire DApp can be developed in a safe and deterministic environment. The code repository containing open source prototype is available in [128].

The EVM has separate storage areas:

- All contracts have state variables, and the state variables are stored on the BC, i.e., the data are recorded into the BC itself. When the contract executes some code, it can access all the previously stored data in its storage area.
- Memory holds temporary values and only exists in the calling function and has less gas price because the stored memory gets erased between calls. Gas price increases with the size of memory scaling quadratically. Though, comparatively cheaper than storage.
- The stack holds small local variables, and here the computations happen. This data can only hold a limited amount of values up to 1024 small local variables.
- Logs store data in an indexed structure with mapping, and with filters, specific data can be accessed. Logs are inaccessible to contract but are mainly used for events that occur on the BC.

6.2.7 System Overview

The proposed BC solution focuses on establishing a secure architecture of transferring arbitrary data for every iteration among the DSE areas based on the established connections on the BC. Fig. 6-6 shows the main participating entities of the system:

- *distributed SE areas*: The control center at each area is responsible for receiving data and then, calculate SE and after that send data to another area.

- *Auditor*: Provides public key infrastructure [129] to all distributed SE areas and is responsible for maintaining smart contracts on the BC and can establish or demolish connection between two areas. In other words, only the auditor can establish communication between two or more than two areas by sending a transaction to the smart contract address that sets communication to *true* between areas on the smart contract. Auditor is like a supervising body of the infrastructure of the distributed SE network. Although, it is responsible for deploying contracts on the blockchain, the distributed SE areas can communicate, i.e., transfer date, with each other via smart contract without interference from the auditor. If any issues arise on the BC, the auditor can resolve this issue with the BC. The distributed SE data transactions are independent, and the auditor is not involved.

6.2.8 System Design

On the BC, two contracts are deployed. First, to establish/demolish connection between the areas. Second, to transfer data per iteration between the areas within the established connection. The following section describes the details.

Establishing/Demolishing Connection

The auditor manages the connections between areas through algorithm 2. The smart contract emit event upon each connection change to inform all the areas.

Data Transfer

Algorithm 3 smart contracts listens to all the transaction call of the first deployed smart contract and as per update the state of the connections of this smart contract. This algorithm takes four parameters i.e., *sender*, *receiver*, *iteration* and *payload*. Each area in our case study has a different data payload size (i.e., state variables which needs to be transferred). With each iteration, data are passed as arrays of

float integers as string type because it is impossible to pass a negative number in a smart contract. With each transaction of the iteration, the transaction event is emitted and notified to the receiving area, who can process the data off-chain as peruse.

Algorithm 2 Establish/Demolish area Connection

Input: address_deployer, address_from, address_to
Initialization: $connection(from, to) \leftarrow bool$
if $(msg.sender \neq address\_deployer)$ **then**
  from ← address from
  to ← address to
  if $(from \neq to)$ **then**
    if $(connection(from, to) \neq True)$ **then**
      Set connection(from,to) ← True
    else
      Revert and show error "Connection exist"
    end if
  else
    Revert and show error "No Self Connection"
  end if
else
  Revert and show error "Only Owner Access"
end if=0

Algorithm 3 Data Transfer

Import: 'Establishing Demolishing Connections'
Input: address_sender, address_to, interation_number, data_String
if $(msg.sender \neq address\_sender)$ **then**
  if $(connection(from, to) = True)$ **then**
    Call function to transact these values on blockchain
    Notify transaction in the network
    Apply the transferred data in the current iteration for state estimation
  else
    Revert and show error "No Connection"
  end if
else
  Revert and show error "Only msg.senders"
end if=0

6.2.9 Security

This architecture provides security because each transaction requires a transaction signature. The connection can only be established by the smart contract owner as there is a specific check-in of the smart contract that requires signature verification. Signature is created using the private key, and address generation also requires a private key. Therefore, losing the private key, especially by the auditor, i.e., controller of the architecture, can compromise the whole system.

6.2.10 Simulation results and discussion

In this section, the test case (i.e., IEEE 14 bus system [1]) results are presente utilizing the proposed method. The system has been divided into four areas and Fig. 6-6 shows the topology of the studied test case.

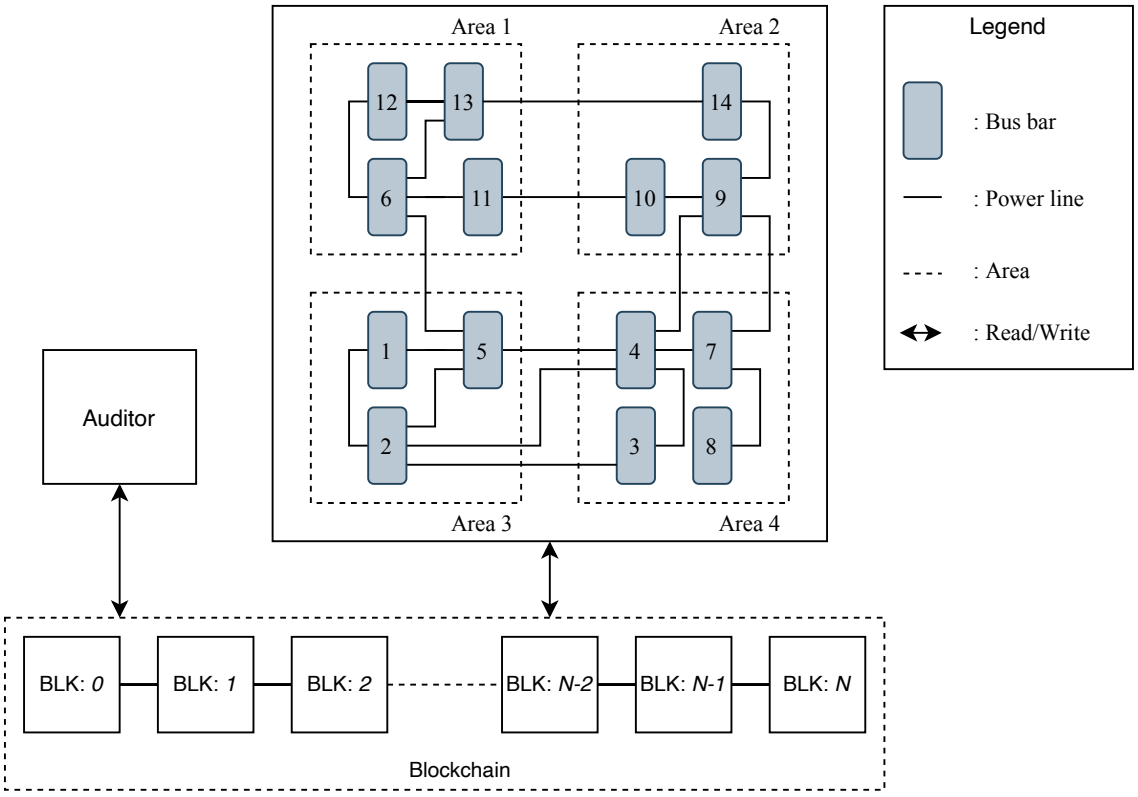

Figure 6-6: Distributed topology of the IEEE 14 bus system [3] integrated with blockchain

In this research, AC SE has been considered, where the state variables would be voltage magnitudes and phase angles at each bus. The number of state variables

and measurements are 27 and 41, respectively. The weighting factor for all measurement units has been considered equal to 10^4. In order to solve (6.1), MATLAB (version *R2018b*) solver (Sequential quadratic programming) has been applied and for initiation of the optimization process the initial value for state variables have been set to flat start, i.e., voltage magnitude of "1" and phase angle value of "0". Moreover, the bus number one has been selected as the slack bus with phase angle zero. To evaluate the prototype's performance, the smart contract was deployed on a local BC server and interacted with the python application. The experiments were performed on a computer with memory 16 GB 2400 MHz DDR4, Intel Core i9 running @2,3GHz.

As mentioned before, we have considered two different cases. The data transfer between areas are simultaneously in the first case and with latency (time delay) in the second case. The graphical and numerical results of both cases are presented.

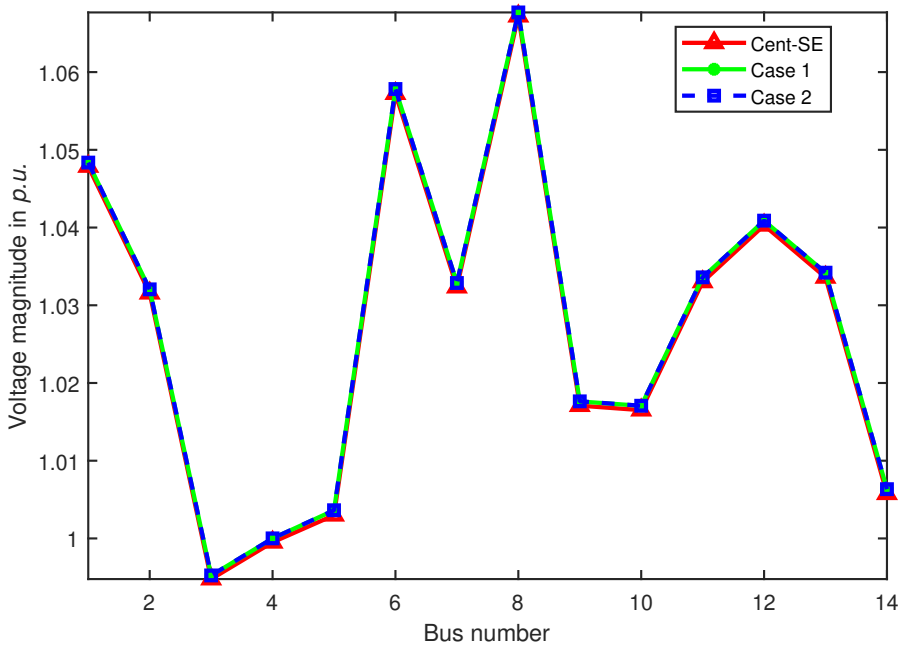

Figure 6-7: IEEE 14 bus system voltage magnitude for centralized (Cent-SE) and distributed state estimation interacting with blockchain (case 1 and case 2)

Fig. 6-7 and Fig. 6-8 represent the comparison of centralized and distributed estimated voltage magnitude and voltage phase angle for IEEE 14 bus test system interacting with BC. The distributed method has succeeded to reach the centralized values in both cases.

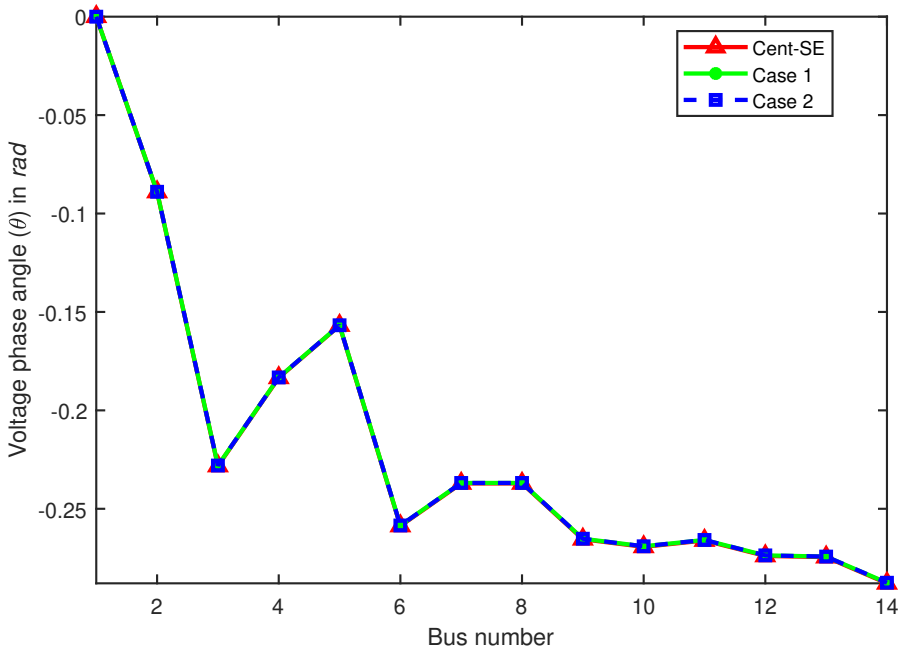

Figure 6-8: IEEE 14 bus system voltage phase angle for centralized (Cent-SE) and distributed state estimation interacting with blockchain (case 1 and case 2)

Fig. 6-9 shows the distributed method objective value during the IEEE 14 bus system's optimization procedure. As proposed in [49], we have considered the state variables convergence rate as convergence criterion. It means that the difference between obtained state variables of two successive iterations are measured at each area and if the value is below the specified threshold (it has been set to 10^{-6} [49]), the optimization stops. It is clear that in case 2, where there is a delay in data transmission, the number of optimization iteration increases.

The numerical results of the comparison between centralized SE and distributed SE are presented in table 6.2. The iteration number and objective value of both centralized and distributed are presented. The objective value for centralized SE is obtained using (3.2) and applying Newton's method. However, for distributed SE, after solving the optimization problem stated in (6.1) for all areas, we gathered the state variables and placed these state variables into (3.2). The objective values are obtained by substituting the distributed SE state variables into (3.2).

As shown in the table, the factual error between these values is approximately 1 percent. The necessity of considering objective value is due to the fact that one of the methods to specify measurement anomalies, so-called *bad data*, is to compare

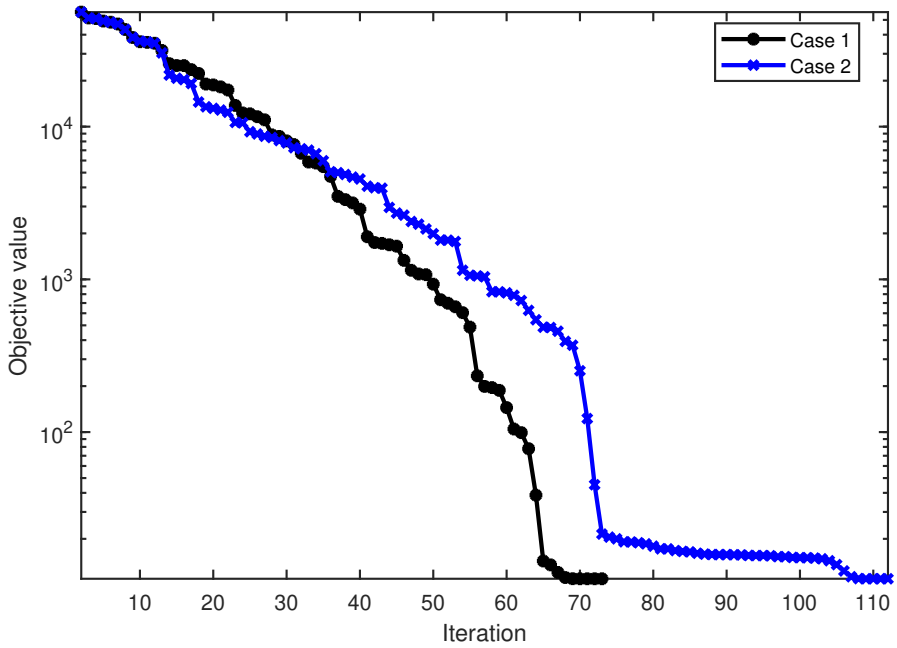

Figure 6-9: Distributed method objective value during iteration for case 1 and case 2

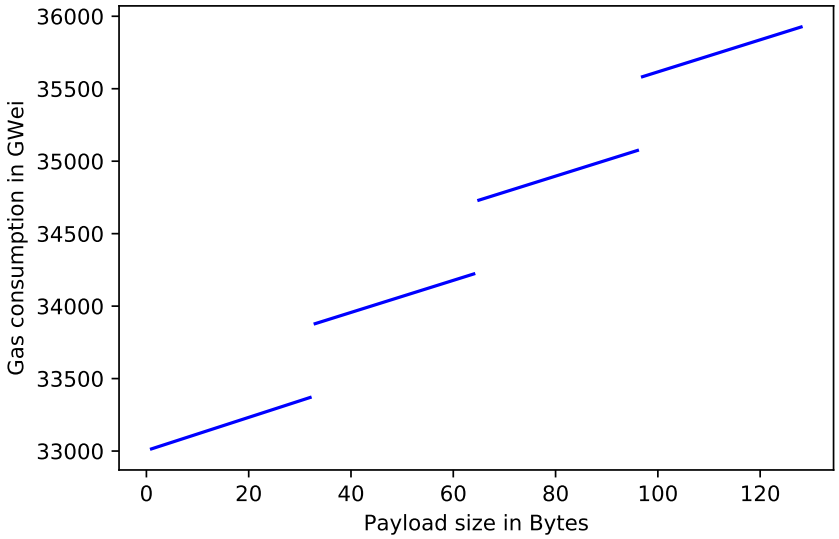

Figure 6-10: Gas consumption in Gwei to transfer bytes with payload size

Table 6.2: Numerical results of comparing centralized and distributed method for case 1 and case 2; centralized SE (CSE); distributed SE (DSE)

	Iteration		**Objective**		**Objective**
	CSE	DSE	CSE	DSE	**error**
case 1	6	73	11.5568	11.6801	1.0559 %
case 2	6	112	11.5568	11.6804	1.0582 %

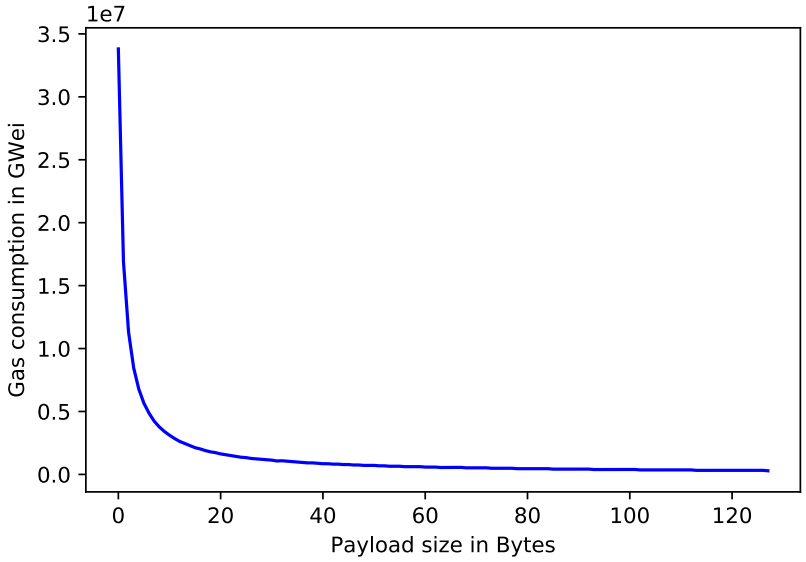

Figure 6-11: Gas consumption to transfer 1024 bytes per payload size in Bytes

objective value with the chi-square value [14]. So, considering measurement residuals distributed method matches the centralized to a great extent as well.

Fig. 6-10 shows the result of the experiment to check the gas consumption, amount of gas used to execute a transaction, with respect to the transaction payload size in bytes of different values in a transaction, i.e., in a hexadecimal value and used to check how it will influence the processing time. Different value precision results in different payload sizes. We executed 128 transactions of payload size one and bytes of size k from 1 to 128. In EVM, a one-word is a maximum of 32 bytes. Zero bytes pad each payload up to the closest factor of 32 bytes and processed as a sequence of 32 bytes words. Most of the operation consumption goes to cryptographic signature checks by the nodes. Gas consumption varies with different byte sizes, and we can see a significant shift for each consecutive 32 bytes, but within each set, the gas fees increased linearly with an increment of a byte.

Fig. 6-11 indicates the optimization of the transfer procedure where several transactions can be concatenated as one string, i.e., bulk data transfer. This would result in less number of transaction to transfer the same amount data without spending extra gas for each execution. For the experiment, we measured the gas consumption to transfer 1024 bytes per 2^{k-1} bytes where $k \in \{1, \ldots, 8\}$, with increase on payload size, the gas consumption reduces for computation at nodes.

6.3 Summary

This chapter was devoted to the practical applications of our proposed methods and tools, showcasing their transformative potential and the broader impact they could have on the research community. We presented not only the concrete applications of our work but also the opportunities for researchers to explore, adapt, and build upon our findings. We invited readers to contribute to the advancement of our proposed model and its applications.

The chapter focused on two key topics: anomaly detection, classification, and identification, and the application of blockchain in distributed power system state estimation. We demonstrated how our methods could enhance anomaly detection and identification capabilities in power systems, contributing to their reliability and resilience. Additionally, we explored the potential of blockchain technology to address data security and integrity challenges in distributed SE, paving the way for more robust and secure power system operations.

"If you optimize everything, you will always be unhappy."

Donald Knuth

Chapter 7

Conclusion

In the pursuit of advancing the topic of power system state estimation, this thesis has embarked on a comprehensive exploration of the intricate challenges inherent in contemporary power grids. The findings that were presented in this research work, have the potential for improving the reliability and resiliency of the power system operation. The outcome of the thesis can be concluded in two categories, i.e., power system anomalies and distributed state estimation.

Power system anomalies

Initially, a novel solution was presented to detect and classify anomalies such as BD, SLC, and FDIA, as well as to identify their origin. Anomalies that bypass the χ^2-test are successfully detected using an anomaly detection index. After that, a ML algorithm is applied to classify anomalies and identify their origin. Based on the obtained results, the proposed algorithm is capable of accurate detection and classification of the anomalies.

It has been demonstrated that utilizing the features associated only with the buses eliminates the need for retraining the ML algorithm once the network topology changes. Furthermore, the application of an optimal feature selection method alleviates the optimization complexity of the ML algorithm. Besides saving time and computational resources, these aspects make the system operator capable of fast response in case an anomaly occurs.

Distributed state estimation

A modified convergence criterion has been presented for distributed SE applications considering features such as iteration number, convergence rate, and needed data to be transmitted between areas. After that, an optimal partitioning method that maintains the security of the system while decreasing the number of auxiliary variables of the distributed SE problem was introduced.

Based on the obtained results, the application of the modified convergence criterion will decrease the number of iterations to a high extent. Additionally, the proposed partitioning method is effective in case of decreasing the number of auxiliary variables of the distributed SE problem and consequently helps to reach an optimal point closer to centralized state estimation.

Additionally, in the context of distributed computation, blockchain technology has attracted research and industrial communities' attention due to its diverse and novel characteristics. Needless to say, the future power grids, so-called smart grids, can benefit from these features in different industrial divisions. In this regard, we tried to point out blockchain application in smart grids' main sector, i.e., the state estimator.

In this work, we have proposed a combination of distributed state estimation and a blockchain designed communication platform for secure data transmission and increasing the system's reliability. Application of the smart contract concept would lead to improving the security of the overall system. Moreover, the robustness of the method against the data transmission latency has been analysed.

Future research direction

Detection and classification of different types of anomalies in case of their simultaneous occurrence along with identification of their origin and designing the suitable countermeasures against them, can be considered as future research directions in the area of power system anomalies.

As mentioned before, we introduced a scheme for the combination of state estimation with blockchain in a distributed transmission system. Therefore, implementing such a combination for the distribution system, in which the applications of renew-

able energy sources are increasing exponentially, can be a future direction. Another research direction for the future can be introducing multi-signature that will make this architecture more secure. Additionally, economical analysis for blockchain's implementation in the power system would be of interest to research and the industrial community and can be considered as another future direction, in the area of distributed state estimation.

Bibliography

[1] R. Christie, "Power systems test case archive. 14 bus power flow test case, 1993," *University of Washington, Department of Electrical Engineering,[Online] Available at https://labs.ece.uw.edu/pstca/pf14/pg_tca14bus.htm.*

[2] R. Christie, "Power systems test case archive. 118 bus power flow test case, 1993," *University of Washington, Department of Electrical Engineering,[Online] Available at https://labs.ece.uw.edu/pstca/pf118/pg_tca118bus.htm.*

[3] A. Minot and N. Li, "A fully distributed state estimation using matrix splitting methods," in *2015 American Control Conference (ACC)*. IEEE, 2015, pp. 2488–2493.

[4] S. Asefi, M. Mitrovic, D. Ćetenović, V. Levi, E. Gryazina, and V. Terzija, "Anomaly detection and classification in power system state estimation: Combining model-based and data-driven methods," *Sustainable Energy, Grids and Networks*, vol. 35, p. 101116, 2023.

[5] N. Živković and A. T. Sarić, "Detection of false data injection attacks using unscented kalman filter," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 5, pp. 847–859, 2018.

[6] A. Minot, Y. M. Lu, and N. Li, "A distributed gauss-newton method for power system state estimation," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3804–3815, 2015.

[7] A. Gómez-Expósito, A. de la Villa Jaén, C. Gómez-Quiles, P. Rousseaux, and T. Van Cutsem, "A taxonomy of multi-area state estimation methods," *Electric Power Systems Research*, vol. 81, no. 4, pp. 1060–1069, 2011.

[8] D. Marelli, B. Ninness, and M. Fu, "Distributed weighted least-squares estimation for power networks," *IFAC-PapersOnLine*, vol. 48, no. 28, pp. 562–567, 2015.

[9] A. J. Conejo, S. de la Torre, and M. Canas, "An optimization approach to multiarea state estimation," *IEEE Transactions on Power Systems*, vol. 22, no. 1, pp. 213–221, 2007.

[10] M. N. Kurt, Y. Yılmaz, and X. Wang, "Secure distributed dynamic state estimation in wide-area smart grids," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 800–815, 2019.

[11] A. Abur and A. G. Exposito, *Power system state estimation: theory and implementation*. CRC press, 2004.

[12] A. S. Musleh, G. Yao, and S. Muyeen, "Blockchain applications in smart grid–review and frameworks," *IEEE Access*, vol. 7, pp. 86 746–86 757, 2019.

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

[14] A. Gomez-Exposito, A. J. Conejo, and C. Canizares, *Electric energy systems: analysis and operation*. CRC press, 2018.

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

[16] F. F. Wu, "Power system state estimation: a survey," *International Journal of Electrical Power & Energy Systems*, vol. 12, no. 2, pp. 80–87, 1990.

[17] M. Ayiad, H. Leite, and H. Martins, "State estimation for hybrid vsc based hvdc/ac transmission networks," *Energies*, vol. 13, no. 18, p. 4932, 2020.

[18] M. Pau, F. Ponci, A. Monti, S. Sulis, C. Muscas, and P. A. Pegoraro, "An efficient and accurate solution for distribution system state estimation with multiarea architecture," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 5, pp. 910–919, 2017.

[19] C. Xu and A. Abur, "Robust linear state estimation for large multi-area power grids," in *2016 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, 2016, pp. 1–5.

[20] T. Zhang, P. Yuan, Y. Du, W. Zhang, and J. Chen, "Robust distributed state estimation of active distribution networks considering communication failures," *International Journal of Electrical Power & Energy Systems*, vol. 118, p. 105732, 2020.

[21] M. Rostami and S. Lotfifard, "Distributed dynamic state estimation of power systems," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 8, pp. 3395–3404, 2017.

[22] V. Kekatos and G. B. Giannakis, "Distributed robust power system state estimation," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1617–1626, 2012.

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

[24] D. U. Case, "Analysis of the cyber attack on the ukrainian power grid," *Electricity Information Sharing and Analysis Center (E-ISAC)*, vol. 388, 2016.

[25] K. Nishiya, J. Hasegawa, and T. Koike, "Dynamic state estimation including anomaly detection and identification for power systems," in *IEE proceedings C (generation, transmission and distribution)*, vol. 129, no. 5. IET, 1982, pp. 192–198.

[26] G. Valverde and V. Terzija, "Unscented kalman filter for power system dynamic state estimation," *IET generation, transmission & distribution*, vol. 5, no. 1, pp. 29–37, 2010.

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

[28] F. Nejabatkhah, Y. W. Li, H. Liang, and R. Reza Ahrabi, "Cyber-security of smart microgrids: A survey," *Energies*, vol. 14, no. 1, p. 27, 2021.

[29] H. T. Reda, A. Anwar, A. N. Mahmood, and Z. Tari, "A taxonomy of cyber defence strategies against false data attacks in smart grid," *arXiv preprint arXiv:2103.16085*, 2021.

[30] D. Mukherjee, "A novel strategy for locational detection of false data injection attack," *Sustainable Energy, Grids and Networks*, vol. 31, p. 100702, 2022.

[31] D. Hock, M. Kappes, and B. Ghita, "Using multiple data sources to detect manipulated electricity meter by an entropy-inspired metric," *Sustainable Energy, Grids and Networks*, vol. 21, p. 100290, 2020.

[32] M. M. Ayiad, H. Leite, and H. Martins, "State estimation for hybrid vsc based hvdc/ac: Unified bad data detection integrated with gaussian mixture model," *IEEE Access*, vol. 9, pp. 91 730–91 740, 2021.

[33] H. Zang, M. Geng, M. Xue, X. Mao, M. Huang, S. Chen, Z. Wei, and G. Sun, "A robust state estimator for integrated electrical and heating networks," *IEEE Access*, vol. 7, pp. 109 990–110 001, 2019.

[34] C. H. Ho, H. Wu, S. Chan, and Y. Hou, "A robust statistical approach to distributed power system state estimation with bad data," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 517–527, 2019.

[35] J. Zhao and L. Mili, "Robust unscented kalman filter for power system dynamic state estimation with unknown noise statistics," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 1215–1224, 2017.

[36] X. Wang and E. E. Yaz, "Second-order fault tolerant extended kalman filter for discrete time nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 64, no. 12, pp. 5086–5093, 2019.

[37] A. L. Da Silva, M. Do Coutto Filho, and J. De Queiroz, "State forecasting in electric power systems," in *IEE Proceedings C (Generation, Transmission and Distribution)*, vol. 130, no. 5. IET, 1983, pp. 237–244.

[38] L. Dang, B. Chen, S. Wang, W. Ma, and P. Ren, "Robust power system state estimation with minimum error entropy unscented kalman filter," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 11, pp. 8797–8808, 2020.

[39] J. A. Massignan, J. B. London, and V. Miranda, "Tracking power system state evolution with maximum-correntropy-based extended kalman filter," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 4, pp. 616–626, 2020.

[40] W. Ma, J. Qiu, X. Liu, G. Xiao, J. Duan, and B. Chen, "Unscented kalman filter with generalized correntropy loss for robust power system forecasting-aided state estimation," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 11, pp. 6091–6100, 2019.

[41] J. Xie, I. Alvarez-Fernandez, and W. Sun, "A review of machine learning applications in power system resilience," in *2020 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2020, pp. 1–5.

[42] A. Sayghe, J. Zhao, and C. Konstantinou, "Evasion attacks with adversarial deep learning against power system state estimation," in *2020 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2020, pp. 1–5.

[43] K. R. Mestav and L. Tong, "Learning the unobservable: High-resolution state estimation via deep learning," in *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2019, pp. 171–176.

[44] B. Liu, H. Wu, Y. Zhang, R. Yang, and A. Bernstein, "Robust matrix completion state estimation in distribution systems," in *2019 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2019, pp. 1–5.

[45] K. R. Mestav, J. Luengo-Rozas, and L. Tong, "Bayesian state estimation for unobservable distribution systems via deep learning," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4910–4920, 2019.

[46] K. Nagaraj, S. Zou, C. Ruben, S. Dhulipala, A. Starke, A. Bretas, A. Zare, and J. McNair, "Ensemble corrdet with adaptive statistics for bad data detection," *IET Smart Grid*, vol. 3, no. 5, pp. 572–580, 2020.

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

[48] M. Ashrafuzzaman, Y. Chakhchoukh, A. A. Jillepalli, P. T. Tosic, D. C. de Leon, F. T. Sheldon, and B. K. Johnson, "Detecting stealthy false data

injection attacks in power grids using deep learning," in *2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC)*. IEEE, 2018, pp. 219–225.

[49] S. Asefi, S. Parsegov, and E. Gryazina, "Distributed state estimation: a novel stopping criterion," *arXiv preprint arXiv:2012.00647*, 2020.

[50] G. N. Korres, "A distributed multiarea state estimation," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 73–84, 2010.

[51] W. Jiang, V. Vittal, and G. T. Heydt, "Diakoptic state estimation using phasor measurement units," *IEEE Transactions on Power Systems*, vol. 23, no. 4, pp. 1580–1589, 2008.

[52] L. Zhao and A. Abur, "Multi area state estimation using synchronized phasor measurements," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 611–617, 2005.

[53] S. Boyd, N. Parikh, and E. Chu, *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011.

[54] X. Li and A. Scaglione, "Robust decentralized state estimation and tracking for power systems via network gossiping," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 7, pp. 1184–1194, 2013.

[55] L. Xie, D.-H. Choi, S. Kar, and H. V. Poor, "Fully distributed state estimation for wide-area monitoring systems," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1154–1169, 2012.

[56] A. Sharma, S. Srivastava, and S. Chakrabarti, "Multi area state estimation using area slack bus angle adjustment with minimal data exchange," in *2013 IEEE Power & Energy Society General Meeting*. IEEE, 2013, pp. 1–5.

[57] Y. Guo, L. Tong, W. Wu, H. Sun, and B. Zhang, "Hierarchical multi-area state estimation via sensitivity function exchanges," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 442–453, 2016.

[58] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," *www.bitcoin.org*, pp. 1–9, 2008. [Online]. Available: https://bitcoin.org/bitcoin.pdf

[59] P. Vigna and M. J. Casey, *The age of cryptocurrency: how bitcoin and digital money are challenging the global economic order*. St. Martin's Press, 2015.

[60] I. Eyal, "Blockchain Technology: Transforming Libertarian Cryptocurrency Dreams to Finance and Banking Realities," *Computer*, vol. 50, no. 9, pp. 38–49, 2017. [Online]. Available: http://ieeexplore.ieee.org/document/8048646/

[61] G. W. Peters and E. Panayi, "Understanding Modern Banking Ledgers Through Blockchain Technologies: Future of Transaction Processing and Smart Contracts on the Internet of Money," in *Banking Beyond Banks and Money*. Springer, Cham, 2016, pp. 239–278.

[62] Y. Madhwal and P. Panfilov, "Blockchain And Supply Chain Management: Aircrafts' Parts' Business Case," in *Annals of DAAAM and Proceedings of the International DAAAM Symposium*, 2017, pp. 1051–1056.

[63] S. Malik, V. Dedeoglu, S. S. Kanhere, and R. Jurdak, "TrustChain: Trust Management in Blockchain and IoT Supported Supply Chains," in *2019 IEEE International Conference on Blockchain (Blockchain)*. IEEE, 7 2019, pp. 184–193. [Online]. Available: https://ieeexplore.ieee.org/document/8946187/

[64] N. Alzahrani and N. Bulusu, "Block-Supply Chain: A New Anti-Counterfeiting Supply Chain Using NFC and Blockchain," in *Proceedings of the 1st Workshop on Cryptocurrencies and Blockchains for Distributed Systems - CryBlock'18*. New York, New York, USA: ACM Press, 2018, pp. 30–35.

[65] D. Korepanova, S. Kruglik, Y. Madhwal, T. Myaldzin, I. Prokhorov, I. Shiyanov, S. Vorobyov, and Y. Yanovich, "Blockchain-Based Solution to Prevent Postage Stamps Fraud," in *2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*. IEEE, 5 2019, pp. 171–175. [Online]. Available: https://ieeexplore.ieee.org/document/8751495/

[66] P. Mamoshina, L. Ojomoko, Y. Yanovich, A. Ostrovski, A. Botezatu, P. Prikhodko, E. Izumchenko, A. Aliper, K. Romantsov, A. Zhebrak, I. O. Ogu, and A. Zhavoronkov, "Converging blockchain and next-generation artificial intelligence technologies to decentralize and accelerate biomedical research and healthcare," *Oncotarget*, vol. 9, no. 5, pp. 5665–5690, 1 2018. [Online]. Available: http://www.oncotarget.com/fulltext/22345

[67] M. Saad, J. Spaulding, L. Njilla, C. Kamhoua, S. Shetty, D. Nyang, and A. Mohaisen, "Exploring the Attack Surface of Blockchain: A Systematic Overview," 4 2019. [Online]. Available: http://arxiv.org/abs/1904.03487

[68] M. Choe, "LONDONCOIN: THE ULTIMATE CRYPTOCURRENCY." [Online]. Available: https://coinmarketcap.com/

[69] Buterin and Vitalik, "Ethereum White Paper: A Next Generation Smart Contract & Decentralized Application Platform," *Etherum*, no. January, pp. 1–36, 2014. [Online]. Available: https://github.com/ethereum/wiki/wiki/White-Paper

[70] "Truffle Suite - Your Ethereum Swiss Army Knife," 2018. [Online]. Available: http://truffleframework.com/

[71] J. Qiu, D. Grace, G. Ding, J. Yao, and Q. Wu, "Blockchain-based secure spectrum trading for unmanned-aerial-vehicle-assisted cellular networks: An

operator's perspective," *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 451–466, 2019.

[72] N. Z. Aitzhan and D. Svetinovic, "Security and privacy in decentralized energy trading through multi-signatures, blockchain and anonymous messaging streams," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 5, pp. 840–852, 2016.

[73] M. Pipattanasomporn, M. Kuzlu, and S. Rahman, "A blockchain-based platform for exchange of solar energy: Laboratory-scale implementation," in *2018 International Conference and Utility Exhibition on Green Energy for Sustainable Development (ICUE)*. IEEE, 2018, pp. 1–9.

[74] G. Liang, S. R. Weller, F. Luo, J. Zhao, and Z. Y. Dong, "Distributed blockchain-based data protection framework for modern power systems against cyber attacks," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3162–3173, 2018.

[75] Z. Dong, F. Luo, and G. Liang, "Blockchain: a secure, decentralized, trusted cyber infrastructure solution for future energy systems," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 5, pp. 958–967, 2018.

[76] Q. Su, H. Wang, C. Sun, B. Li, and J. Li, "Cyber-attacks against cyber-physical power systems security: State estimation, attacks reconstruction and defense strategy," *Applied Mathematics and Computation*, vol. 413, p. 126639, 2022.

[77] A. Presekal, A. Ştefanov, V. S. Rajkumar, and P. Palensky, "Attack graph model for cyber-physical power systems using hybrid deep learning," *IEEE Transactions on Smart Grid*, 2023.

[78] A. Dairi, F. Harrou, B. Bouyeddou, S.-M. Senouci, and Y. Sun, "Semi-supervised deep learning-driven anomaly detection schemes for cyber-attack detection in smart grids," in *Power Systems Cybersecurity: Methods, Concepts, and Best Practices*. Springer, 2023, pp. 265–295.

[79] T. Chen, F. Liu, P. Li, L. Sun, and G. A. Amaratunga, "A distributed multi-area power system state estimation method based on generalized loss function," *Measurement Science and Technology*, vol. 34, no. 11, p. 115010, 2023.

[80] F. Mohammadi and M. Saif, "Blockchain technology in modern power systems: a systematic review," *IEEE Systems, Man, and Cybernetics Magazine*, vol. 9, no. 1, pp. 37–47, 2023.

[81] J. Zhao, A. Gómez-Expósito, M. Netto, L. Mili, A. Abur, V. Terzija, I. Kamwa, B. Pal, A. K. Singh, J. Qi *et al.*, "Power system dynamic state estimation: Motivations, definitions, methodologies, and future work," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 3188–3198, 2019.

[82] M. Göl and A. Abur, "A modified chi-squares test for improved bad data detection," in *2015 IEEE Eindhoven PowerTech*. IEEE, 2015, pp. 1–5.

[83] J. Zhao, G. Zhang, Z. Y. Dong, and K. P. Wong, "Forecasting-aided imperfect false data injection attacks against power system nonlinear state estimation," *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 6–8, 2015.

[84] P. Frasca, H. Ishii, C. Ravazzi, and R. Tempo, "Distributed randomized algorithms for opinion formation, centrality computation and power systems estimation: A tutorial overview," *European journal of control*, vol. 24, pp. 2–13, 2015.

[85] V. Kekatos and G. B. Giannakis, "Distributed robust power system state estimation," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1617–1626, 2013.

[86] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[87] D. N. Ćetenović and A. M. Ranković, "Optimal parameterization of kalman filter based three-phase dynamic state estimator for active distribution networks," *International Journal of Electrical Power & Energy Systems*, vol. 101, pp. 472–481, 2018.

[88] V. Basetti, A. K. Chandel, and R. Chandel, "Power system dynamic state estimation using prediction based evolutionary technique," *Energy*, vol. 107, pp. 29–47, 2016.

[89] Z. Jin, S. Chakrabarti, J. Yu, L. Ding, and V. Terzija, "An improved algorithm for cubature kalman filter based forecasting-aided state estimation and anomaly detection," *International Transactions on Electrical Energy Systems*, vol. 31, no. 5, p. e12714, 2021.

[90] P. Del Moral, "Nonlinear filtering: Interacting particle resolution," *Comptes Rendus de l'Académie des Sciences-Series I-Mathematics*, vol. 325, no. 6, pp. 653–658, 1997.

[91] N. Bretas, "An iterative dynamic state estimation and bad data processing," *International Journal of Electrical Power & Energy Systems*, vol. 11, no. 1, pp. 70–74, 1989.

[92] P. L. Houtekamer and H. L. Mitchell, "Data assimilation using an ensemble kalman filter technique," *Monthly Weather Review*, vol. 126, no. 3, pp. 796–811, 1998.

[93] A. Gelb, J. Kasper Jr, R. Nash Jr, C. Price, and A. Sutherland Jr, "Applied optimal estimation, 374 pp," 1974.

[94] V. Basetti, A. K. Chandel, and C. K. Shiva, "Square-root cubature kalman filter based power system dynamic state estimation," *Sustainable Energy, Grids and Networks*, vol. 31, p. 100712, 2022.

[95] J. H. Friedman, "Stochastic gradient boosting," *Computational statistics & data analysis*, vol. 38, no. 4, pp. 367–378, 2002.

[96] D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein, and M. Klein, *Logistic regression*. Springer, 2002.

[97] R. Short and K. Fukunaga, "The optimal distance measure for nearest neighbor classification," *IEEE transactions on Information Theory*, vol. 27, no. 5, pp. 622–627, 1981.

[98] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[99] J. R. Quinlan, "Induction of decision trees," *Machine learning*, vol. 1, no. 1, pp. 81–106, 1986.

[100] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[101] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[102] J. Opitz and S. Burst, "Macro f1 and macro f1," *arXiv preprint arXiv:1911.03347*, 2019.

[103] D. Tran, H. Mac, V. Tong, H. A. Tran, and L. G. Nguyen, "A lstm based framework for handling multiclass imbalance in dga botnet detection," *Neurocomputing*, vol. 275, pp. 2401–2413, 2018.

[104] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Journal of bioinformatics and computational biology*, vol. 3, no. 02, pp. 185–205, 2005.

[105] Z. Zhao, R. Anand, and M. Wang, "Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform," in *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 2019, pp. 442–452.

[106] R. Smith, "A mutual information approach to calculating nonlinearity," *Stat*, vol. 4, no. 1, pp. 291–303, 2015.

[107] T. D. Gauthier, "Detecting trends using spearman's rank correlation coefficient," *Environmental forensics*, vol. 2, no. 4, pp. 359–362, 2001.

[108] K. Hassine, A. Erbad, and R. Hamila, "Important complexity reduction of random forest in multi-classification problem," in *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*. IEEE, 2019, pp. 226–231.

[109] J. Yan, F. Guo, and C. Wen, "False data injection against state estimation in power systems with multiple cooperative attackers," *ISA transactions*, vol. 101, pp. 225–233, 2020.

[110] V. Kekatos, G. Wang, H. Zhu, and G. B. Giannakis, "Psse redux: Convex relaxation, decentralized, robust, and dynamic approaches," *arXiv preprint arXiv:1708.03981*, 2017.

[111] D. Du, X. Li, W. Li, R. Chen, M. Fei, and L. Wu, "Admm-based distributed state estimation of smart grid under data deception and denial of service attacks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2019.

[112] M. Glavic and T. Van Cutsem, "Tracking network state from combined scada and synchronized phasor measurements," in *2013 IREP Symposium Bulk Power System Dynamics and Control-IX Optimization, Security and Control of the Emerging Power Grid*. IEEE, 2013, pp. 1–10.

[113] A. Minot, Y. M. Lu, and N. Li, "A distributed gauss-newton method for power system state estimation," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3804–3815, 2016.

[114] R. D. Zimmerman, C. E. Murillo-Sanchez, "Matpower." [Online]. Available: https://matpower.org

[115] S. Asefi, M. Mitrovic, D. Ćetenović, V. Levi, E. Gryazina, and V. Terzija, "Power system anomaly detection and classification utilizing WLS-EKF state estimation and machine learning," *arXiv preprint arXiv:2209.12629*, 2022.

[116] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux, "API design for machine learning software: experiences from the scikit-learn project," in *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, 2013, pp. 108–122.

[117] A. Abur, "Power education toolbox (p.e.t): An interactive software package for state estimation," in *2009 IEEE Power & Energy Society General Meeting*, 2009, pp. 1–4.

[118] S. Haber and W. S. Stornetta, "How to time-stamp a digital document," *Journal of Cryptology*, vol. 3, no. 2, pp. 99–111, 1 1991. [Online]. Available: https://link.springer.com/article/10.1007/BF00196791

[119] P. Jones and D. Eastlake, "US Secure Hash Algorithm 1 (SHA1)," September 2001.

[120] S. Gueron, S. Johnson, and J. Walker, "SHA-512/256 ," 2011.

[121] "NODES AND CLIENTS." [Online]. Available: https://ethereum.org/en/developers/docs/nodes-and-clients/

[122] "Infura: The foundation for decentralized applications." [Online]. Available: https://infura.io/

[123] "Alchemy." [Online]. Available: https://www.alchemyapi.io/

[124] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," *Ethereum project yellow paper*, vol. 151, pp. 1–32, 2014.

[125] Q. ShenTu and J. Yu, "A Blind-Mixing Scheme for Bitcoin based on an Elliptic Curve Cryptography Blind Digital Signature Algorithm," *arxiv*, 2015.

[126] C. Chen, Y. Chen, K. Zhang, M. Ni, S. Wang, and R. Liang, "System redundancy enhancement of secondary frequency control under latency attacks," *IEEE Transactions on Smart Grid*, 2020.

[127] "Ganache: ONE CLICK BLOCKCHAIN." [Online]. Available: https://www.trufflesuite.com/ganache

[128] Y. Madhwal, "code repository," 2020. [Online]. Available: https://github.com/yashmadhwal/secureDataTransmission

[129] U. Maurer, "Modelling a public-key infrastructure," pp. 325–350, 1996.

[130] D. Simon, *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons, 2006.

Appendix A

Additional material

I. Statistical properties of normalized measurement residuals

In the proposed methodology, χ^2-test is performed over the measurement residuals obtained by WLS estimator in order to detect bad data presence. This is feasible if normalized measurement residuals follow Standard Gaussian distribution.

WLS estimator utilizes measurement model only. Instead of (4.45), consider the linear measurement model first (time index t is omitted to simplify the notation):

$$\boldsymbol{z} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{e} \tag{A.1}$$

In this case, estimated state can be obtained directly as:

$$\widehat{\boldsymbol{x}} = \left[\boldsymbol{H}^T \boldsymbol{R}^{-1} \boldsymbol{H}\right]^{-1} \boldsymbol{H}^T \boldsymbol{R}^{-1} \boldsymbol{z} \tag{A.2}$$

Now, measurement residuals can be expressed as follows:

$$\begin{aligned}
\boldsymbol{r} &= \boldsymbol{z} - \boldsymbol{H}\widehat{\boldsymbol{x}} \\
&= \boldsymbol{z} - \boldsymbol{H}\left[\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\right]^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{z} \\
&= \left[\boldsymbol{I} - \boldsymbol{H}\left[\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\right]^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1}\right]\boldsymbol{z} \\
&= \left[\boldsymbol{I} - \boldsymbol{H}\left[\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\right]^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1}\right]\left[\boldsymbol{H}\boldsymbol{x} + \boldsymbol{e}\right] \\
&= \boldsymbol{H}\boldsymbol{x} + \boldsymbol{e} - \boldsymbol{H}\left[\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\right]^{-1}\left[\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\right]\boldsymbol{x} - \\
&\quad \boldsymbol{H}\left[\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\right]^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{e} \\
&= \boldsymbol{H}\boldsymbol{x} + \boldsymbol{e} - \boldsymbol{H}\boldsymbol{x} - \boldsymbol{H}\left[\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\right]^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{e} \\
&= \boldsymbol{e} - \boldsymbol{H}\left[\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\right]^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{e} \\
&= \left[\boldsymbol{I} - \boldsymbol{H}\left[\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\right]^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1}\right]\boldsymbol{e} \\
&= \boldsymbol{S}\boldsymbol{e}
\end{aligned} \tag{A.3}$$

where $\boldsymbol{S} = \boldsymbol{I} - \boldsymbol{H}\left[\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\right]^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1}$ is residual sensitivity matrix. Thus, the relationship between measurement residuals $\boldsymbol{r}$ and measurement noise $\boldsymbol{e}$ is linear and determined by the residual sensitivity matrix $\boldsymbol{S}$. Measurement noise $\boldsymbol{e}$ is random variable assumed to be Gaussian distributed with zero mean ($E[\boldsymbol{e}] = \boldsymbol{0}$) and covariance matrix $\boldsymbol{R}$ ($E[\boldsymbol{e}\boldsymbol{e}^T] = \boldsymbol{R}$). When a Gaussian random variable undergoes a linear transformation, the result will be new random variable that is also Gaussian distributed; proof for this can be found in [130]. Considering the relationship between measurement residuals and measurement noise is linear, and the probability distribution of measurement noise is Gaussian, measurement residuals will also follow Gaussian distribution with mean and covariance matrix obtained as:

$$E[\boldsymbol{r}] = E\left[\boldsymbol{S}\boldsymbol{e}\right] = \boldsymbol{S}E[\boldsymbol{e}] = \boldsymbol{0} \tag{A.4}$$

$$
\begin{aligned}
E\left[\boldsymbol{r}\boldsymbol{r}^T\right] &= E\left[\boldsymbol{S}\boldsymbol{e}\left[\boldsymbol{S}\boldsymbol{e}\right]^T\right] \\
&= E\left[\boldsymbol{S}\boldsymbol{e}\boldsymbol{e}^T\boldsymbol{S}^T\right] \\
&= \boldsymbol{S}E\left[\boldsymbol{e}\boldsymbol{e}^T\right]\boldsymbol{S}^T \\
&= \boldsymbol{S}\boldsymbol{R}\boldsymbol{S}^T \\
&= \boldsymbol{S}\boldsymbol{R} \\
&= \boldsymbol{R} - \boldsymbol{H}\left[\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\right]^{-1}\boldsymbol{H}^T \\
&= \boldsymbol{\Omega}
\end{aligned}
\tag{A.5}
$$

If measurement residuals are normalized with square root of the corresponding diagonal entries of residual covariance matrix $\boldsymbol{\Omega}$, the normalized measurement residuals defined by (3.7) will have Standard Gaussian distribution (mean is zero vector and covariance matrix is identity matrix).

Since measurement model (4.45) is nonlinear, the relationship between measurement residuals and measurement noise is also nonlinear. Thus, the true probability distribution of normalized measurement residuals is hard to find analytically. However, model (4.45) can be linearized yielding [11]:

$$\Delta\boldsymbol{z} = \boldsymbol{H}\Delta\boldsymbol{x} + \boldsymbol{e} \tag{A.6}$$

where $\boldsymbol{H} = \frac{\partial \boldsymbol{h}(\boldsymbol{x})}{\partial \boldsymbol{x}}$ is calculated at the point of linearization. In this case, estimate of the linearized state will be:

$$\Delta\widehat{\boldsymbol{x}} = \left[\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\right]^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1}\Delta\boldsymbol{z} \tag{A.7}$$

and estimated value of $\Delta\boldsymbol{z}$ is:

$$\Delta\widehat{\boldsymbol{z}} = \boldsymbol{H}\Delta\widehat{\boldsymbol{x}} \tag{A.8}$$

Now, measurement residuals can be expressed as:

$$\boldsymbol{r} = \Delta\boldsymbol{z} - \Delta\widehat{\boldsymbol{z}} \tag{A.9}$$

Combining (A.6) - (A.8) into (A.9) lead to the same conclusion as before:

$$\boldsymbol{r} = \left[\boldsymbol{I} - \boldsymbol{H}\left[\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\right]^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1}\right]\boldsymbol{e} = \boldsymbol{S}\boldsymbol{e} \tag{A.10}$$

In this case, since original measurement function $\boldsymbol{h}(\boldsymbol{x})$ is approximated via Taylor series with higher order terms neglected, WLS approximates the true probability distribution of normalized measurement residuals with Gaussian distribution. The accuracy of this approximation depends on the level of nonlinearity in (4.45) and the level of measurement noise. Less nonlinearity and more accurate measurements will lead to more accurate approximation. Considering that the measurement set is composed of telemetered SCADA measurements with low noise levels, it can be expected that the approximation is accurate enough. The approximate probability distribution of normalized measurement residuals will not be exactly Standard Gaussian, but closely. To demonstrate this, samples of normalized measurement residuals are collected at four different time instants during simulation of normal operating conditions. In Fig. A.1, for each instant, the approximate probability density function of normalized measurement residuals is plotted against the probability density function of the Standard Gaussian distribution. As can be seen, the approximate distribution of the data is very close to Standard Gaussian distribution.

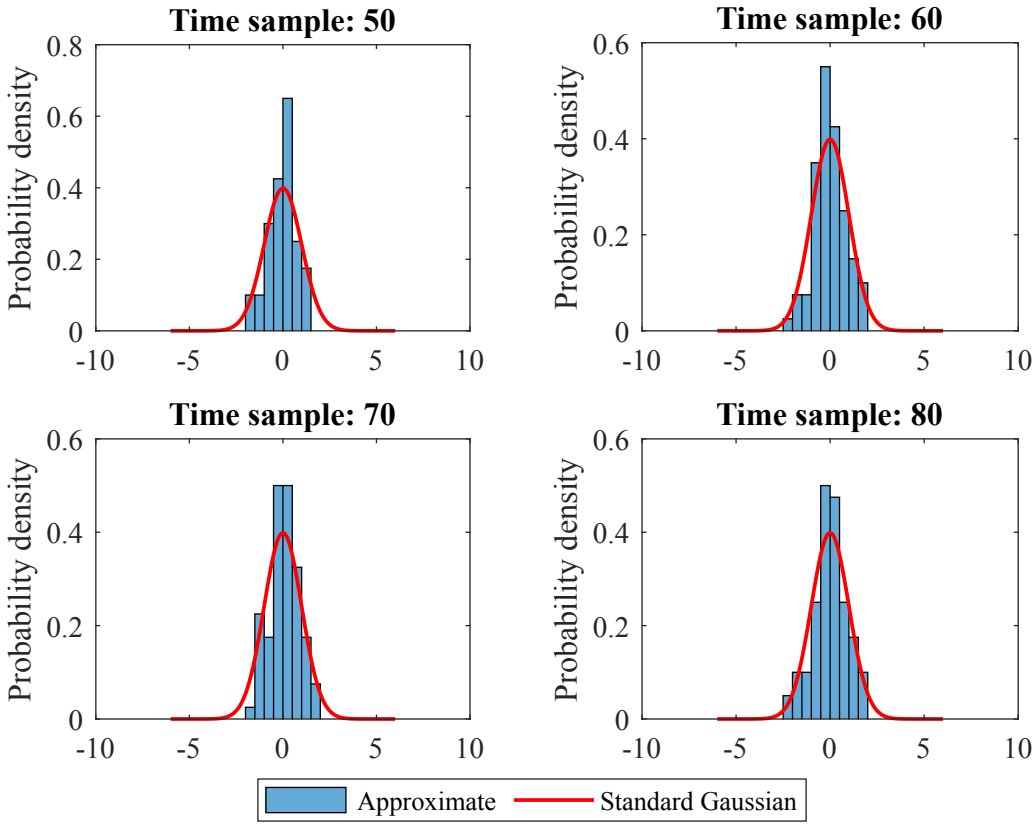

Figure A.1: Approximate probability density function of normalized measurement residuals (blue histogram) against Standard Gaussian probability density function (red curve) for four time instants during the normal operation.