

University of Nevada, Reno

**Analysis of Privacy-aware Data Sharing in Cyber-physical Energy
Systems**

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in
Computer Science and Engineering

by

Md Tamjid Hossain

Dr. Hung (Jim) La/Thesis Advisor
May 2022



THE GRADUATE SCHOOL

We recommend that the thesis
prepared under our supervision by

Md Tamjid Hossain

entitled

**Analysis of Privacy-aware Data Sharing in Cyber-physical
Energy Systems**

be accepted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

Hung (Jim) La, Ph.D.
Advisor

Shahriar Badsha, Ph.D.
Co-advisor

Mohammed Ben-Idris, Ph.D.
Graduate School Representative

David W. Zeh, Ph.D., Dean
Graduate School

May, 2022

Abstract

In this thesis, we determine the key factors and correlations among the privacy, security, and utility requirements of grid networks to ensure effective inter-and intra-actions within physical layer equipment (e.g., distributed energy resources (DERs), intelligent electronic devices (IEDs), etc.). We have conducted a comprehensive analysis of the existing consensus mechanisms in blockchain-enabled smart grids while pointing out the potential research gaps. We develop a practical and effective consensus mechanism for a private and permissioned blockchain-enabled Supervisory control and data acquisition (SCADA) system. Moreover, we bridge a common and popular industrial control system (ICS) protocol, distributed network protocol 3 (DNP3) with the blockchain network to ensure smooth operation. In addition, we develop differential privacy (DP)-enabled strategies to achieve data security, privacy, and utility requirements of the power system network under an adversarial setting. Specifically, we aim to analyze and develop a provable correlation between privacy loss and other DP parameters considering the variations of attacks and their impacts along with DP constraints. This will enable modern power grid designers to develop, design, and employ DP-based fault-tolerant models in data-driven power grid operation and control. Furthermore, we conduct feasibility and quality-of-service (QoS) analysis of the DP mechanism and the grid to achieve certified robustness. Feasibility analysis of the privacy measure provides an assessment of the practicability of differential privacy in grid operation and warns the operators about the possible failures and incoming attacks on physical layer operations. QoS is analyzed in the power grid in terms of data accuracy, computational overhead, and resource utilization.

Acknowledgments

I would like to thank my advisor, Dr. Hung (Jim) La, who funded my research and allowed me to pursue my passion for the research on data privacy and security. His guidance and recommendations have been instrumental throughout my degree. I would also like to thank Dr. Shahriar Badsha wholeheartedly for his support and guidance from the beginning of my journey here. Moreover, I am grateful to him for providing funding support throughout the major part of my academic years at the University of Nevada, Reno. Without Dr. La and Dr. Badsha, I would not have been able to pursue a master's degree, or have learned so much about research and teaching. I am also thankful to my other supervisor, Dr. Haoting Shen to provide me the counseling and support during his tenure at the University of Nevada, Reno. I would also like to thank my committee member Dr. Mohammed Ben-Idris for his advice and for taking the time to review the contents of my thesis. Finally, I would like to thank the people in my personal life, particularly my parents, my spouse and my friends for being as supportive as they are and helping me when I needed it most. I am fortunate to have the company of such wonderful people in my life, my success is in part due to the influence and impact of these people. Thank you!

Table of Contents

1	Introduction	1
1.1	Focused Privacy-aware Data Sharing Techniques	2
1.1.1	Blockchain	2
1.1.2	Differential Privacy (DP)	3
1.2	Motivations	4
1.3	Literature Review and Research Gap	5
1.3.1	Blockchain-enabled Data Sharing Frameworks for SCADA	6
1.3.2	DP-assisted Privacy Preservation	7
1.3.3	Adversarial Classification of the DP mechanism	8
1.4	Contributions	9
1.5	Thesis Organization	11
2	Blockchain-enabled SCADA systems for Power Grids	12
2.1	Introduction	12
2.2	Communication Protocol Design	14
2.2.1	Request Packet	15
2.2.2	Response Packet	16
2.3	Component Description of SCADA systems	16
2.3.1	Relay Servers (or Smart Meters)	16
2.3.2	Data Aggregator (DA)	17

2.3.3	Control Center (CC)	17
2.4	Proposed Practical Blockchain Framework	18
2.4.1	Hashing Mechanism	18
2.4.2	Developing Consensus Mechanism	20
2.5	Performance Evaluation	26
2.5.1	Experimental Setup	26
2.5.2	Experimental Procedure	26
2.5.3	Experimental Result Analysis	31
2.6	Summary	35
3	Privacy, Security, and Utility Analysis of Differentially Private CPES	37
3.1	Introduction	37
3.1.1	Adversarial Exploitation of Differential Privacy	38
3.1.2	Impacts of DP integration	38
3.1.3	Contributions	39
3.2	Problem Statement and Threat Model	40
3.2.1	Problem Statement	40
3.2.2	Threat Model	43
3.3	Optimal Attack and Defense Strategy	45
3.3.1	Attacker's Objectives	45
3.3.2	Defender's Objectives	45
3.3.3	Optimal Attack Strategy	46
3.3.4	Parameter Design for Effective Defense Mechanism	47
3.3.5	Modelling the Criterion for Feasibility and QoS Analysis	50
3.4	Experimental Analysis	51
3.4.1	False Data injection Attack Simulation in a Synchronphasor Network of a Power Grid	52

3.4.2	Privacy, Security, and Utility Analysis	53
3.5	Summary	59
4	Conclusion and Future Work	61
4.1	Conclusion	61
4.2	Published Research Articles	62
4.3	Future Work	63

List of Tables

2.1	Measurement data mapping	28
2.2	Network topology mapping	29
3.1	List of symbols and their description	40

List of Figures

1.1	(a) Energy Systems as a subset of cyber-physical systems (b) Cyber-physical energy system (CPES) is fueling other industries and smart cities	1
1.2	Practical implementation of blockchain technology in industries (image source: https://7wdata.be/digital-transformation/what-companies-are-using-blockchain-technology/)	3
1.3	Concept and implementation of Differential Privacy (DP)	4
2.1	Application Layer Structure of the DNP3m Protocol	15
2.2	Example blockchain.	19
2.3	Traditional mining node selection procedure	22
2.4	Proposed mining node selection process	23
2.5	Operational process flowchart of the proposed scheme. The process follows the number sequence of the flowchart.	25
2.6	IEEE-9 bus system topology	27
2.7	Network topology mapping	29
2.8	Number of total nodes vs computational cost (in ms) of the sub-processes	33
2.9	Comparison of our scheme with (a) PPEA-BPC [1], (b) PPM-HDA [2] and (c) DG-APED [3]	34

3.1	Basic operation of differential privacy mechanism. Laplacian noise is added to the query to make the query result ε -indistinguishable. For each query, the differentially private query results are different.	41
3.2	Differential privacy in an adversarial setting. Differentially private query results are manipulated by a malicious actor and sent to the analyst. A small amount of manipulation is difficult to detect	44
3.3	Hourly average user consumption (KWh) recorded by a PMU. Small level of privacy loss (ε) leads to higher deviation. The DP-noise has been drawn from a Laplace distribution with zero mean and $2b^2$ variance ($\theta = 0, \sigma^2 = 2b^2$)	53
3.4	Correlation between the attack impact (μ_a) and the DP-parameters. Attack impact (μ_a) decreases when- (i) privacy loss (ε) increases, (ii) attackers' tolerance (γ) decreases, and (iii) data sensitivity (Δf) decreases.	54
3.5	Utility cost analysis of DP method for varying privacy loss (ε). Privacy cost is higher than the security cost. Moreover, the overall cost is small when epsilon is large (i.e., $\varepsilon = 0.9$)	57
3.6	QoS analysis of a DP-based system under attack scenario. (a) 'DP-predicted' vary from the 'original-predicted' due to privacy cost (b) The attack impact on the prediction accuracy is negligible for a short-time FDI attack (circled in the graph) (c) The 'DP-FDI-predicted' values are very close to the 'DP-predicted' values as the security cost is low.	58

Funding Support

This work is partially supported by the U.S. National Science Foundation (NSF) under grants NSF-CAREER: 1846513 and NSF-PFI-TT: 1919127. The views, opinions, findings, and conclusions reflected in this publication are solely those of the authors and do not represent the official policy or position of the NSF.

Chapter 1

Introduction

Cyber and physical worlds are overlapping at a tremendous rate in this modern era as we continue to look for new ways to control the composition of the physical world. The industries and organizations that are utilizing this mesh network of cyber and physical components include but are not limited to transportation, manufacturing, health, smart factories, and energy systems. Among these, the energy sector (particularly, power grids) is prominent as it enables all the other industries and organizations. Fig. 1.1(a) demonstrates how the energy system is a subset of cyber-physical systems (CPSs) as a whole and Fig. 1.1(b) illustrates how it is enabling other industries.

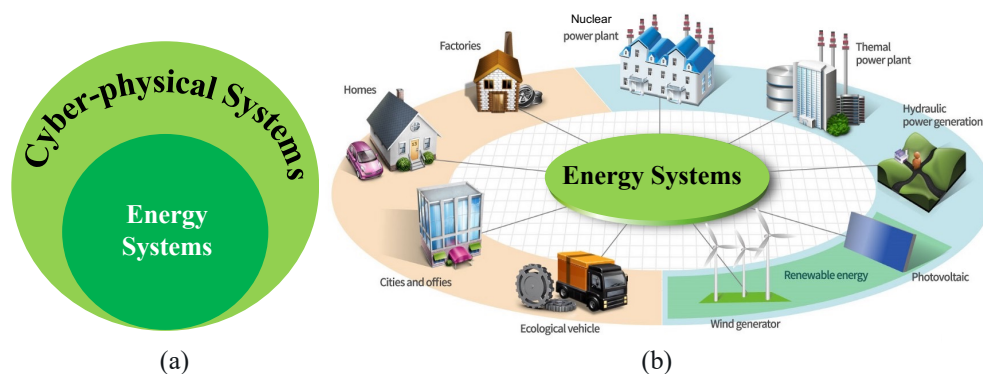


Figure 1.1: (a) Energy Systems as a subset of cyber-physical systems (b) Cyber-physical energy system (CPES) is fueling other industries and smart cities

However, for the successful functioning of the cyber-physical energy systems (CPES), they need to be self-optimized, self-cognitive, and self-customized. And for that, a continuous, secure, and automated flow of data is essential. At the same time, the data acquisition processes of the energy systems need to be trustworthy and faster without the presence of a third party to protect the client’s confidential information and prevent unauthorized data manipulation. Particularly, as the power grids get frequent attention from the cyber attackers (e.g., Dragonfly [ICS-ALERT-14-176-02A], Stuxnet [4]), they need additional protection from data privacy and security attacks. Vulnerable data sharing and data acquisition process in the power grid operations may bring catastrophic consequences.

Over the years, a significant number of research including data-driven privacy-preservation methods [5–7], blockchain-based data sharing [1, 8–10] and anomaly detection techniques [11–13] have been conducted to ensure the data privacy and security of power grid. Nevertheless, these techniques are still vulnerable to adversarial manipulation [14–16], and thus, need proper adversarial analysis. In this thesis, we focus on two data integrity and privacy-preservation techniques, blockchain [17] and differential privacy [18] for privacy-aware data sharing in the context of the CPES domain. Moreover, we analyze the potential vulnerabilities of these methods and subsequently, propose the countermeasures.

1.1 Focused Privacy-aware Data Sharing Techniques

1.1.1 Blockchain

Blockchain is an open, distributed, and immutable ledger system that was first conceptualized as a public transaction ledger of the cryptocurrency bitcoin by Satoshi

Nakamoto in 2008 [17]. Since then, blockchain technology has been proposed to apply in several domains, including but not limited to, task allocation [19], ride-sharing [20], cybersecurity information sharing [21], cyber insurance [22], etc. (Fig. 1.2).



Figure 1.2: Practical implementation of blockchain technology in industries (image source: <https://7wdata.be/digital-transformation/what-companies-are-using-blockchain-technology/>)

1.1.2 Differential Privacy (DP)

Differential privacy (DP) is an emerging data privacy-preservation technique that hides an individual contribution through a randomized noise addition mechanism following some well-known statistical distributions (e.g., Laplace, Gaussian, Exponential, etc.) and privacy conditions. It was first introduced by a group of Microsoft researchers back in 2006 and named ϵ -differential privacy [18]. Since then, it has

gained a lot of research attention from the privacy and security research community.

The main idea behind the DP mechanism lies in the indistinguishable nature of the query results on databases that differ by a single entry or row. More specifically, the mechanism ensures that no additional information about a participating client/node/entry can be obtained by analyzing the query results (Fig. 1.3(a)). Hence, DP essentially provides a provable data privacy guarantee. Since data privacy can be quantified through DP mechanism, several industries including Apple [23], Google [24], Uber [25], Microsoft [26] have already implemented it in their data-driven applications (Fig. 1.3(b)).

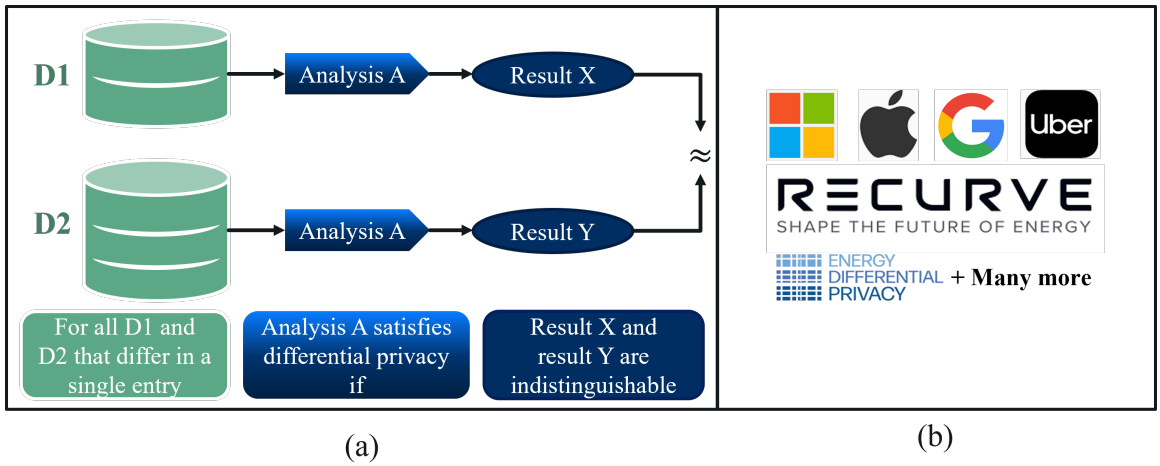


Figure 1.3: Concept and implementation of Differential Privacy (DP)

1.2 Motivations

A common threat to the privacy of the grid data is inferring valuable and sensitive information about the clients and industrial operations. For example, multiple aggregated queries on the smart meters dataset with other available peripheral information (e.g., client's location, number of smart meters, etc.) can reveal sensitive information (e.g., energy consumption pattern, absence time of the tenants from

home, etc.) which in turn, can facilitate malicious manipulation of the consumption, and even, house theft [27]. Similar types of attacks are also possible on a masked and anonymized dataset with some additional information [28, 29]. Thus, it is important to preserve the privacy of sensitive information, which includes but is not limited to the consumer’s electricity usage data [30], utility power generation data [31], electrical appliance location data, etc. [32]. To tackle the threat against malicious activity in power grid, existing privacy preservation schemes [33–36] include data encryption [37], blockchain-based data sharing [38], differential privacy [39], and k-anonymity [37].

However, the major drawbacks are (1) encryption-based methods: high resource utilization, less resiliency, attack-prone nature, cryptographic key generation and distribution complexity, latency, etc. [40, 41], (2) anonymization-based techniques: may reveal true identity [28]. On the other hand, blockchain and differential privacy do not have such drawbacks. Moreover, blockchain is particularly effective in data integrity preservation, verification, and nodal authentication whereas differential privacy provides a provable privacy guarantee with a minimal computational cost. These motivate us to choose blockchain and differential privacy over data anonymization and encryption techniques for privacy-aware data sharing analysis in this thesis.

1.3 Literature Review and Research Gap

In this part, we discuss some notable research on blockchain and differential privacy in the context of privacy-aware data sharing in the CPES domain. We also point out some potential research gaps in existing frameworks and methods.

1.3.1 Blockchain-enabled Data Sharing Frameworks for SCADA

Supervisory control and data acquisition (SCADA) is a system of software and hardware elements that enables industrial organizations (in our case, power grids) to control industrial operations locally or at remote locations. It monitors, collects, and processes real-time data through direct interaction with devices such as sensors, valves, pumps, motors, and more using human-machine interface (HMI) software. This data acquisition process needs to be trustworthy and faster without the presence of a third party to prevent data privacy leakage and unauthorized data manipulation.

Blockchain technology [17] can pave the way for such secured and privacy-preserved data acquisition in SCADA systems. Moreover, due to some inherent properties (e.g., immutability), blockchain can preserve data integrity. Numerous research has been conducted to apply blockchain frameworks in SCADA systems [8, 9, 42, 43]. For example, In [8], blockchain has been applied to the industrial control systems (ICS) network of device nodes where every device node executes the smart contract and records the transaction data. Any device node on the ICS network can record all transaction data. *However, it may then be possible for an adversary to compromise a device node and manipulate the data. So, there has to be a novel and decentralized way of tackling cyber-attacks before they take place and cause a single-point-of-failure and data manipulation.*

Taking this into account, [9] proposes to enhance the integrity and confidentiality of synchrophasor communication networks in an ICS environment through blockchain and essential cryptographic tools (e.g. bloom filter, ECDSA, etc.). Similarly, [1] proposes a customized mining node selection procedure that follows the process of selecting a node as a mining node if its energy consumption data is the closest to the average. *However, in a practical scenario, the electricity consumption behavior of any user may not change rapidly and remain the same at least for a while. Thus,*

in a particular time slot, there can be some nodes (e.g., smart meters) that hold the same (or at least similar) data as previous time slots. Now, if any of those nodes have data closest to the average, then it would remain as a mining node for concurrent time slots. An adversary, knowing this confidential information, can compromise that node and exploit these concurrent time slots to design and execute malicious attacks.

Likewise, [44] addresses the problem of reaching consensus in distributed social networks. *However, they have not incorporated the mining technique, rather they've used wired connected RSUs (roadside units). Their proposed consensus mechanism is lightweight but not suitable enough to apply in such SCADA systems that require a near-real-time operation through wireless connectivity.* From an urgent need of developing a practical consensus protocol, PoAh (Proof-of-Authentication) has been proposed in [45]. It introduces the trusted third party in the validation of blocks and solves several problems of previous consensus mechanisms. However, this and the other discussed schemes have not bridged the operational technology (OT) protocols (e.g., Modbus, RS-232, Profinet, Profibus, DNP3, etc.) with the communication network of the blockchain which in turn, requires further research in this direction.

1.3.2 DP-assisted Privacy Preservation

Unlike other privacy-preserving techniques, DP preserves data privacy by perturbing a small amount of noise in the query result while making sure not to overly degrade data utility [41, 46–56]. Over the years, DP has been proposed as a solution to passive attacks (e.g., eavesdropping) in various CPSs including smart grids [46, 48]. However, the adversarial exploitation of DP mechanisms has not been noticed until recently, *so little focus has been given to explore in the following aspects of DP: security breach due to active attacks, utility degradation due to large noise, resource exhaustion due to defense strategies, in the smart grid domain. In this context, we focus on the*

adversarial analysis of active attacks (particularly, false data injection (FDI) attacks) in DP-based power grids, which facilitates the development of novel and effective defense strategies.

Another related line of research has focused on the impacts and defenses of active attacks on state estimation, future consumption prediction, billing, pricing, etc. in smart grids [49, 57–60]. Most of this work has analyzed scenarios where false data has been used to modify the original results. *However, these attacks have not been formulated in a power grid network that uses DP as a privacy-preserving tool, which we aim to do in this research.*

1.3.3 Adversarial Classification of the DP mechanism

Recent work [61–63] considered active attacks (particularly, false data injection attacks) in DP-based CPSs (e.g., smart grids, transportation systems, etc.). More specifically, the authors discussed and successively solved the optimal false data injection attacks (FDIA) that degrade the anomaly detection capabilities of the system while allowing the attacker to remain undetected by “hiding” false data into the DP noise. However, these works mostly considered defense mechanisms based on anomaly detection schemes for the post-attack phase. Thus, these existing methods propose to sanitize the manipulated outcomes after the attack instead of taking steps to prevent the attacks beforehand, which could open a window for the attacker to conduct real-time attacks. *In contrast, we aim to analyze the correlation of the parameters affecting DP under adversarial settings; then, leveraging the correlation, we aim to facilitate the deployment of the desired level of privacy, utility, and security among the nodes in each layer of the power system network. Following the adversarial analysis, the attack surface can be minimized, i.e., the incentive of the attacker can be reduced, which in turn reduces attack motivations while assisting attack prevention.* In short,

we aim to develop a defense strategy as a part of the design process (pre-attack phase) to prohibit FDI attacks.

In addition, our method will offer feasibility and quality-of-service (QoS) analysis of DP-based systems in an adversarial setting and will enable smart grid operators to achieve certified robustness of DP-based power system networks through adversarial analysis, effective defense strategy development, and QoS analysis, which has been difficult to attain [64,65]. DP can be a tool for the attacker to conduct active attacks in the smart grid domain, so it is necessary to investigate the feasibility of the DP mechanism in power grid operations. At the same time, it is also essential to carry out a QoS analysis of a DP-based smart grid in adversarial settings to determine the most suitable applications for use of DP as a secured privacy-preserving tool, which makes our proposed work advanced compared to the state of the art.

1.4 Contributions

The **main goal of this research** is to determine the key factors and correlations among the privacy, security, and utility requirements of grid networks to ensure effective inter-and intra-actions within physical layer equipment (e.g., distributed energy resources (DERs), intelligent electronic devices (IEDs), etc.). To this end, our major contributions to this research are:

- **Conducting a comprehensive analysis on the existing consensus mechanisms and mining node selection processes** in blockchain-enabled smart grids while pointing out the potential research gap.
- **Develop a practical and effective mining node selection algorithm to assist the consensus process of the nodes** in a private and permissioned blockchain-enabled SCADA system.

- **Bridging** the common and popular ICS (Industrial Control System) protocols, **DNP3 (Distributed Network Protocol 3) with the blockchain network** to ensure smooth operation.
- **Developing differential privacy (DP)-enabled data sharing strategy to achieve data security, privacy, and utility requirements** of the power system network under adversarial settings. Specifically, we aim to analyze and develop a provable correlation between privacy loss, data sensitivity, and other DP parameters considering the variations of attacks and their impacts along with DP constraints. This will enable power grid designers to develop, design and employ DP-based fault-tolerant models in data-driven power grid operation and control.
- **Performing adversarial analysis of the differential privacy** in cyber-physical energy systems. Particularly, we aim to explore the malicious manipulation opportunities through the exploitation of differential noise and develop countermeasures against such manipulation.
- **Performing feasibility and QoS analysis of the DP mechanism and the grid** to achieve certified robustness. The analysis is performed in terms of data accuracy (cost), and computational overhead (latency). Both feasibility and QoS analysis of the differential privacy provides an assessment of the practicability of the mechanism in grid operation and warns the operators about the possible failures and incoming attacks on physical layer operations.

1.5 Thesis Organization

The organization of the remainder of this thesis is as follows. Chapter 2 details the comparative analysis of the existing practical methods to ensure data integrity and privacy along with their limitations in the context of blockchain-enabled cyber-physical energy systems. In addition, it discusses an effective consensus protocol for the blockchain-assisted data acquisition process. Chapter 3 covers the feasibility and quality-of-service requirements and their evaluation for the integration of privacy measures in energy systems. In addition, it highlights how one of the focused privacy measures, differential privacy, can be exploited to conduct a data integrity attack. In chapter 4, we summarize our research work with some potential research impacts and future research scopes. We also enlist some of our published research articles in this direction.

Chapter 2

Blockchain-enabled SCADA systems for Power Grids

2.1 Introduction

The SCADA systems need to be privacy-protected, secured, and fast for the trustworthy operation of the power grids. Blockchain technology can assist to fulfil these objectives due to its inherent properties such as distributed, immutable, non-repudiation, etc. records transactions between two parties efficiently and in a verifiable and permanent way [66]. For these certain characteristics and potentiality, blockchain technology has gained a lot of attention to be implemented in the data acquisition process of the future SCADA systems in power grid industries. This can be a giant leap toward the fourth industrial revolution. The motivations behind the blockchain-enabled SCADA systems are as follows:

- The data acquisition process needs to be secured, automated, quick, error-free, and robust for any SCADA system in the power grid domain.

- Blockchain technology can help the data acquisition process to be a smart fit in the context of modern power grids (smart grids) and Industry 4.0 (I4.0).
- Traditional consensus mechanisms oftentimes incorporate reward-penalty for the nodes which can be proven unnecessary. Moreover, this would add a computation burden to the data acquisition process. Hence, a practical, random, and fair consensus mechanism is essential for effective blockchain-enabled SCADA systems.

From the perspective of a blockchain-enabled data acquisition system, the consensus mechanism can be considered one of the most crucial parts as it deals with nodal authentication and data verification. Hence, we aim to develop a realistic consensus protocol through a customized mining node selection process. Additionally, we discuss other essential steps of a private blockchain creation including, but not limited to, hashing mechanism, block creation, verification, and addition.

Our proposed **Proof-of-Random-Count-in-Hashes** (PoRCH) mechanism counts a random number that appears in the hash value of the measurement data and selects the node/server as a mining node that has the largest count. In case of not having a single node bearing the largest count of random number appearance, the scheme utilizes a cryptographically secure random node selection algorithm for mining purposes from all available nodes. In this way, the mechanism ensures randomness and fairness in the mining node selection process of a private blockchain network. The performance evaluation shows that the entire process requires very low computational overhead while preserving data security, privacy, and trust. In a nutshell, our major contributions in this chapter are –

- Highlighting the major components of the blockchain framework and conducting a comprehensive analysis of the existing process.

- Analyzing the compatibility of the **DNP3** protocol and bridging it with the blockchain framework.
- Developing and presenting a novel consensus mechanism, PoRCH, and with a simplified demo. PoRCH incorporates a customized mining node selection procedure for a private and permissioned blockchain-enabled data acquisition system where incentive or penalty is not required for the validators/miners. Thus, it ensures a low computational burden over the network.

Throughout this chapter, ‘nodes’, ‘edge servers’, ‘field devices’, ‘relay servers’, and ‘smart meters’ are used alternately depending upon the context.

2.2 Communication Protocol Design

The communication system of an ICS can be broadly categorized into information technology (IT) and operational technology (OT). IT systems use ERP systems that mostly use SMTP, SNMP, OPC, SMB, HTTP, and XML protocols whereas OT systems encompass MES, DCS, SCADA, RTUs, PLCs, sensors, etc. and use MODBUS, PROFIBUS, PROFINET, DNP3, Ethernet/IP, RS-232, etc. In particular, we focus on **DNP3m** (DNP3 minus, a simplified version of DNP3) protocol. The protocol is used while sharing the data among the control center, data aggregator, and smart meters. It is an application layer protocol built on top of the TCP protocol and widely used in the US power grid infrastructure [67]. The application layer structure is depicted in Fig. 2.1.

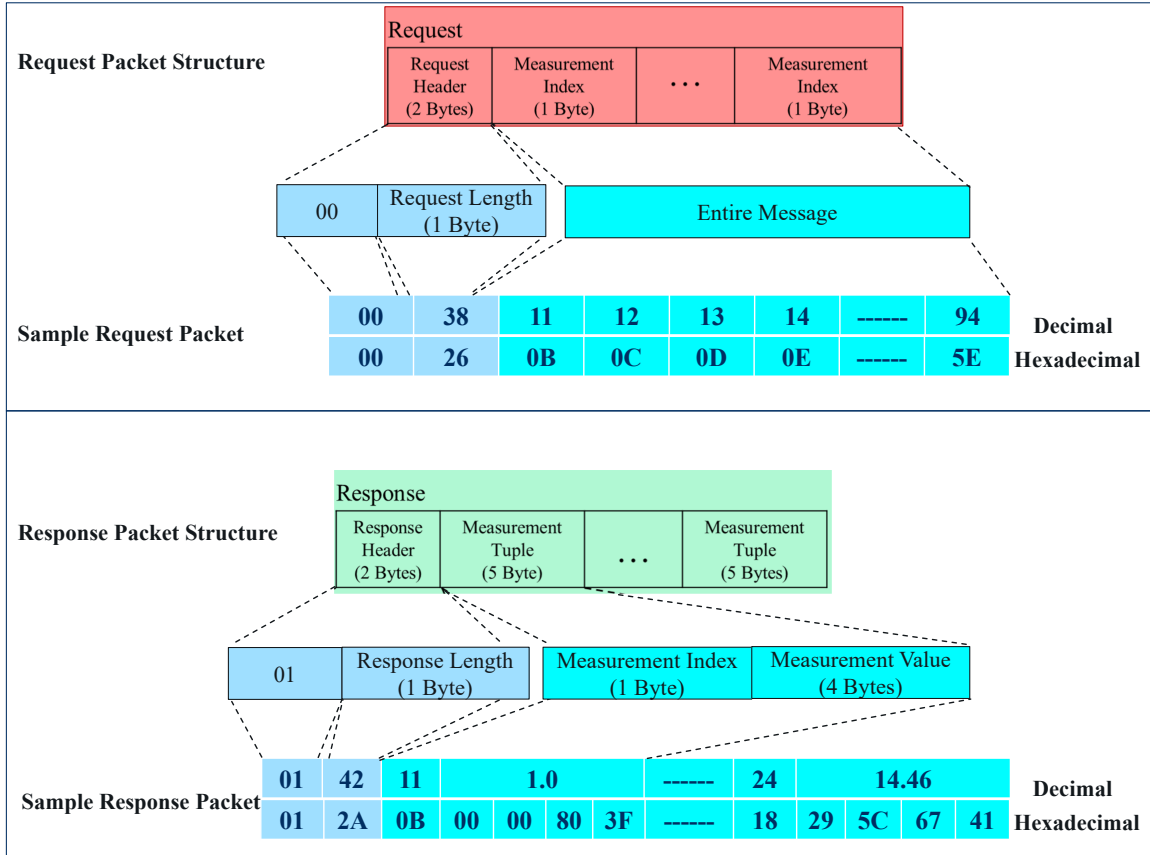


Figure 2.1: Application Layer Structure of the DNP3m Protocol

2.2.1 Request Packet

The “Request Header” contains two bytes. The first byte indicates that this is the request of the DNP3m protocol; the value of this byte should be set as “00”. The second byte of the “Request Header” contains the length of this whole request message. Following the “Request Header,” the request contains requested data indices or messages. For example, if the control center wants to collect the value of voltage magnitude (indexed as 11), voltage phasor (indexed as 12), real power injection (indexed as 13), reactive power injection (indexed as 14) from Bus-1 of an IEEE 9-Bus system in a particular instance, then the request packet should include indices 11, 12, 13 & 14 in the message part. Each index should occupy one byte.

2.2.2 Response Packet

The “Response Header” contains two bytes. The first byte indicates that this is the response of the DNP3m protocol; the value of this byte should be set as “01”. The second byte of the “Response Header” contains the length of this whole response message. Following the “Response Header,” the response contains multiple measurement tuples. Each measurement tuple contains a measurement index occupying one byte and a measurement value occupying 4 bytes. In this tuple, the measurement index corresponds to the index included in the request packet and the measurement value is the corresponding value measured from the IEEE bus systems.

For example, if the control center wants to collect the value of voltage magnitude (indexed as 11), voltage phasor (indexed as 12), real power injection (indexed as 13), reactive power injection (indexed as 14) from Bus-1 of an IEEE 9-Bus system in a particular instance, then the request packet should include indices 11, 12, 13 & 14 in the message part. Each index should occupy one byte.

2.3 Component Description of SCADA systems

2.3.1 Relay Servers (or Smart Meters)

Relay servers are the primary sources of measurement data. They are associated with different field devices, sensors, actuators, phasor measurement units (PMUs), and valves. In our experiment, we use four nodes to represent the relay servers. The relay servers can act both as servers and clients based on operational needs. Here, relay servers use the DNP3m protocol to send the response with measurement data. 8-bit Unicode Transformation Format (UTF-8) has been used as an encoding mechanism for sending string responses. Another important task of the relay servers

is to take part in the voting process of mining node selection. To cast vote for the appropriate node during the mining node selection process, the relay servers first use hashing technique to hash their measurement data. Then, they count the appearance of the random number in their hash values and finally vote for the node to be selected as the mining node that has the largest count.

2.3.2 Data Aggregator (DA)

The data aggregator (DA) plays the most vital role in the proposed data acquisition system. It is responsible for collecting the data, then preprocessing it, coordinating mining node selection, and all kinds of verification processes. DA should preferably reside inside the internal network with the relay servers for efficient operation and low computational overhead and time. In our experiment, based on the operational requirements, DA can act as a server during some periods and a client in other periods. As soon as it gets a data acquisition request from the control center (CC), it starts collecting the data from relay servers and finally completes the cycle by sending an updated blockchain to CC and other relay servers.

Another important task of DA is to generate random numbers to facilitate the mining node selection process. Moreover, during counting the random number's appearance in the hash value, there could be multiple nodes having the same random count in their hash values. Then, although it might occur rarely, the DA delegates one node from all available nodes based on a random selection algorithm.

2.3.3 Control Center (CC)

The control center (CC) is responsible for initializing and controlling the operation. CC can be any physical or cloud server and may have a human-machine interface

(HMI) attached to it for better user experiences. At first, CC sends the data acquisition command to DA and finally receives the updated copy of the blockchain containing measured data at the end of each cycle. For simplicity, in our experiment, we consider each data acquisition cycle as 15 seconds. Nonetheless, this data acquisition cycle is configurable according to the requirement of the data acquisition system, data generation, and processing time.

2.4 Proposed Practical Blockchain Framework

In this part, we describe the proposed practical blockchain framework for future SCADA systems in the power grid domain. We divide the framework into two parts: (a) hashing mechanism, (b) core consensus mechanism. Hashing mechanism performs generating hash operations. Core consensus mechanism covers (b.1) counting random number appearance in hash value, (b.2) mining node selection, (b.3) block creation, (b.4) block verification, and (b.5) block addition. A brief overview of the underlying structure of our proposed blockchain framework is illustrated in Fig. 2.2. The data preprocessing, data aggregation, existing blockchain verification, hashing, new block creation, block verification, and block addition occur as an intermediate process.

2.4.1 Hashing Mechanism

Hashing refers to a process of generating an output value of a fixed or predefined length regardless of the length of the input value. The result of this computational function is called a hash. Normally, a hash is encoded in hexadecimal so the resulting string is easier to work with and debug. In a blockchain, the process of using a specific hash function to process a transaction or data is called hashing. We use the hashing mechanism due to its following properties:

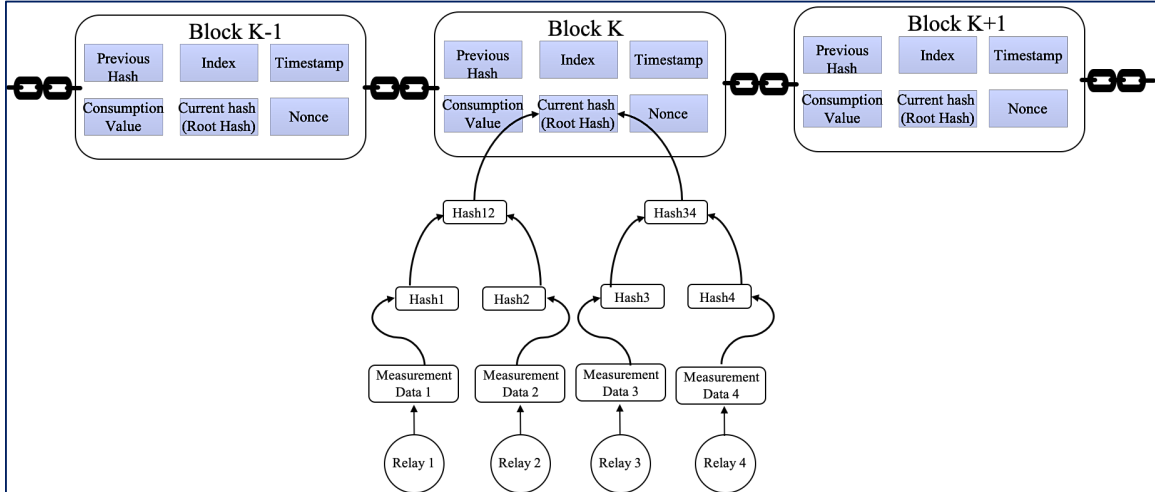


Figure 2.2: Example blockchain.

- **Deterministic.** The same output will be generated for the same input regardless of the number of times one processes that input using a hash function
- **Pre-image resistance** The hashing mechanism works as a one-way function. That means, it is practically impossible to find out the original input from an output hash number. The only possible way to find the original input is to hash all the possible combinations of inputs until hashing the correct input which is impractical and would create a huge amount of computational overhead
- **Second pre-image resistance.** For every different input, even if they differ slightly by only a letter or a digit, there will be different hashes with a predefined length
- **Fast computation.** Hashing mechanism works very fast. While it is impractical and may take years for a normal computer to reverse the hash function to find out the original input from a hash number, it would take the only fraction of a second to get the hash value of a given input of any length
- **Collision resistance.** Rarely, a hash collision may occur as a result of the

‘Birthday Box’ incident [68]. A hash collision is defined as an instance in which two or more observations are hashed to the same value [69].

In our experiment, we use SHA-256 as a hashing algorithm. To implement SHA-256, we utilize Python 3.0 *hashlib* module.

2.4.2 Developing Consensus Mechanism

A consensus mechanism can be described as the process of adding new blocks to the blockchain after testing the validity by validators/miners to ensure trust in the network. Over the years, numerous consensus mechanisms have been proposed, developed, and implemented in several blockchain networks [70]. Both public and private blockchains use a consensus mechanism. Popular consensus mechanisms are Proof of Work (PoW), Proof of Stake (PoS), Delegated Proof of Stake (DPoS), Practical Byzantine Fault Tolerance (PBFT), and Proof of Authority (PoA), RAFT [70], etc. Each one has its advantages and drawbacks. For instance, PoW requires high computational overheads whereas PoS is biased toward the wealthiest validators and thus unfair for new participants. Similarly, DPoS is less decentralized and less resilient and PBFT only works on a permissioned blockchain due to the lack of anonymity [71].

Though [71] proposes a consensus mechanism named Proof of Driving (PoD) that aims to improve PBFT in vehicular ad hoc network applications, they mainly focus on implementing it in a public blockchain network. However, they overlook the exclusion of a reward-penalty mechanism that unnecessarily adds the computational burden over a private blockchain network in a small industry. For example, in a power grid’s SCADA systems, the participating nodes (i.e., smart meters) do not need any incentive mechanism to take part in the mining process since the power grid authority usually owns them and the blockchain network. Adding the incentive mechanisms, in this case, would only add a heavy computational burden in terms of bandwidth,

latency, and storage capacity. Considering that, we design a lightweight consensus mechanism with a novel mining node selection algorithm that does not require the inclusion of any incentive (i.e., reward-penalty) process for the participating nodes.

Similarly, the consensus process developed by [1] possesses a security vulnerability. In their scheme, they select the node as a mining node whose data is closest to the average electricity consumption data. They also mention that there could be multiple pseudonyms (nodes with pseudonymous identities) whose average electricity consumption data have the same distance to the average, which means all of them would be the mining nodes in that time slot. Nevertheless, in a real-life scenario, the energy consumption behavior of any client does not change drastically for any concurrent time slots. For example, when the users turn on a light or fan at their houses, they may not turn it off momentarily; rather they keep their loads (i.e., lights, fans, heaters, etc.) active for at least some concurrent time slots. Therefore, it's a realistic assumption that the overall load recorded by a smart meter at time t_n remains similar for the next t_{n+k} time slots.

A malicious actor can exploit this information by compromising any pseudonym and then spending longer time (at least $t_{n+k} - t_n$ time slots) to craft other attacks (e.g., false data injection, membership inference, etc.), or even hiding the attack identity. Moreover, traditional schemes (e.g., [1]) consider selecting multiple mining nodes in any communication round of the blockchain. It may lead to easier attack (impersonating) opportunities for any adversary. Fig. 2.4 illustrates such a mining node selection process of the traditional consensus models.

Mining node selection:

Our proposed mining node selection process follows algorithm 1. The process is also illustrated as a flowchart in Fig. 2.5. This mining node selection algorithm can be

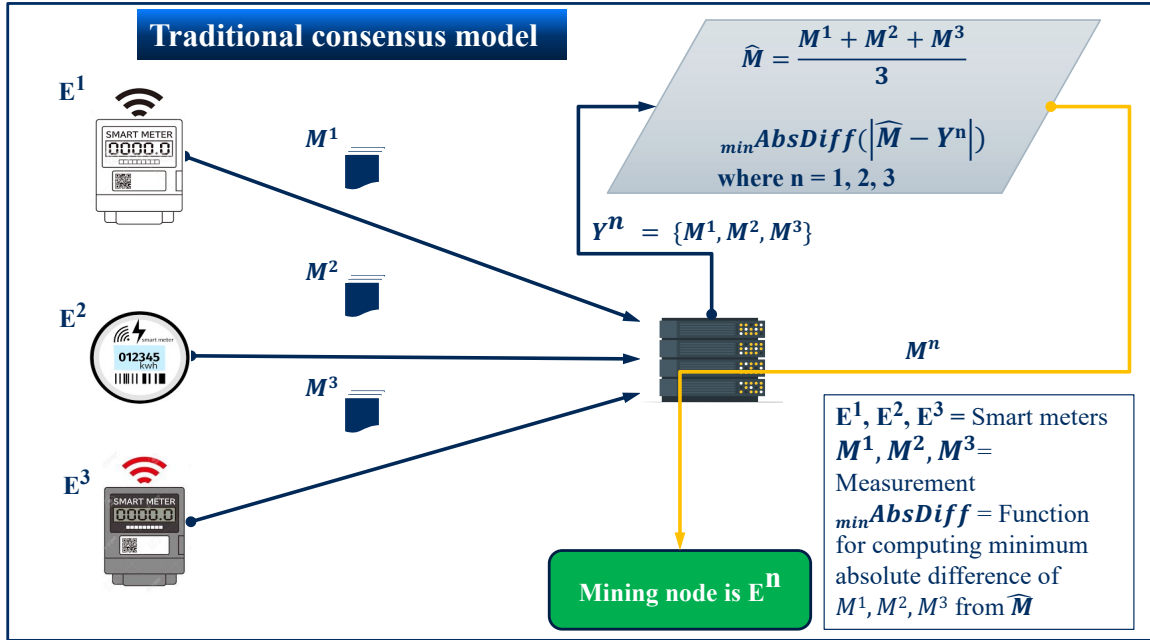


Figure 2.3: Traditional mining node selection procedure

described below.

- First, the DA server requests data from the relay servers (or smart meters).
- Then the relay servers send the encrypted measurement data to DA.
- After getting the measurement data from each relay, DA generates a random number and broadcasts this random number to all the relay servers.
- Relay servers then hash their measurement data and find out that random number appearance count from their respective hash values. For example if the hash value of a particular node in a particular round is f5c75be1e8cafe74cdb3b1a500-89259146cee4334924d6f131a2e2c06829eb39 and the random number is 3, then, the random number appearance count of that particular node is five (5). Then, they share their associated measurement with a random number count to DA.
- Next, the relay servers share the count with each other. Meanwhile, the DA also hashes each measurement and finds out the same counts against each hash

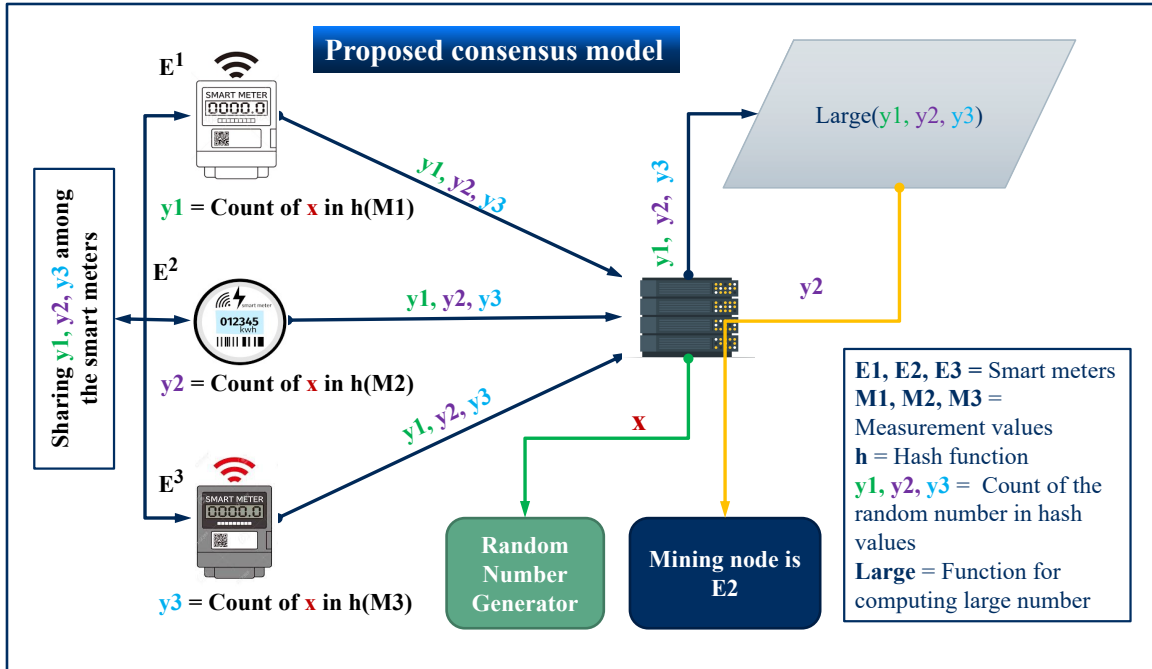


Figure 2.4: Proposed mining node selection process

value. In this stage, every server including the DA has the same counts.

- At the final stage, all the servers cast their votes for the relay server having the largest count of random number appearance. If all the relay servers reach a consensus, then the DA selects the relay having the largest count as the mining node for that cycle. However, if multiple relay servers have the same largest counts or all of them have the count as zero (0), then the DA selects the mining node randomly following a cryptographically secure random selection algorithm. The functions $random.randint(start, end)$ and $random.choice(list)$ of the module *random* from the python standard library has been used to generate random numbers and select random nodes in our simplified demonstration. Nonetheless, for achieving more security and randomness, the proposed scheme can be extended to adopt a cryptographically secure pseudorandom number generator (CSPRNG) algorithm.

Algorithm 1: Mining node selection	
	Input : Measurement data (D_m), Relay (N_r), Random number (R)
	Output: Mining node (N_m)
1	Function <i>Generate_Hash</i> (D_m) :
2	$hash \leftarrow \text{hashlib.sha256}(\text{block}).\text{hexdigest}()$
3	return $hash$;
4	end
5	Function <i>Count</i> ($hash, R$) :
6	$C \leftarrow hash.\text{count}(R)$
7	return C ;
8	end
9	Create a dictionary: $R_v \leftarrow \{N_r : C\}; C \leftarrow \text{Count}(hash, R)$
10	Sort the dictionary, R_v in descending order:
	$S_{R_v} \leftarrow \text{sorted}(R_v.\text{items}(), \text{key} = \text{operator.itemgetter}(1), \text{reverse} = \text{True})$
11	Find largest count: $C_l \leftarrow \text{list}(S_{R_v}.\text{values}()).\text{count}(S_{R_v}[\text{list}(S_{R_v}.\text{keys}())[0]])$
12	if $C_l == 1$ then
13	$N_m \leftarrow \text{list}(S_{R_v})[0]$ # select the node having largest count as mining node
14	else
15	DA selects N_m randomly
16	end
17	return N_m

Block creation

As soon as the mining node selection is completed, the DA sends the DNP3-encrypted and aggregated measurement data to the mining node. The measurement data is then hashed by the mining node using the technique described in 2.4.1. We follow a Merkle tree-like structure during the hashing operation. Then, the mining node creates a new block (e.g., block K in Fig. 2.2) with the previous hash, index, timestamp, consumption value, current hash (root hash), and nonce.

Block verification

All the relay servers and the DA participate in the block verification process. For the verification process, the servers exchange several types of responses (e.g., ‘already

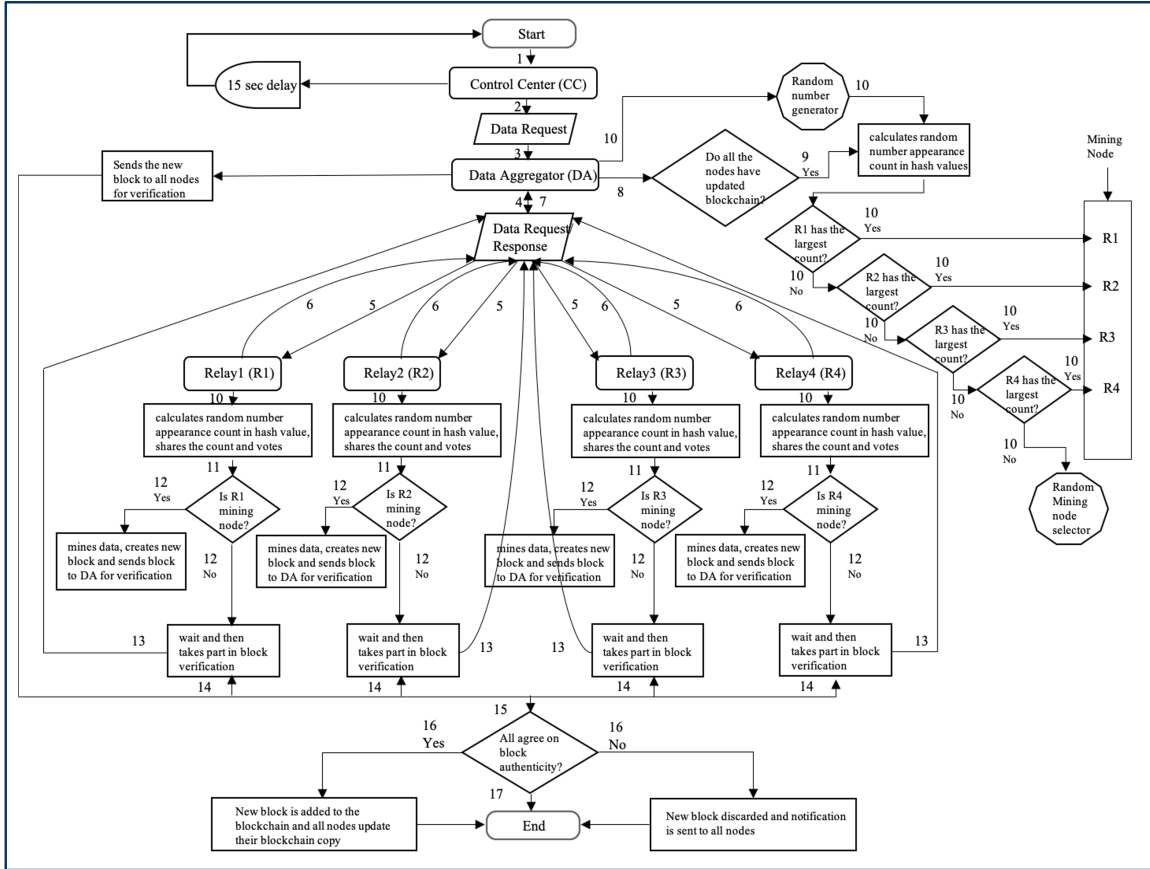


Figure 2.5: Operational process flowchart of the proposed scheme. The process follows the number sequence of the flowchart.

updated’, ‘blockchain manipulated’, ‘chain verified’, etc.) with the nodes. Although this message exchange may add some latency to the overall process, this is particularly crucial for the block verification process.

Block addition

If all parties agree on the validity of the block, the DA sends the block addition request to all the nodes and they follow. Finally, after the successful addition of the block to the blockchain, the DA shares the updated blockchain with the control center and other nodes.

2.5 Performance Evaluation

2.5.1 Experimental Setup

For experimental purposes, we set up a testbed in a virtual machine. We use a Ubuntu 20.04(64 – bit) operating system with a base memory of 8 GB and storage of 20 GB. To create and mimic the real servers/nodes, Mininet version 2.2.2 is used. We use Python 3.0 as the programming language. Moreover, we use Wireshark (previously known as Ethereal) to capture the network packet. DNP3 runs over TCP and UDP. The Wireshark DNP dissector registers for TCP and UDP ports 20000 by default and checks whether the TCP segment contains at least 2 bytes of data or not.

The measurement data is coming from an IEEE 9-bus system [72] (Fig. 2.6). We list the measurement data in Table 2.1. For simplicity, we remove the units of all measurements. Here, V_m & V_p represent the magnitude and phasor angles of voltage whereas P & Q represent the real and reactive power injection respectively. We calculate the power injections by subtracting the power consumption from power generations. Our network topology is shown in Fig. 2.7. We assign each node a specific IP address and port. Table 2.2 presents this network topology mapping with corresponding bus association and nodal mode. This simple network topology mimics the communication infrastructure and is commonly found in the data acquisition system of the modern power grid system.

2.5.2 Experimental Procedure

- *Step-1:* Process starts
- *Step-2:* CC initiates the data acquisition process by sending specific indices to DA. Here, for the demonstration purpose, we consider all indices are requested

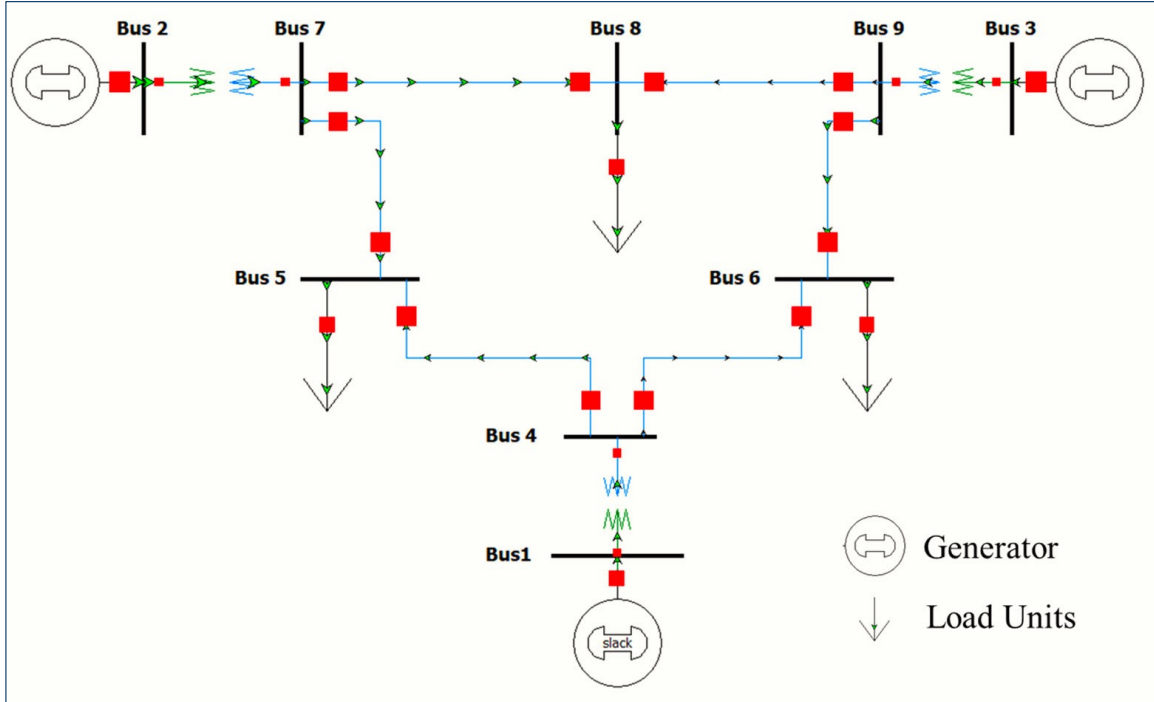


Figure 2.6: IEEE-9 bus system topology

by CC to DA. Also, CC packs the indices in a request header following the DNP3m protocol and sends it to DA over a TCP socket. The IP address of CC is 10.0.0.1 that acts as a client to the DA server (10.0.0.20) in this step.

- *Step-3:* DNP3-encrypted requests reach the DA server
- *Step-4:* DA unpacks the combined request packet from CC and then, packs it into four separate requests for field devices (i.e., relays/smart meters). These request packets include measurement indices according to the assigned relays. DA also creates a TCP socket connection with four relays with four different ports. The IP address of DA is 10.0.0.20 and the port is 20000. In this step, it acts as a server to CC and client to the field devices (relays)
- *Step-5:* DA sends the request packets to relays and waits for the responses
- *Step-6:* Each relay unpacks the DA request and creates a response pack with

Table 2.1: Measurement data mapping

Node	Substation	Type	Index	Values
SM-1	Bus-1	V_m	11	1.00
		V_p	12	0.00
		P	13	71.95
		Q	14	24.07
	Bus-2	V_m	21	1.00
		V_p	22	9.67
		P	23	163.00
		Q	24	14.46
SM-2	Bus-3	V_m	31	1.00
		V_p	32	4.77
		P	33	85.00
		Q	34	-3.65
	Bus-4	V_m	41	0.99
		V_p	42	-2.41
		P	43	0.00
		Q	44	0.00
SM-3	Bus-5	V_m	51	0.98
		V_p	52	-4.02
		P	53	-90.00
		Q	54	-30.00
	Bus-6	V_m	61	1.01
		V_p	62	1.93
		P	63	0.00
		Q	64	0.00
SM-4	Bus-7	V_m	71	0.99
		V_p	72	0.62
		P	73	-100.00
		Q	74	-35.00
	Bus-8	V_m	81	1.00
		V_p	82	3.80
		P	83	0.00
		Q	84	0.00
	Bus-9	V_m	91	0.96
		V_p	92	-4.35
		P	93	0.00
		Q	94	0.00

their measured data according to the indices following the DNP3m protocol. Then they send the responses to DA. The IP addresses of relays are

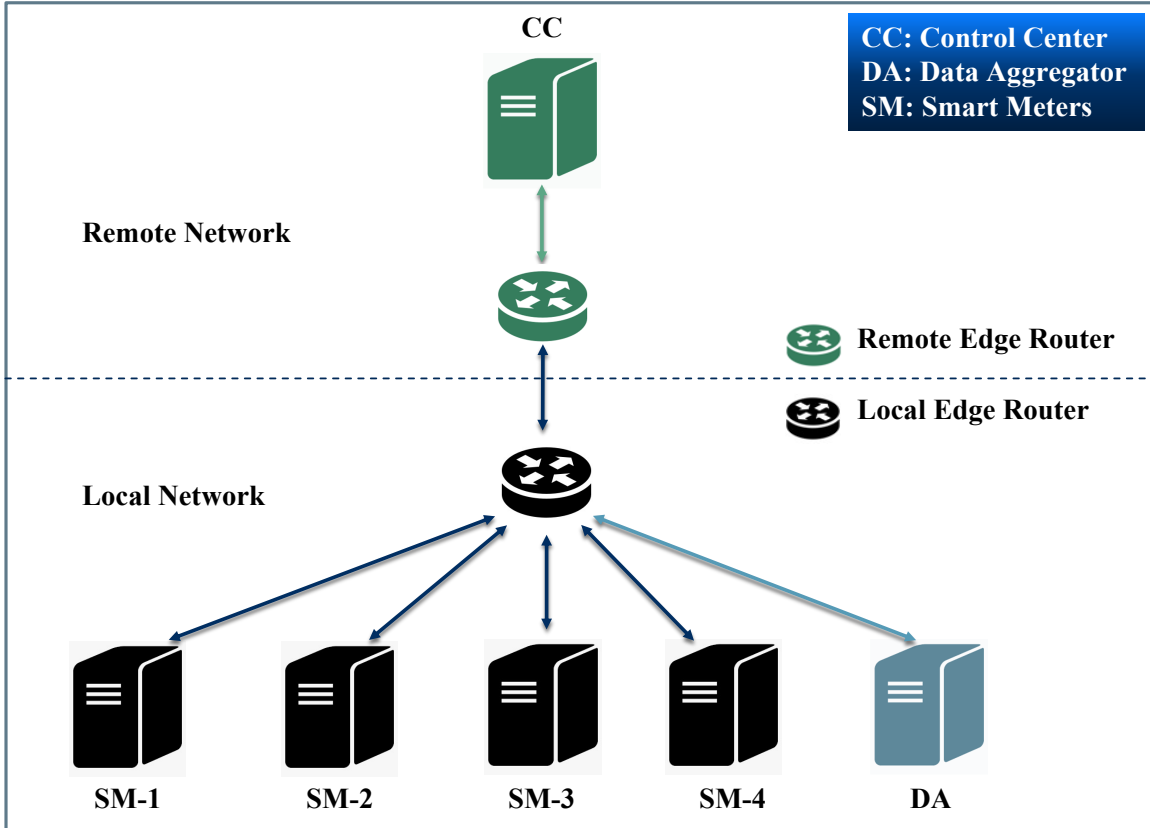


Figure 2.7: Network topology mapping

10.0.0.11, 10.0.0.12, 10.0.0.13, 10.0.0.14 and the ports are 20001, 20002, 20003, 20004.

The relays act as a client to DA

Table 2.2: Network topology mapping

Node	IP Address	Port	Nodal Mode	Bus Association
SM-1	10.0.0.11	20001	Server & Client	Bus-1 & Bus-2
SM-2	10.0.0.12	20002	Server & Client	Bus-3 & Bus-4
SM-3	10.0.0.13	20003	Server & Client	Bus-5 & Bus-6
SM-4	10.0.0.14	20004	Server & Client	Bus-7 & Bus-8 & Bus-9
DA	10.0.0.20	20000	Server & Client	
CC	10.0.0.1		Client	

- *Step-7*: Response packs reach to DA server
- *Step-8*: In this step, DA sends its copy of the blockchain to all nodes (relays) to check whether they have the same copy of the blockchain or not. If they

have the same copy of blockchain as DA, they send replies as ‘already updated’; otherwise ‘Blockchain manipulated’

- *Step-9:* If any node does not have an updated blockchain, it immediately stops the program. On the other hand, if all the nodes have the updated blockchain, the DA starts the mining node selection process
- *Step-10:* As described in the mining node selection process, after receiving measurement data from all relays, the DA sends the random number. All the nodes including DA count the random number of appearance values in the hashed value of measurement data. Then, every node shares its counts with each other and votes for the mining node. The relay server that has the largest count is selected as the mining node for that cycle. If there are multiple largest counts or every node has a count as zero, the DA randomly chooses a mining node from the relay servers
- *Step-11:* If a node is being selected as a mining node, it first aggregates the data into four categories (Voltage Magnitude, V_m ; Voltage Phasor, V_p ; Real Power, P ; Reactive Power, Q). So, even if the mining node is compromised, the attacker would only get the sum value of data for all relays
- *Step-12:* In this step, the mining node mines the data, hashes the data, and creates a new block. While doing so, it also verifies the previous block. On the other hand, the other three nodes wait for the DA to command them to verify this new block
- *Step-13:* Mining node sends the new block to DA for verification purpose
- *Step-14:* DA sends this new block with the previous block to all nodes except the mining node for verification. In the verification part, nodes match their

copies of blockchain with the copy from DA and also match the hash values between the newly created block (previous hash value) and the previous block (current hash value)

- *Step-15:* All the nodes except the mining node send their verification responses ('chain verified') to DA
- *Step-16:* DA gets all the responses and starts its verification process. If the verification is successful, DA sends a command to all nodes except the mining node (since the mining node has already added the block in its chain) to add that new block to their chain. On the other hand, if any of the received verification responses are negative, the new block is discarded, and the program comes to an end
- *Step-17:* All nodes update their chain with a new block, Then, DA sends the updated copy of the blockchain to CC and one cycle completes. But CC initiates the data acquisition process again after every 15 sec (adjustable) and the cycle repeats. For simplicity, when 10 blocks have been added to any chain, we store the latest chain into a separate file following an incremental naming convention.

2.5.3 Experimental Result Analysis

As this is a no reward-no penalty-based private and permissioned blockchain model for modern power grids, miners/validators do not need to compete for mining opportunities. So, a complex mathematical problem-solving approach is avoided. Moreover, unlike PoW, the proposed PoRCH mechanism saves energy that may incur due to huge computational overhead. At the same time, the mining process is random and fair, and thus, provides a level of operational efficiency, data integrity, privacy, and security to the data acquisition system. In short, our PoRCH scheme has the following

advantages over the traditional consensus schemes in this direction:

- PoRCH does not require any incentive mechanism for the participating nodes. Therefore, it solves the huge computational overhead as in PoW. Moreover, through this no reward-no penalty model of PoRCH, biasness in node selection (as in the PoS scheme) is also avoided
- A level of fairness is also ensured as all the relay servers get the chance of voting. At the same time, the random number generator and the random node selector ensure the randomness of the mining node selection process
- As all the servers have the same counts, even if an adversary compromises a server, he/she cannot alter the count to become a mining node without revealing his/her identity. This ensures security in the mining node selection process

Computational Cost

Our experiment shows that an entire operation cycle with four relay servers and four measurement data (one from each server) per block takes only 382 milliseconds in total. In this total cycle completion time, the mining node selection incurs 75% of entire computational burden (Fig. 2.8). Moreover, computational costs of the sub-processes also increase with nodal increment. For example, increasing the number of participating nodes (in our case smart meters) from 04 to 22 significantly increases the total computational cost by more than 06 times. However, The impact of incriminating the number of participating nodes is higher in ‘authentication & verification’ and ‘mining node selection’ processes than in the ‘block creation’ process. It is because, in these sub-processes, both the communication and computational complexities grow; whereas in the ‘block creation’ sub-process, the aggregator does not need to communicate with individual nodes and only records the transaction into blocks.

However, for a large network with a huge number of nodes, the computational cost of the sub-processes can be further reduced by limiting only a subset of total nodes to be eligible for mining nodes in a particular round. For example, if there are total N nodes in the network, we can limit only K nodes (where $K \subset N$ and $K < N$) to be eligible to take part in the mining node selection sub-process. Besides, to make the data acquisition process more secure and trustworthy, other measures that can be adopted include (but are not limited to) data encryption, data anonymization using pseudonyms, node identity anonymization, etc. Similarly, faster authentication could be achieved by using a bloom filter.

We also compare our scheme to (a) PPEA-BPC [1], (b) PPM-HDA [2] and (c) DG-APED [3] in terms of computational cost. As Fig. 2.9 shows, our scheme performs significantly better than the other three schemes.

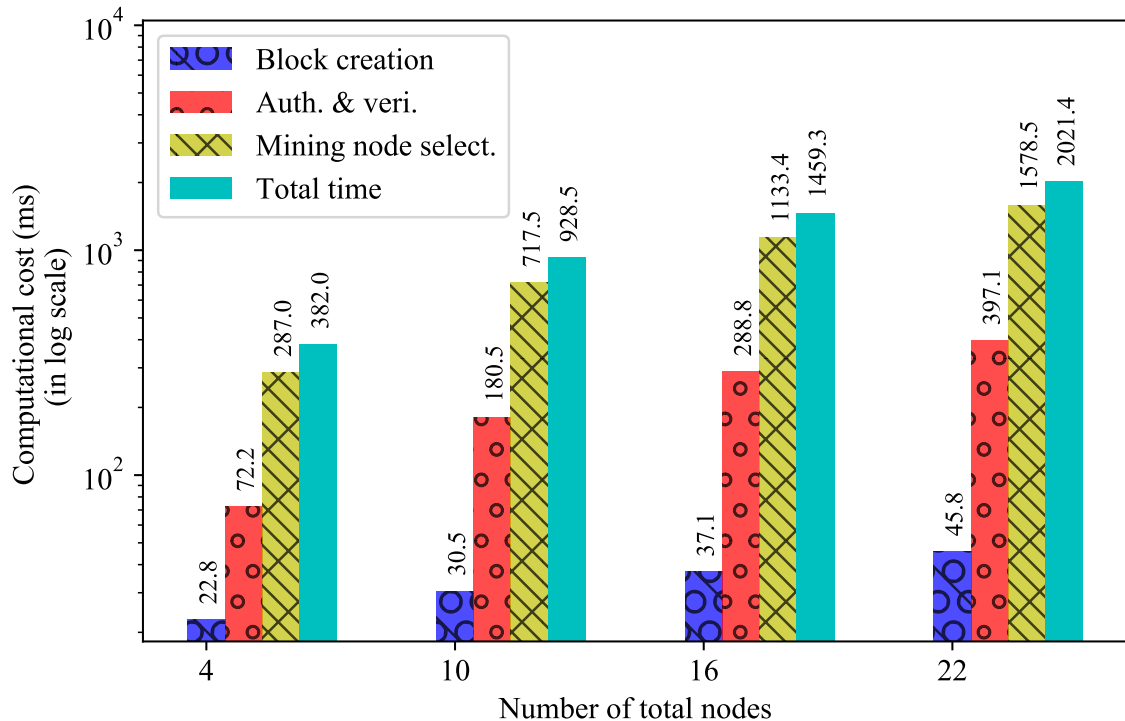


Figure 2.8: Number of total nodes vs computational cost (in ms) of the sub-processes

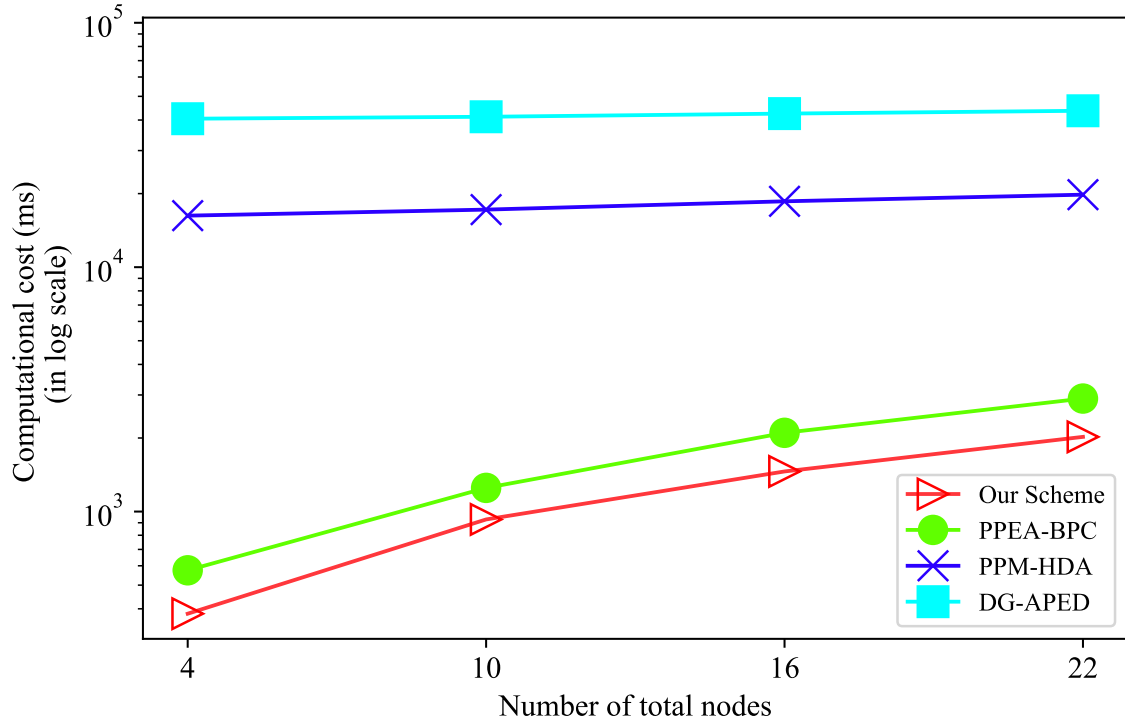


Figure 2.9: Comparison of our scheme with (a) PPEA-BPC [1], (b) PPM-HDA [2] and (c) DG-APED [3]

Hash Collision Avoidance

Usually, there is a very small probability of a hash collision occurring during hashing mechanism. Nevertheless, the probability of hash collision can be further reduced by increasing the array size of the bloom filter [1]. Another way is to use the double-SHA-256 as a hashing technique. Nonetheless, there is an ongoing debate on using double-SHA-256 to overcome the hash collision though bitcoin is already using it as $SHA256(SHA256(x))$ [73], where x is the block_header. Our proposed scheme can also be extended to support measurement data that are hashed by double-SHA-256 using simple python programs.

Limitations

Despite the lightweight operation of our PoRCH mechanism, questions may arise regarding the complexity and computational time for this process. Undoubtedly, a small processing time is more desirable, especially for operations where near-real-time data acquisition is necessary. Selecting the mining node through a simple random process by the DA would reduce both the complexity and computational time. But, in that case, an adversary may easily become a mining node by only compromising the DA. Therefore, a trade-off is essential between the tolerable computational burden and security. Another limitation of our scheme is the lack of resiliency. Our model is still susceptible to some security attacks (e.g., membership inference and false data injection attacks). These can be overcome through the integration of additional privacy and security measures (e.g., differential privacy, secure multi-party computation, etc.); however, they are out of the scope of this research.

2.6 Summary

Modern power grids or simply, smart grids are neither a dream nor a near-future thing anymore, rather it is happening now. The integration of more and more embedded systems and IoT devices is pushing existing power grids forward at a breakneck speed. To cope with this, the modification of age-old SCADA systems is necessary in terms of decentralization, near-real-time operation, security, and privacy. In this context, blockchain technology has the potential of providing not only these essential features of the data acquisition process of future SCADA systems but also many other useful add-ons. On the other side, it is evident that various type of security breach tends to take place more during any economic turmoil. These can cause even more serious devastation to the global economy and human life. Thus, it is necessary to make

our industries robust, automated, and resilient with secured and immutable data acquisition systems.

Taking those into consideration, in this chapter, we discuss the implementation scopes of blockchain in the data acquisition part of SCADA systems in the area of the modern power grid. Several consensus mechanisms have been developed over the years to support blockchain integration in the field of cryptocurrencies, vehicular networks, healthcare systems, e-commerce, etc. But little attention has been paid to developing efficient and easy-to-implement consensus mechanisms in the field of SCADA systems. From this perspective, we propose a novel consensus mechanism (we call it ‘**PoRCH**’) with an effective mining node selection scheme in this chapter. PoRCH bridges the DNP3 protocol to blockchain technology while maintaining randomness and fairness in the process. We also present a small-scale prototype of a blockchain-enabled data acquisition system. The performance evaluation of our proposed PoRCH mechanism verifies the low computational burden of the process. PoRCH also outperforms some existing state-of-the-art models (e.g., PPEA-BPC [1], PPM-HDA [2], DG-APED [3]) in this direction.

Chapter 3

Privacy, Security, and Utility Analysis of Differentially Private CPES

3.1 Introduction

Cyber-Physical Energy System (CPES) such as smart grid data is used for various mission-critical (e.g., state estimation, microgrid islanding, and synchronization, load balancing, etc.) and non-mission-critical applications (e.g., energy consumption prediction, power outage forecasting, etc.) [74–77]. The granular level grid data with numerous features pave the way for state-of-the-art security and privacy research. However, grid data also carries the personal and confidential information of the customers. In most cases, releasing such data is restricted due to security and legal issues [48]. Therefore, the data needs privacy preservation, especially during data sharing or data exchanging operations [78].

Over the years, several research works have been conducted on the data privacy-preservation mechanisms (e.g., secure multi-party communication, k-anonymization,

l-diversity, etc.) [79] and the privacy-violating attacks on those mechanisms [27–29]. To overcome such privacy-violating attacks while preserving data confidentiality and maintaining a level of data utility, the concept of differential privacy (DP) has been introduced by Dwork et al. [18]. The DP mechanism adds sufficient randomized noise to the aggregation result. This randomized noise prevents the attacker from revealing the identity of an entity by obscuring the contribution of a single record. This privacy assurance can motivate the data providers to provide open access to their differentially private databases to the research community and others.

3.1.1 Adversarial Exploitation of Differential Privacy

Although the DP mechanism achieves a privacy guarantee, a well-known drawback of it is the degradation of data utility with the increments of the data privacy [18]. Moreover, from recent research by Giraldo et al., we find evidence that differentially private data (i.e., DP-data) can be exploited by the attacker to conduct a false data injection (FDI) attack on a DP-based system in a smart grid domain [63]. These data integrity attacks on differentially private data have led us to the question- “*What should be the design approach to defend a DP-based cyber-physical energy system (CPES) against FDI attack?*”. To answer these questions, it is essential to find out the factors that facilitate this FDI attack. Likewise, to design a fault-tolerant, resilient, privacy-preserved, and well-secured defense mechanism, a provable relationship among these factors, outcomes, and parameters of the DP method needs to be developed.

3.1.2 Impacts of DP integration

Besides, any active or passive attack degrades the overall QoS of a system. So, it is also essential to raise and subsequently answer the question- “*What are the impacts of*

the FDI attacks on the QoS of a DP-based system?". Based on this impact analysis of the FDI attack, the feasibility analysis of the DP technique in the CPES domain needs to be carried out in an adversarial setting. Successively, the feasibility analysis would help the CPES designers to determine the proper applicable areas of the DP mechanism. As DP can be a tool for the attacker to conduct the FDI attack, it should not be applied in the mission-critical operations of any CPES just to enhance data privacy. On the contrary, the method must be applied with proper security measures. Otherwise, the entire CPES can be comprised of any malicious actor. Our motivations behind this research are also influenced by these defense perspectives.

3.1.3 Contributions

Earlier research has focused on building robust anomaly detection algorithms to either defend a DP-based CPES or improve the accuracy of an ML-based prediction algorithm [63, 80, 81]. However, if the DP technique is applied in several layers of the CPES network, then identifying the false data is very difficult (if not impossible) for the traditional anomaly detectors. Therefore, novel defense strategies need to be devised against FDI attacks in DP-based smart grid applications, and for that more research should be carried out to analyze the correlation among DP parameters.

In this chapter, we address and analyze the concern of the FDI attack exploiting the differentially private data through adversarial analysis of the DP-based mechanisms in the smart grid domain. Our main contributions to this research work can be summarized as follows:

- We demonstrate the formulation of a provable relationship between the attack impact and the parameters (data sensitivity, privacy loss, and attackers' tolerance to be detected) associated with the DP mechanism. This relationship enables the designer to design a robust DP-based privacy-preserving system.

- We, both theoretically and empirically, show that the maximum attack impact of the FDI attack could be minimized if the designer uses our proposed approach to design the process in the first place.
- We rigorously analyze the feasibility of the DP mechanism in CPES data under an adversarial setting from a QoS point of view. We further evaluate the usability of the data considering an attack scenario where the attacker manipulates the data despite all the privacy and security measures. The evaluation shows that the manipulated data are usable for many non-mission-critical operations of CPESs like smart grids.

3.2 Problem Statement and Threat Model

In this section, we formulate the main research problem, as well as the objectives. Also, a threat model is developed. Table 3.1 provides the symbols and their description used in this chapter.

3.2.1 Problem Statement

Fig. 3.1 depicts the basic operational principle of a DP-based system. A mechanism is ε - indistinguishable if for all pairs $X, X' \in D^n$ which differ in only one entry, for

Table 3.1: List of symbols and their description

Symbols	Description	Symbols	Description
X	Dataset contains all the measurement value	μ_a^*	Optimal attack impact
X'	Dataset differs by a single record from X	τ	Query result deviation threshold
t	Transcripts	V	Measured voltage by PMU
ε	Privacy loss	f_0	PDF of Laplace distribution
Δf	Data sensitivity	f_a	PDF of attack distribution
Q	Differentially private data	f_a^*	PDF of optimal attack distribution
η	Noise drawn from Laplace distribution	θ	Mean or location parameter of Laplace distribution
η_a	Noise drawn from Optimal attack distribution	k_1	Kullback-Leibler divergence
b	Scale parameter of Laplace distribution	y	Query result
μ_a	Attack impact	γ	Attackers' tolerance to be detected

all adversaries A , and for all transcripts t [18]:

$$\ln \left| \frac{Pr[T_A(X) = t]}{Pr[T_A(X') = t]} \right| \leq \varepsilon \quad (3.1)$$

Equation (3.1) provides the privacy guarantee of DP mechanism. Applying DP during a sum query over a database containing $\{x_1, x_2, x_3, \dots, x_i, \dots\}$ as input, we get-

$$\sum_i x_i + Y; Y \sim Lap\left(\frac{\Delta f}{\varepsilon}\right) \quad (3.2)$$

Here, Lap represents the Laplacian distribution mechanism. Δf is the sensitivity of the data which is inherent in the dataset. For desired privacy, the noise is calibrated according to the sensitivity of the data. Larger noise yields a smaller value of privacy loss and vice versa. However, more noise leads to less data utility. Besides, applying large noise into the data for ensuring better privacy provides an attacker the opportunity of injecting false data into it. So, if the DP technique is applied in several

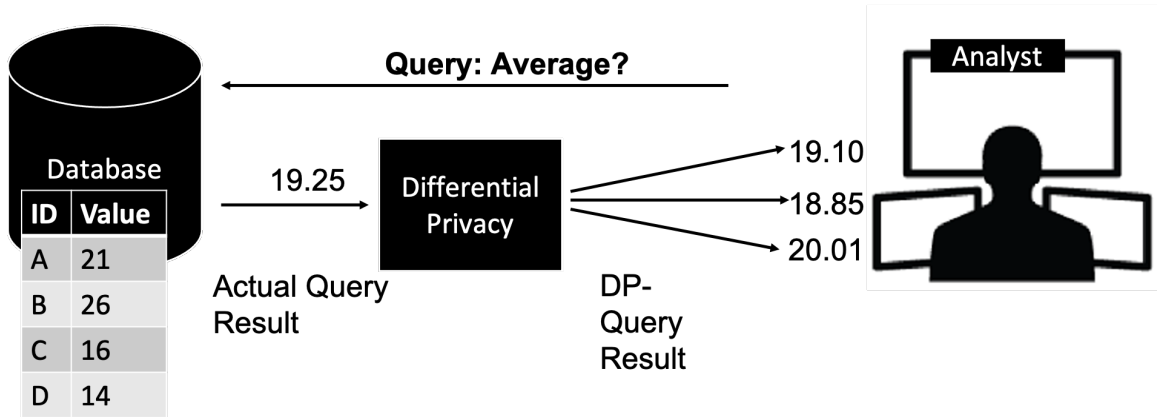


Figure 3.1: Basic operation of differential privacy mechanism. Laplacian noise is added to the query to make the query result ε -indistinguishable. For each query, the differentially private query results are different.

layers of a CPES network and an attacker injects a low amount of false data into the DP-data (differentially private data), then the traditional anomaly detectors cannot detect the manipulation.

An alternative defense strategy that can minimize the attack impact is to design and apply the DP technique targeting minimum attack impact and maximum privacy. For this, a verifiable correlation between the outcome (attack impact) and the parameters of the DP method needs to be developed that can enable the designer to design a resilient and privacy-preserved system as a part of the defense mechanism.

The primary objective of our analysis is to enable the defender to design a privacy-preserved and secured DP-based system (e.g., synchrophasor network) against FDI attacks. The secondary objective of our research is to analyze the feasibility of the DP mechanism in synchrophasor networks considering the vulnerability of the DP-based system against FDI attacks. We also analyze the QoS of the DP-based synchrophasor network in an adversarial setting. More precisely, we want to raise and subsequently answer the following questions.

Design objective question

Given a dataset (D), what will be the value of scale parameter (b) and epsilon (ϵ) such that the attack impact (μ_a) becomes equal or close to the actual query result (i.e., $\mu_a = Q(D) \vee \mu_a \rightarrow Q(D)$)?

Feasibility analysis objective question

Would the DP technique remain feasible over other privacy-preserving mechanisms (e.g., encryption, masking, anonymization, etc.) considering its security challenges (i.e., the chance of FDI attacks exploiting DP-noise)?

QoS analysis objective question

What would be the QoS of the DP-based power grid in terms of computational overhead and data utility under the FDI attack?

3.2.2 Threat Model

The proposed threat model, as depicted in Fig. 3.2, considers the FDI attack on the data packets while transferring from one node to another node through the spoofing technique. Throughout the attack, the attacker pretends to be an honest node to all other nodes and injects false measurements as a form of noise into DP data. The model is developed based on an attack scenario, where the attacker can conduct the FDI attack despite all the privacy and security measures. More formally, for a DP-result of any aggregated query over a database consisting of grid measurement, the FDI attack is expressed as follows:

$$Q_a = Q(D) + \eta_a \text{ s.t. } Q(D) = \sum_{i=1}^n v_i + \eta \forall v_i \in V \quad (3.3)$$

where, $V = \{v_1, v_2, \dots, v_n\}$ is the dataset containing n number of measurement values obtained from a μ PMU dataset, η is the measure of Laplace noise, $Q(D)$ is the non-manipulated differentially private query result over database D , η_a is the false measurement injected by the attacker as a form of noise and Q_a is the manipulated query result. Moreover, the Laplace noise, η is a random variable with a probability distribution that satisfies the condition of the differential privacy (as stated by (3.1)). Other than the Gaussian and the Exponential distribution, the Laplace distribution is also a good choice for extracting random noise as it satisfies $(\epsilon, 0)$ or (ϵ, δ) – differential privacy. Also, the Probability Density Function (PDF) of Laplace distribution has a fatter tail which is why Laplacian noise provides better privacy. The Laplace

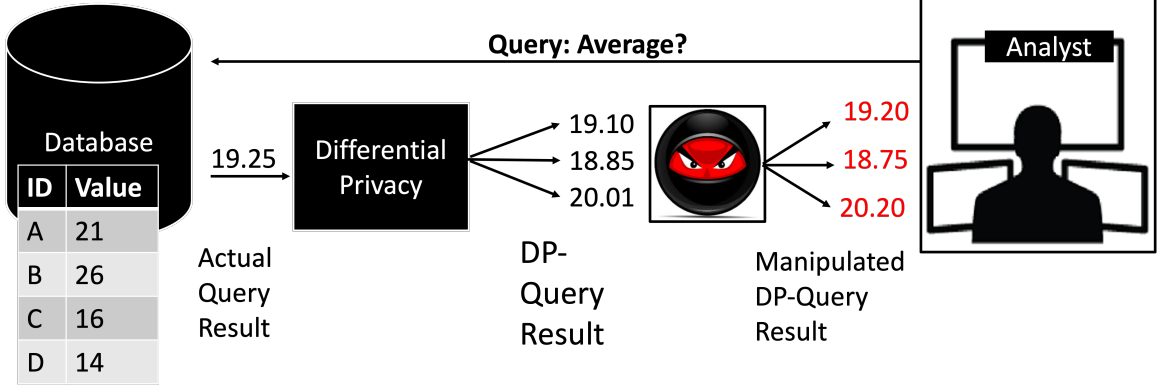


Figure 3.2: Differential privacy in an adversarial setting. Differentially private query results are manipulated by a malicious actor and sent to the analyst. A small amount of manipulation is difficult to detect

distribution with mean ($\theta = 0$), variance ($\sigma^2 = 2b^2$) and PDF, f_0 can be described by (3.4).

$$f_0 = \frac{1}{2b^2} \exp \frac{-|y - \theta|}{b} \text{ s.t. } b = \frac{\Delta f}{\epsilon} \quad (3.4)$$

Here, b is the scale parameter and can be represented as the ratio between the sensitivity of the data (Δf) and the privacy loss (ϵ). We consider that the attacker varies the value of η_a according to her attack tolerance level (i.e., the willingness to be detected or remain undetected). Now, if η_a is sufficiently small, then it is unlikely for the recipient to detect the manipulation without a well-designed and robust anomaly detector. On the other hand, the anomaly detectors also add extra latency to the network and slow down the system performance. This is a potential threat to the system's performance that needs to be minimized.

3.3 Optimal Attack and Defense Strategy

In this section, we first describe the attackers' and the defenders' objectives. We then compute the correlation among the DP parameters to facilitate defense against the optimal attack.

3.3.1 Attacker's Objectives

An attacker can be a passive attacker or an active attacker. A passive attacker eavesdrops on the communication path while an active attacker directly interferes with the data (e.g., masquerades, modification of messages, denial of services, etc.). The DP mechanism gives protection against passive attacks. However, the attacker can still modify the DP data by injecting false data into them, which becomes a concerning security issue.

Generally, the attacker of our threat model has two major objectives while carrying out the attack—“(a) *maximum damage* (b) *avoid detection*”. These two goals are contradictory to each other in the sense that it becomes difficult for an attacker to hide the attack or hide her identity if she wants to carry out maximum devastation to the system.

3.3.2 Defender's Objectives

Defender's main objective is to defend the system against passive and active attacks. The defense can be carried out from various fronts of the system. One such technique can be setting up the anomaly detectors to minimize the attack impact. But, this would add extra latency to the system performance. Another way could be designing the system as a specific fault-tolerant in the first place considering the possible FDI attacks in the future. In our model, the defender has two major objectives—“(a)

design the process as fault-tolerant against FDI attacks (b) preserve the data privacy and QoS of the system".

These specific design requirements have led us to the quest of finding a provable relationship among the DP parameters to achieve desired privacy, utility, security as well as the feasibility of the DP mechanism.

3.3.3 Optimal Attack Strategy

An optimal attack would take place when the objectives of the attacker are achieved with maximum possible payoffs. In [63], Giraldo et al. have discussed this optimization problem. They show that an optimistic attacker would draw noise η_a from a probability density function, f_a^* , which can be represented as follows:

$$f_a^* = \frac{k_1^2 - b^2}{2bk_1^2} \exp \frac{-|y - \theta|}{b} + \frac{(y - \theta)}{k_1} \quad (3.5)$$

Here, b is the scale factor and θ is the location parameter. k_1 is a solution to KL (Kullback–Leibler) divergence between two probability distributions $D_{KL}(f_a || f_0)$ and can be found by solving below equation –

$$\frac{2b^2}{k_1^2 - b^2} + \ln\left(1 - \frac{b^2}{k_1^2}\right) = \gamma \quad \forall k_1 > b \quad (3.6)$$

Here, γ is the attackers' tolerance to be detected. A large γ would mean that the attacker does not care to be detected whereas a small γ means hiding her entity is crucial for the attacker. To achieve her second objective which is to remain undetected as long as possible the value of γ should be as small as zero. However, then the attackers' probability distribution would be similar to the original probability distribution of the Laplace mechanism (i.e., $f_a^* \simeq f_0$). So, the attacker must go for a

trade-off between the two objectives. If the attacker chooses a small value of γ , i.e., to remain undetected as to fulfill the second objective, k_1 approaches infinity. Again, as k_1 approaches to infinity, the PDF, f_a^* as described by (3.5) approaches to f_0 . The optimal attack impact is given by the following formula [63].

$$\mu_a^* = \frac{b^2(\theta - 2k_1) - \theta k_1^2}{b^2 - k_1^2} \quad (3.7)$$

Here, μ_a^* is the amount of optimal bias introduced by the attacker. It depends on four parameters (mean, θ ; attacker's tolerance, γ ; data sensitivity, Δf and privacy loss, ε). The optimal attack impacts (μ_a^*) largely vary with the privacy loss and attackers' tolerance. More specifically, a low privacy loss (or a large amount of noise) and a high attacker's tolerance help the attacker to conduct a more devastating attack (i.e., the high value of μ_a^*) by injecting large false data and hiding behind the large noise of differential privacy.

3.3.4 Parameter Design for Effective Defense Mechanism

Among the four parameters, the mean (θ , also called the 'location' in PDF) and the data sensitivity (Δf) can be calculated from the targeted dataset. Hence, these two parameters are not adjustable by the user as per the design requirements. However, these parameters are crucial for tuning the noise while applying differential privacy so that, either, the attack impact cannot be large, or the attack can be easily detected.

During the design process, the attackers' tolerance (γ) should be selected in such a way that considers the maximum possible attack impact or deviation from the actual data. The possible incidents that can take place are given below.

- The attacker chooses a very large γ to maximize the attack impact without paying much attention to detection. In that case, a carefully designed anomaly

detector can easily identify the attack.

- The attacker chooses a very small γ to minimize the disclosure of her identity. In that case, the attack impact or deviation would be negligible (as f_a^* becomes equal to f_0) and can be ignored.
- The attacker chooses an optimal value of γ for an optimal attack impact and identity disclosure. In that case, the attack impact is not only tough to detect (if not impossible) but also impractical to be neglected totally. We elaborate and analyze this point in section 3.4.

We are now ready to present one of the main contributions (i.e., correlation among the DP-parameters considering adversarial presence).

Theorem 4.1: For any optimal attack strategy (stealthiness: γ , attack impact: μ_a^*), the correlation among the DP-parameters that solves the design problem stated by 3.2.1 can be represented as follows:

$$\varepsilon = \frac{\Delta f}{b} \text{ s.t. } b = \sqrt{\frac{k_1^2(\mu_a^* - \theta)}{2k_1 + (\mu_a^* - \theta)}} \quad \forall |\mu_a^* - \theta| \geq 0 \quad (3.8)$$

Proof: The scale factor, b needs to be greater than zero, otherwise (3.4) becomes undefined. Moreover, according to the optimal attack strategy described in 3.3.3, the optimal attack follows the Laplace distribution. Therefore, the absolute deviation of the attack impact from the mean of the query data (alternatively, Laplace density) cannot be less than zero (i.e., $|\mu_a - \theta| \geq 0$). Under these conditions, we solve for the scale parameter, b from (3.7) following simple algebraic rules and get–

$$b = \sqrt{\frac{k_1^2(\mu_a^* - \theta)}{2k_1 + (\mu_a^* - \theta)}} \quad \forall |\mu_a^* - \theta| \geq 0 \quad (3.9)$$

Since it is an optimization problem in a constrained environment, we model k_1 as a Lagrange multiplier and compute k_1 numerically using (3.6) and (3.9). Thus, the general expression of (3.6) can be rewritten as (3.10). Moreover, we can solve k_1 numerically from (3.10) and successively compute ε from (3.11).

$$\frac{\mu_a^* - \theta}{k_1} + \ln \left| \frac{2k_1}{2k_1 + (\mu_a^* - \theta)} \right| = \gamma \quad (3.10)$$

$$\varepsilon = \frac{\Delta f}{b} \quad (3.11)$$

We can further evaluate the proposed theorem through boundary conditions.

Case-1 (lower bound, $\mu_a^* - \theta \rightarrow 0$): When $\mu_a^* - \theta \rightarrow 0$ (i.e., when we want to minimize the attack impact and make the deviation zero), (3.10) becomes-

$$0 + \ln \frac{2k_1}{2k_1 + 0} = \gamma \implies \gamma \rightarrow 0 \quad (3.12)$$

Here, $\gamma \rightarrow 0$ indicates that the attacker does not want to be detected at all and hence the payoff of the attacker (i.e., $\mu_a^* - \theta$) approaches to zero. After rewriting (3.9), we can apply limit $(\mu_a^* - \theta) \rightarrow 0$ as follows:

$$\lim_{(\mu_a^* - \theta) \rightarrow 0} b = \lim_{(\mu_a^* - \theta) \rightarrow 0} \sqrt{\frac{k_1^2}{\frac{2k_1}{\mu_a^* - \theta} + 1}} \implies b \rightarrow 0 \quad (3.13)$$

Remark 4.1: When $b \rightarrow 0$, then $\varepsilon \rightarrow \infty$ since $\varepsilon = \frac{\Delta f}{b}$ Equation (3.13) and Remark 4.1 indicate that if the privacy loss, ε approaches infinity (alternatively, noise approaches zero) and the attacker does not want to be detected ($\gamma \rightarrow 0$), then the attack impact is minimum and the deviation from the mean is close to zero (i.e.,

$\mu_a^* - \theta \rightarrow 0$).

Case-2 (upper bound, $\mu_a^* - \theta \rightarrow \infty$): If we consider the maximum attack impact (i.e., $\mu_a^* - \theta \rightarrow \infty$), from (3.10), it can be inferred that $\gamma \rightarrow \infty$, which means the attacker does not care about to be exposed and goes for maximum attack impact. From 3.3.3, we know that, when $\gamma \rightarrow \infty$, then $k_1 \rightarrow 0$. Now, when $k_1 \rightarrow 0$ is very small and $(\mu_a^* - \theta) \rightarrow \infty$, then the scale factor (b) is very large.

Remark 4.2: When $b \rightarrow \infty$, then $\varepsilon \rightarrow 0$ since $\varepsilon = \frac{\Delta f}{b}$. Therefore, if the privacy loss, ε approaches zero (0) (alternatively, noise approaches infinity) and the attacker does not care to be exposed ($\gamma \rightarrow \infty$), then the attack impact is maximum and the deviation from the mean very large (i.e., $\mu_a^* - \theta \rightarrow \infty$).

In summary, (3.9) – (3.11) can be used to adjust the privacy loss (ε) for a given sensitivity, Δf , an optimal attack impact, μ_a^* , and a level of attackers' tolerance, γ . The defender can use the correlation of (3.9) and (3.11) to design a robust, fault-tolerant, privacy-preserved, and well-secured DP-based system. We discuss the effectiveness of our proposed design approach analytically in section 3.4.

3.3.5 Modelling the Criterion for Feasibility and QoS Analysis

Analyzing QoS of the DP-based system under attack scenario is another key objective of our research. Here, QoS refers to the measurement of the overall performance of the system. To quantitatively measure the QoS of the system, several parameters of the network are considered, such as overall latency, congestion, data availability, data utility, system complexity, network speed, operational overhead, data redundancy, system resiliency, etc. As we have not considered any data availability attacks (e.g., DDoS attack, delay attack, etc.) in our model, we have eliminated the data availability from our QoS criterion list. Moreover, the DP technique is a better choice than other privacy-preserving techniques in terms of system complexity, system resiliency,

and operational overhead. The cryptographic encryption techniques incorporate key generation, key distribution, and third-party involvement while the anonymization techniques require huge time for a large dataset. Consequently, we measure the QoS of the attack-prone DP-based synchrophasor network through overall data utility along with the privacy and security cost of DP.

The feasibility analysis is an assessment of the practicability of a proposed plan. The analysis is carried out by asking and subsequently answering the question- "*Is the proposed method feasible?*". We analyze this feasibility question under an adversarial scenario. A method is labeled as feasible if it preserves the specific requirements (e.g., privacy, security, etc.) as well as the QoS of the system. From this perspective, the DP technique is feasible for the synchrophasor network if it provides a substantial amount of data privacy, does not violate the security of the system, and does not degrade the QoS of the system below a satisfactory level. At the same time, it is also necessary to compare the DP technique with other existing privacy-preserving techniques in terms of computational overhead. We carry out such a comparison in section 3.4.

3.4 Experimental Analysis

For the experimental purpose, we have used the same dataset used by [82]. The dataset includes the real-time measurement of dedicated PMUs connected on the medium voltage side of the network secondary substations in a smart grid on the EPFL campus. We have selected this data because similar data are being used in various applications in the smart grid domain, e.g., state estimation, modeling power networks, predicting future needs, etc. The data contains historical measurement values from 5 PMUs starting from the year 2014 to 2019 with a high data missing

percentage. However, since the frequency of the measurement is the fraction of seconds (ms), the size of the dataset is sufficient for the practical demonstration purpose of our proposed model.

3.4.1 False Data injection Attack Simulation in a Synchronphasor Network of a Power Grid

The FDI attack can take place in different nodes (master nodes, fog nodes, edge nodes, etc.) of a synchronphasor network. The attacker can conduct the attack either by compromising nodes (or sensors) through direct access or by compromising the communication channel through the traditional Man-in-the-Middle (MITM) attack approach [83]. As physical access to a grid network is difficult for an attacker to achieve, an FDI attack by physically accessing the node is unlikely to occur. Therefore, for our proposed model we have considered the FDI attacks conducted by compromising the communication channel of the synchronphasor network.

In a synchronphasor network, the queries among the master nodes (controller), fog nodes (PDCs-Phasor Data Concentrates), and edge nodes (PMUs) flow from the master nodes towards the edge nodes. When DP is applied over the entire synchronphasor network, hierarchically, the receiving node (e.g., PDC) gets the DP data from the lower-level nodes (e.g., PMUs). However, as there are multiple layers in the network, the DP technique adds random noise multiple times in the same query result. So, it becomes difficult for the anomaly detectors to identify the false data.

In our experiment, the query is made to find out the hourly average of the user consumption, P (in KWh) recorded by a PMU on a particular day. Fig. 3.3 depicts the differentially private results (DP-results) of the query along with the actual result for varying privacy loss (ϵ). The DP results differ from the actual query results due

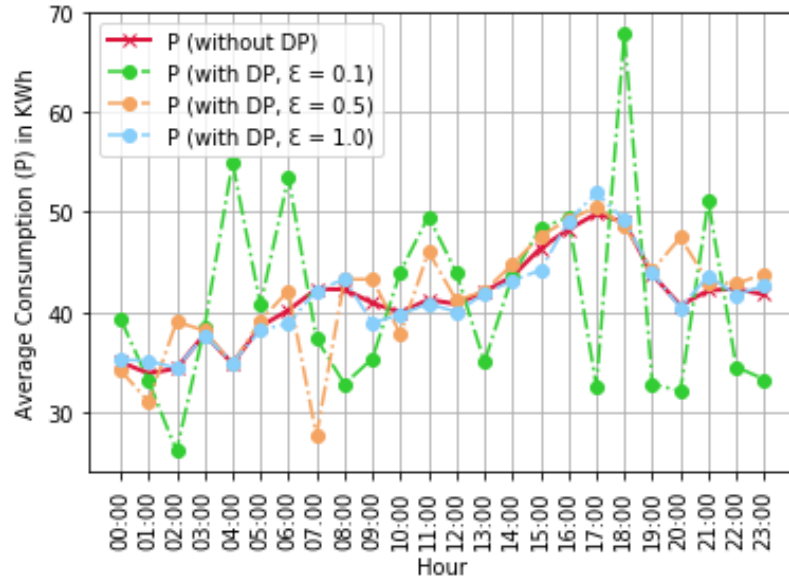


Figure 3.3: Hourly average user consumption (KWh) recorded by a PMU. Small level of privacy loss (ϵ) leads to higher deviation. The DP-noise has been drawn from a Laplace distribution with zero mean and $2b^2$ variance ($\theta = 0, \sigma^2 = 2b^2$)

to the random Laplacian noise added through the DP process. For demonstrating the FDI attack, some false measurement is injected deliberately as a form of noise into the DP result. The false measurement is drawn from the attackers' probability distribution following the optimal attack strategy of section 3.3.3. Corresponding attack impact is illustrated in Fig. 3.4 for various level of ϵ . The attack impact decreases for the following cases: (i) the privacy loss increases, (ii) the attacker's tolerance decreases, and (iii) data sensitivity decreases.

3.4.2 Privacy, Security, and Utility Analysis

According to the ϵ -indistinguishability condition, a DP-result differs from the actual query result (Fig. 3.3). When the noise is large (alternatively, ϵ is lower), the DP-result of average user consumption deviates significantly from the true result and vice versa. The amount of random noise also depends on the data sensitivity. Highly

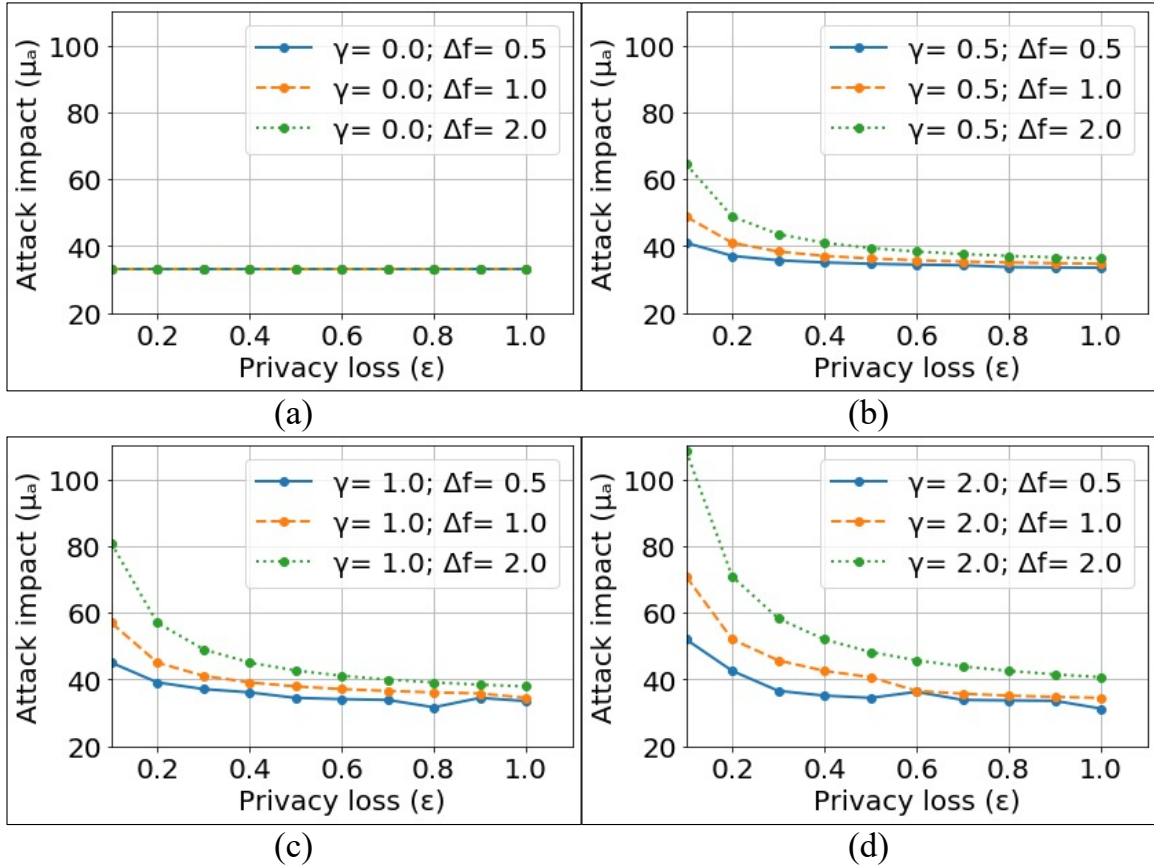


Figure 3.4: Correlation between the attack impact (μ_a) and the DP-parameters. Attack impact (μ_a) decreases when- (i) privacy loss (ϵ) increases, (ii) attackers' tolerance (γ) decreases, and (iii) data sensitivity (Δf) decreases.

sensitive data requires a large amount of noise to preserve data privacy. On the contrary, low sensitive data needs a small amount of noise.

Consequently, for a small amount of noise (alternatively, a high value of ϵ), the DP-result differs from the true result only by small fractions (curve without DP and with $\epsilon = 1.0$ in Fig. 3.3). So, for the non-critical applications requiring less sensitive data in the smart grid domain, the required amount of noise to preserve the data privacy is also low.

Correlation between the attack impact and the DP-parameters

The data sensitivity level has an important role in the DP mechanism. If the defender designs to add large noise to highly sensitive data, then the attacker can conduct devastating attacks. For example, in Fig. 3.4, the privacy is very large when $\varepsilon = 0.1$. Also, for highly sensitive data (e.g., $\Delta f = 2.0$), if the defender keeps the privacy loss low ($\varepsilon = 0.1$) to preserve privacy, the attack impact is as large as 110. However, the attacker may not create a large attack to avoid the attack detection; rather she would follow the optimal attack distribution described by (3.5) for an optimal payoff.

In our experiment, the true query result (true average user consumption) is 33.18. Following the optimal attack distribution interpreted by (3.5), the attacker can add maximum noise of 76.82 if her tolerance level (γ) is 2.0. Now, if γ is increased by 0.5 (e.g., 0.5 to 1.0) while Δf remains same (e.g., 0.5), the maximum possible attack impact is increased by approximately 4.0 (e.g., 41.0 to 45.0). On the contrary, if Δf rises by 0.5 (e.g., 0.5 to 1.0) while γ remains same (e.g., 0.5), the attack impact increases by approximately 8.0 (e.g., 41.0 to 49.0). So, it is perceived that the attack impact depends more on Δf than γ .

Moreover, we see that the actual mean (θ) has no significant effect on the attack impact. For any mean (θ), the attack impact (μ_a) follows the same principle. That means if the mean is lower than 33.18, the attack impact also decreases by the same amount following the DP principle. Based on this correlation between the data sensitivity (Δf), the attackers' tolerance (γ), and privacy loss (ε), the designer can design his system targeting minimal attack impact, high data privacy, and high data utility. More specifically, the designer can calibrate the noise following (3.9)-(3.11) and adjust the tolerance of the system up to the calculated attack impact value.

QoS of the DP-based synchrophasor network under adversarial setting

We use the defense cost and the data utility as indicators to evaluate the QoS of the DP-based system in an adversarial setting. The defense cost is the overall cost that arises from the defensive measures against cyber-attacks. It represents the distance of the DP data from actual data. It consists of two separate costs- the privacy cost and the security cost. Privacy cost (i.e., the distance of the DP data from actual data due to privacy measures) takes place when the DP technique is applied to the query result. It increases with the number of times the DP mechanism has been applied in the nodes. On the other hand, the security cost is added to the defense cost when the FDI attacks occur. Besides analyzing the defense cost of a DP-based synchrophasor network under an attack framework, we also measure the data utility in terms of prediction accuracy. The data utility measurement is performed with both the DP-data and the FDI-DP-data (i.e., the DP-data that is manipulated by the FDI attack).

Motivated by [84], we build the proposed prediction model to evaluate the above-mentioned properties. The model simply executes a time-series prediction algorithm on the PMU dataset for different levels of privacy loss (ϵ). The experiment is conducted with the original data, the DP-data, and the FDI-DP-data. After executing the experiment with DP data, the defense cost (i.e., the privacy and the security cost) is measured on the predicted values. Next, the FDI attack is conducted deliberately on both datasets (with and without the DP mechanism). After that, the manipulated datasets are fed into the prediction model. Finally, the security cost (i.e., the prediction deviation) due to the FDI attack on the DP data is measured.

We first study the defense cost (i.e., the sum of privacy cost and security cost) and find that the privacy cost is higher than the security cost when the attacker conducts a stealthy attack (as shown in Fig. 3.5). However, the overall defense cost reduces

significantly when the privacy loss (ϵ) is increased. Then, we investigate the utility

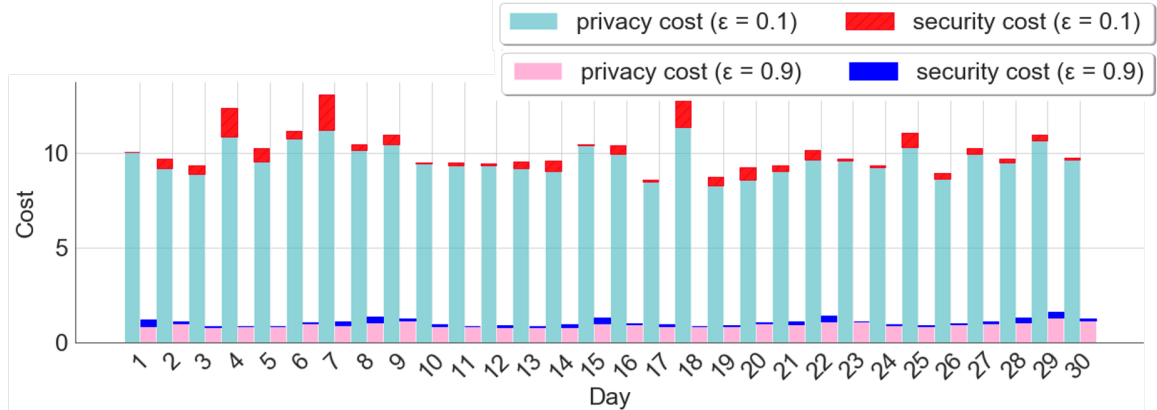


Figure 3.5: Utility cost analysis of DP method for varying privacy loss (ϵ). Privacy cost is higher than the security cost. Moreover, the overall cost is small when epsilon is large (i.e., $\epsilon = 0.9$)

of both the DP-data and FDI-DP-data through a time-series prediction algorithm. We find that the “DP: predicted” value is higher than the “Original: predicted” value due to the privacy cost of the DP mechanism as shown in Fig. 3.6(a). After that, we deliberately conduct the FDI attack (from “2018–10–14” to “2018–11–14” in Fig. 3.6(b) on the original data. Note that the FDI attack has a significant impact on the predicted values. Finally, the same degree of FDI attack is conducted intentionally on the DP-data to find the changes in the predicted values. The findings indicate that the distance between prediction accuracy with DP-data and FDI-DP-data is nominal as the security cost is very low (Fig. 3.6(c)). We find that the impact of the FDI attack for a short period is low on the predicted values and can be negligible for the non-critical applications of the smart grid domain.

Feasibility of DP in a synchrophasor network of a power grid

The criterion for our feasibility analysis of the DP mechanism is the comparison of overall latency with other privacy-preservation techniques. The latency comparison is carried out between the DP technique and the AES-256 encryption technique under an

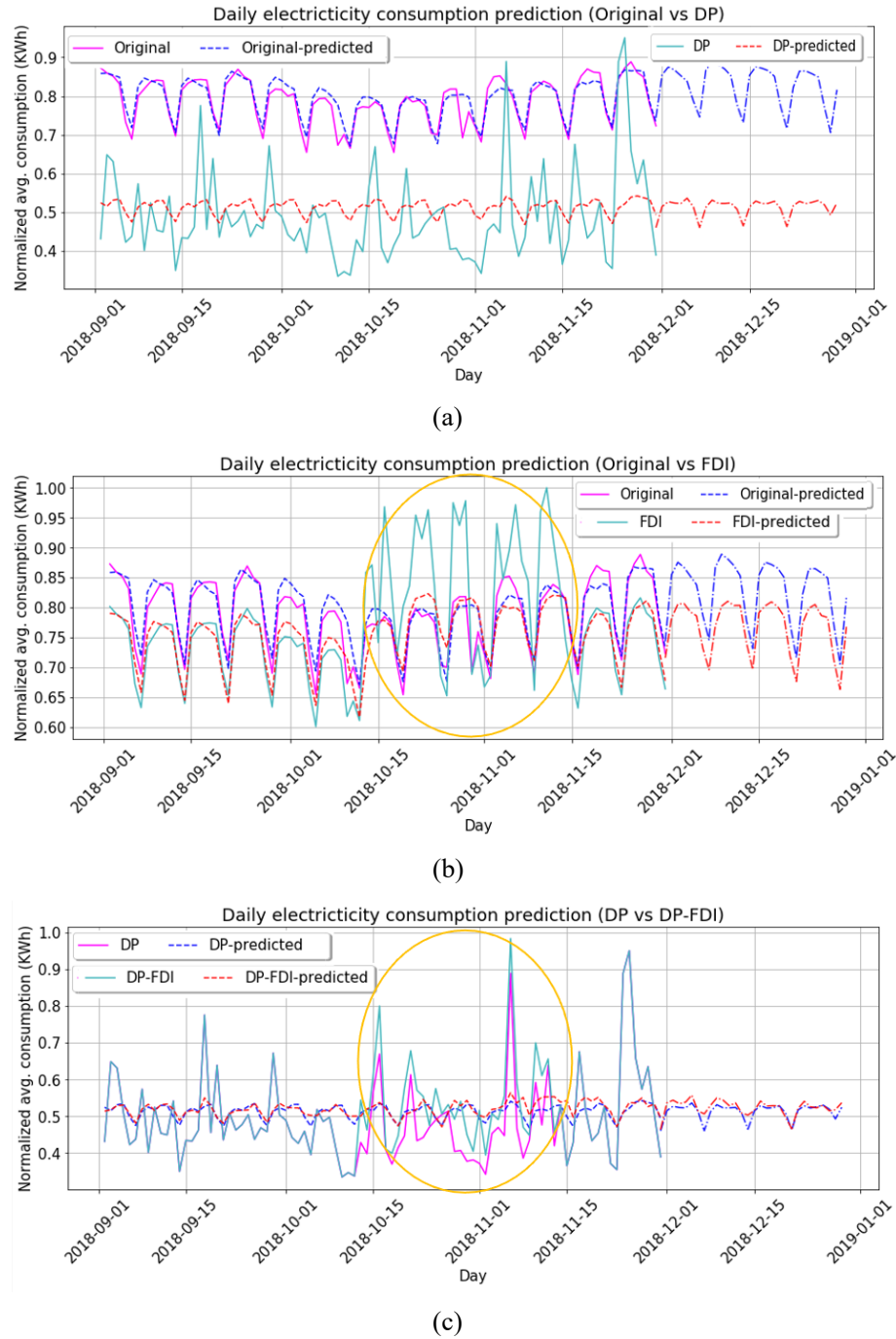


Figure 3.6: QoS analysis of a DP-based system under attack scenario. (a) ‘DP-predicted’ vary from the ‘original-predicted’ due to privacy cost (b) The attack impact on the prediction accuracy is negligible for a short-time FDI attack (circled in the graph) (c) The ‘DP-FDI-predicted’ values are very close to the ‘DP-predicted’ values as the security cost is low.

attack scenario. AES-256 is chosen because it is computationally faster than most of the other encryption techniques (e.g. RSA). Also, the anonymization techniques incur more latency than AES-256 encryption over the large dataset. So, for the feasibility analysis of the DP technique under attack scenario, the latency comparison between the DP and AES-256 technique should suffice. Both the passive and the active attacks add a latency burden to the overall system performance. In the case of the FDI attack, the attacker manipulates the data by adding false information. This process takes some time and adds latency to data flow speed. For any encryption technique, to avoid detection, the attacker needs to decrypt the data, inject false data and again encrypt the falsified data with the cryptographic key. However, for the DP technique, the attacker only injects false data and remains undetected. As a consequence, the DP mechanism does not add as much latency as AES-256. Our experimental evaluation also supports this latency comparison under the FDI attack scenario. We find that the DP technique is approximately 121.49 times faster (on average) than the AES-256 in the same adversarial setting (i.e., the computational time of DP is only 0.0016 seconds whereas the computational time of AES-256 is 0.1944).

3.5 Summary

The differential privacy (DP) can be exploited as a tool for conducting false data injection attacks. To overcome this challenge, only a handful of research has been carried out till this point that mostly focuses on developing a defense mechanism based on the anomaly detection schemes in DP-based cyber-physical energy systems. However, most of the existing anomaly detectors cannot detect the bad data added by exploiting the DP mechanism in the power grid. We address this research gap in this chapter. Subsequently, we formulate and develop the defense strategy as a part

of the design process. We propose a provable correlation among the factors affecting the DP mechanism which enables the defender to design the DP-based system as fault-tolerant against the false data injection attacks. To experimentally evaluate and prove the effectiveness of the proposed correlation, we simulate the false data injection attack in a DP-based metering network. The evaluation indicates that the attack can be minimized using the proposed correlation in the design process. The quality-of-service analysis of the DP-based cyber-physical energy systems indicates the applicability of the DP mechanism in many non-critical operations. Furthermore, the feasibility analysis of the DP mechanism in the cyber-physical energy system domain infers that the DP technique is feasible over other privacy-preserving mechanisms in terms of computational overheads.

Chapter 4

Conclusion and Future Work

4.1 Conclusion

In this thesis, we determine the key factors and correlations among the privacy, security, and utility requirements of grid networks to ensure effective inter-and intra-actions within physical layer equipment (e.g., distributed energy resources (DERs), intelligent electronic devices (IEDs), etc.). We conduct a comprehensive analysis of the existing consensus mechanisms in blockchain-enabled smart grids while pointing out the potential research gap. We develop a practical and effective consensus mechanism for a private and permissioned blockchain-enabled SCADA system. Our novel mining node selection strategy reduces the security vulnerabilities of the traditional mining node selection processes. Moreover, we bridge the common and popular industrial control system (ICS) protocols, Distributed Network Protocol 3 (DNP3) with the blockchain network to ensure smooth operation.

In addition, we develop differential privacy (DP)-enabled strategies to achieve data security, privacy, and utility requirements of the power system network under

adversarial settings. Specifically, we aim to analyze and develop a provable correlation between privacy loss and other DP parameters considering the variations of attacks and their impacts along with DP constraints. This will enable power grid designers to develop, design and employ DP-based fault-tolerant models in data-driven power grid operation and control. Furthermore, we conduct feasibility and QoS analysis of the DP mechanism and the grid to achieve certified robustness. Feasibility analysis of the privacy measure provides an assessment of the practicability of the DP mechanism in grid operation and warns the operators about the possible failures and incoming attacks on physical layer operations. QoS is analyzed in the power grid in terms of data accuracy, computational overhead, and resource utilization.

4.2 Published Research Articles

To this end, we have published our research results described in this thesis in two IEEE conferences: (1) IEEE IEMTRONICS 2020, Venue: Vancouver, Canada, Conference date: 9th-12th September 2020, and (2) IEEE CNS 2021, Venue: Virtual, Conference date: 4th-6th October 2021. The articles are as follows:

- M. T. Hossain, S. Badsha and H. Shen, "PoRCH: A Novel Consensus Mechanism for Blockchain-Enabled Future SCADA Systems in Smart Grids and Industry 4.0," 2020 IEEE International IoT, Electronics and Mechatronics Conference (IEMTRONICS), 2020, pp. 1-7, doi: 10.1109/IEMTRONICS51293.2020.9216438
- M. T. Hossain, S. Badsha and H. Shen, "Privacy, Security, and Utility Analysis of Differentially Private CPES Data," 2021 IEEE Conference on Communications and Network Security (CNS), 2021, pp. 65-73, doi: 10.1109/CNS53000.2021.9705022.

4.3 Future Work

Blockchain and differential privacy has the potential to preserve data privacy in other fields also such as edge computing and cloud computing [85, 86]. We hope that this research will contribute to those fields too. Continuing this line of research, we hope to analyze the impact of distributed active attacks on the DP-based system. In particular, we want to investigate the impact of distributed FDI attacks on personalized-privacy-aware grid architecture. Additionally, we hope to extend our model from a game-theoretic viewpoint where multiple defenders and attackers play dynamic strategies in a repeated manner.

Bibliography

- [1] Z. Guan, G. Si, X. Zhang, L. Wu, N. Guizani, X. Du, and Y. Ma, “Privacy-preserving and efficient aggregation based on blockchain for power grid communications in smart communities,” *IEEE Communications Magazine*, vol. 56, no. 7, pp. 82–88, 2018.
- [2] S. Han, S. Zhao, Q. Li, C.-H. Ju, and W. Zhou, “Ppm-hda: privacy-preserving and multifunctional health data aggregation with fault tolerance,” *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 9, pp. 1940–1955, 2015.
- [3] Z. Shi, R. Sun, R. Lu, L. Chen, J. Chen, and X. S. Shen, “Diverse grouping-based aggregation protocol with error detection for smart grid communications,” *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2856–2868, 2015.
- [4] D. U. Case, “Analysis of the cyber attack on the ukrainian power grid,” *Electricity Information Sharing and Analysis Center (E-ISAC)*, vol. 388, 2016.
- [5] Y. Li, Y. Zhou, A. Jolfaei, D. Yu, G. Xu, and X. Zheng, “Privacy-preserving federated learning framework based on chained secure multi-party computing,” *IEEE Internet of Things Journal*, 2020.

- [6] F. Mo and H. Haddadi, “Efficient and private federated learning using tee,” in *EuroSys*, 2019.
- [7] C. Zhang, S. Li, J. Xia, W. Wang, F. Yan, and Y. Liu, “Batchcrypt: Efficient homomorphic encryption for cross-silo federated learning,” in *2020 {USENIX} Annual Technical Conference ({USENIX} {ATC} 20)*, 2020, pp. 493–506.
- [8] M. Mao and H. Xiao, “Blockchain-based technology for industrial control system cypersecurity,” in *2018 International Conference on Network, Communication, Computer Engineering (NCCE 2018)*. Atlantis Press, 2018, pp. 903–907.
- [9] A. Bhattacharjee, S. Badsha, A. R. Shahid, H. Livani, and S. Sengupta, “Blockphasor: A decentralized blockchain framework to enhance security of synchrophasor,” in *2020 IEEE Kansas Power and Energy Conference (KPEC)*. IEEE, 2020, pp. 1–6.
- [10] A. Maw, S. Adepu, and A. Mathur, “Ics-blockops: Blockchain for operational data security in industrial control system,” *Pervasive and Mobile Computing*, vol. 59, p. 101048, 2019.
- [11] P. Nader, P. Honeine, and P. Beuseroy, “ $\{l_p\}$ -norms in one-class classification for intrusion detection in scada systems,” *IEEE Transactions on Industrial Informatics*, vol. 10, no. 4, pp. 2308–2317, 2014.
- [12] D. I. Urbina, J. A. Giraldo, A. A. Cardenas, N. O. Tippenhauer, J. Valente, M. Faisal, J. Ruths, R. Candell, and H. Sandberg, “Limiting the impact of stealthy attacks on industrial control systems,” in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 1092–1105.

- [13] M. Brown, C. Barrington-Leigh, and Z. Brown, “Kernel regression for real-time building energy analysis,” *Journal of Building Performance Simulation*, vol. 5, no. 4, pp. 263–276, 2012.
- [14] P. Singh, M. Masud, M. S. Hossain, and A. Kaur, “Blockchain and homomorphic encryption-based privacy-preserving data aggregation model in smart grid,” *Computers & Electrical Engineering*, vol. 93, p. 107209, 2021.
- [15] A. Bhattacharjee, S. Badsha, and S. Sengupta, “Personalized privacy preservation for smart grid,” in *2021 IEEE International Smart Cities Conference (ISC2)*. IEEE, 2021, pp. 1–7.
- [16] Z. Zheng, T. Wang, A. K. Bashir, M. Alazab, S. Mumtaz, and X. Wang, “A decentralized mechanism based on differential privacy for privacy-preserving computation in smart grid,” *IEEE Transactions on Computers*, 2021.
- [17] S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system,” 2008.
- [18] C. Dwork, F. McSherry, K. Nissim, and A. Smith, “Calibrating noise to sensitivity in private data analysis,” in *Theory of cryptography conference*. Springer, 2006, pp. 265–284.
- [19] I. Vakiliinia, S. Vakiliinia, S. Badsha, E. Arslan, and S. Sengupta, “Pooling approach for task allocation in the blockchain based decentralized storage network,” in *2019 15th International Conference on Network and Service Management (CNSM)*. IEEE, 2019, pp. 1–6.
- [20] S. Kudva, R. Norderhaug, S. Badsha, S. Sengupta, and A. Kayes, “Pebers: Practical ethereum blockchain based efficient ride hailing service,” in *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIOT)*. IEEE, 2020, pp. 422–428.

- [21] S. Badsha, I. Vakilinea, and S. Sengupta, “Blocynfo-share: Blockchain based cybersecurity information sharing with fine grained access control,” in *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2020, pp. 0317–0323.
- [22] I. Vakilinea, S. Badsha, and S. Sengupta, “Crowdfunding the insurance of a cyber-product using blockchain,” in *2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. IEEE, 2018, pp. 964–970.
- [23] J. Tang, A. Korolova, X. Bai, X. Wang, and X. Wang, “Privacy loss in apple’s implementation of differential privacy on macos 10.12,” *arXiv preprint arXiv:1709.02753*, 2017.
- [24] Ú. Erlingsson, V. Pihur, and A. Korolova, “Rappor: Randomized aggregatable privacy-preserving ordinal response,” in *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, 2014, pp. 1054–1067.
- [25] J. Near, “Differential privacy at scale: Uber and berkeley collaboration,” in *Enigma 2018 (Enigma 2018)*, 2018.
- [26] J. Allen, B. Ding, J. Kulkarni, H. Nori, O. Ohrimenko, and S. Yekhanin, “An algorithmic framework for differentially private data analysis on trusted processors,” *Advances in Neural Information Processing Systems*, vol. 32, 2019.
- [27] L. Sweeney, “Simple demographics often identify people uniquely,” *Health (San Francisco)*, vol. 671, no. 2000, pp. 1–34, 2000.
- [28] A. Narayanan and V. Shmatikov, “Robust de-anonymization of large sparse datasets,” in *2008 IEEE Symposium on Security and Privacy (sp 2008)*. IEEE, 2008, pp. 111–125.

- [29] P. Ohm, “Broken promises of privacy: Responding to the surprising failure of anonymization,” *UCLA l. Rev.*, vol. 57, p. 1701, 2009.
- [30] N. K. Singh and V. Mahajan, “End-user privacy protection scheme from cyber intrusion in smart grid advanced metering infrastructure,” *International Journal of Critical Infrastructure Protection*, p. 100410, 2021.
- [31] P. Kumar, Y. Lin, G. Bai, A. Paverd, J. S. Dong, and A. Martin, “Smart grid metering networks: A survey on security, privacy and open research issues,” *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2886–2927, 2019.
- [32] D. Li, Q. Yang, D. An, W. Yu, X. Yang, and X. Fu, “On location privacy-preserving online double auction for electric vehicles in microgrids,” *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 5902–5915, 2018.
- [33] J. He, L. Cai, and X. Guan, “Preserving data-privacy with added noises: Optimal estimation and privacy analysis,” *IEEE Transactions on Information Theory*, vol. 64, no. 8, pp. 5677–5690, 2018.
- [34] Y. Liu, W. Guo, C.-I. Fan, L. Chang, and C. Cheng, “A practical privacy-preserving data aggregation (3PDA) scheme for smart grid,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 3, pp. 1767–1774, 2018.
- [35] M. A. Ferrag, L. A. Maglaras, H. Janicke, and J. Jiang, “A survey on privacy-preserving schemes for smart grid communications,” *arXiv preprint arXiv:1611.07722*, 2016.
- [36] R. Lu, X. Liang, X. Li, X. Lin, and X. Shen, “EPPA: An efficient and privacy-preserving aggregation scheme for secure smart grid communications,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 9, pp. 1621–1631, 2012.

- [37] Z. Guan, J. Li, L. Wu, Y. Zhang, J. Wu, and X. Du, "Achieving efficient and secure data acquisition for cloud-supported internet of things in smart grid," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1934–1944, 2017.
- [38] Z. Guan, X. Zhou, P. Liu, L. Wu, and W. Yang, "A blockchain based dual side privacy preserving multi party computation scheme for edge enabled smart grid," *IEEE Internet of Things Journal*, 2021.
- [39] M. U. Hassan, M. H. Rehmani, and J. Chen, "Differential privacy techniques for cyber physical systems: a survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 746–789, 2019.
- [40] R. Lu, H. Zhu, X. Liu, J. K. Liu, and J. Shao, "Toward efficient and privacy-preserving computing in big data era," *IEEE Network*, vol. 28, no. 4, pp. 46–50, 2014.
- [41] P. Barbosa, A. Brito, and H. Almeida, "A technique to provide differential privacy for appliance usage in smart metering," *Information Sciences*, vol. 370, pp. 355–367, 2016.
- [42] R. Colelli, C. Foglietta, R. Fusacchia, S. Panzieri, and F. Pascucci, "Blockchain application in simulated environment for cyber-physical systems security," in *2021 IEEE 19th International Conference on Industrial Informatics (INDIN)*. IEEE, 2021, pp. 1–7.
- [43] A. S. Sani, D. Yuan, K. Mens, and Z. Y. Dong, "Diacs: A blockchain-based model for systematic data integrity assessment and control," in *2021 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, 2021, pp. 1–5.

- [44] Z. Zheng, J. Pan, and L. Cai, "Lightweight blockchain consensus protocols for vehicular social networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 6, pp. 5736–5748, 2020.
- [45] D. Puthal, S. P. Mohanty, V. P. Yanambaka, and E. Kougianos, "Poah: A novel consensus algorithm for fast scalable private blockchain for large-scale iot frameworks," *arXiv preprint arXiv:2001.07297*, 2020.
- [46] H. Sandberg, G. Dán, and R. Thobaben, "Differentially private state estimation in distribution networks with smart meters," in *2015 54th IEEE conference on decision and control (CDC)*. IEEE, 2015, pp. 4492–4498.
- [47] B. Pinte, M. Quinlan, and K. Reinhard, "Low voltage micro-phasor measurement unit (μ PMU)," in *2015 IEEE Power and Energy Conference at Illinois (PECI)*. IEEE, 2015, pp. 1–4.
- [48] F. Fioretto, T. W. Mak, and P. Van Hentenryck, "Differential privacy for power grid obfuscation," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1356–1366, 2019.
- [49] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security (TISSEC)*, vol. 14, no. 1, pp. 1–33, 2011.
- [50] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics*. PMLR, 2017, pp. 1273–1282.
- [51] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, "Blockchain and federated learning for privacy-preserved data sharing in industrial IoT," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4177–4186, 2019.

- [52] H. Cao, S. Liu, R. Zhao, and X. Xiong, "IFed: A novel federated learning framework for local differential privacy in Power Internet of Things," *International Journal of Distributed Sensor Networks*, vol. 16, no. 5, p. 1550147720919698, 2020.
- [53] R. C. Geyer, T. Klein, and M. Nabi, "Differentially private federated learning: A client level perspective," *arXiv preprint arXiv:1712.07557*, 2017.
- [54] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, "Differentially private asynchronous federated learning for mobile edge computing in urban informatics," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 2134–2143, 2019.
- [55] A. Bhowmick, J. Duchi, J. Freudiger, G. Kapoor, and R. Rogers, "Protection against reconstruction and its applications in private federated learning," *arXiv preprint arXiv:1812.00984*, 2018.
- [56] Y. Zhao, J. Zhao, M. Yang, T. Wang, N. Wang, L. Lyu, D. Niyato, and K.-Y. Lam, "Local differential privacy based federated learning for internet of things," *IEEE Internet of Things Journal*, 2020.
- [57] R. Deng, G. Xiao, R. Lu, H. Liang, and A. V. Vasilakos, "False data injection on state estimation in power systems—attacks, impacts, and defense: A survey," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 411–423, 2016.
- [58] Q. Yang, L. Jiang, W. Hao, B. Zhou, P. Yang, and Z. Lv, "PMU placement in electric transmission networks for reliable state estimation against false data injection attacks," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1978–1986, 2017.

- [59] M. A. Rahman and H. Mohsenian-Rad, "False data injection attacks against nonlinear state estimation in smart power grids," in *2013 IEEE Power & Energy Society General Meeting*. IEEE, 2013, pp. 1–5.
- [60] L. Xie, Y. Mo, and B. Sinopoli, "False data injection attacks in electricity markets," in *2010 First IEEE International Conference on Smart Grid Communications*. IEEE, 2010, pp. 226–231.
- [61] J. Giraldo, A. A. Cardenas, and M. Kantarcioglu, "Security vs. privacy: How integrity attacks can be masked by the noise of differential privacy," in *2017 American Control Conference (ACC)*. IEEE, 2017, pp. 1679–1684.
- [62] F. Farokhi and P. M. Esfahani, "Security versus privacy," in *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 7101–7106.
- [63] J. Giraldo, A. Cardenas, M. Kantarcioglu, and J. Katz, "Adversarial classification under differential privacy," in *Network and Distributed Systems Security (NDSS) Symposium 2020*, 2020.
- [64] S. R. Pokhrel, Y. Qu, and L. Gao, "QoS-aware personalized privacy with multipath TCP for industrial IoT: Analysis and design," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 4849–4861, 2020.
- [65] Y.-H. Jeon, "QoS requirements for the smart grid communications system," *International Journal of Computer Science and Network Security*, vol. 11, no. 3, pp. 86–94, 2011.
- [66] M. Iansiti and K. Lakhani, "The truth about blockchain:," *Harvard business review*, vol. 95, pp. 118–127, 01 2017.
- [67] "Overview of dnp3 protocol," (Accessed on 03/16/2022). [Online]. Available: <https://www.dnp.org/About/Overview-of-DNP3-Protocol>

- [68] Y. Zhao, “Optimizing hash strategy to avoid birthday attack,” in *Journal of Physics: Conference Series*, vol. 1486, no. 3. IOP Publishing, 2020, p. 032004.
- [69] C. S. Gao, *Classification with Hash Collision Networks*. University of California, Los Angeles, 2018.
- [70] D. Ongaro and J. Ousterhout, “In search of an understandable consensus algorithm,” in *2014 USENIX Annual Technical Conference (Usenix ATC 14)*, 2014, pp. 305–319.
- [71] M. B. Mollah, J. Zhao, D. Niyato, K.-Y. Lam, X. Zhang, A. M. Ghias, L. H. Koh, and L. Yang, “Blockchain for future smart grid: A comprehensive survey,” *IEEE Internet of Things Journal*, vol. 8, no. 1, pp. 18–43, 2020.
- [72] K. Loji, I. E. Davidson, and R. Tiako, “Voltage profile and power losses analysis on a modified ieeec 9-bus system with pv penetration at the distribution ends,” in *2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA)*. IEEE, 2019, pp. 703–708.
- [73] R. P. Naik and N. T. Courtois, “Optimising the sha256 hashing algorithm for faster and more efficient bitcoin mining,” *MSc Information Security Department of Computer Science UCL*, pp. 1–65, 2013.
- [74] A.-R. Sadeghi, C. Wachsmann, and M. Waidner, “Security and privacy challenges in industrial internet of things,” in *2015 52nd ACM/EDAC/IEEE Design Automation Conference (DAC)*. IEEE, 2015, pp. 1–6.
- [75] C.-C. Sun, C.-C. Liu, and J. Xie, “Cyber-physical system security of a power grid: State-of-the-art,” *Electronics*, vol. 5, no. 3, p. 40, 2016.

- [76] R. F. Berriel, A. T. Lopes, A. Rodrigues, F. M. Varejao, and T. Oliveira-Santos, “Monthly energy consumption forecast: A deep learning approach,” in *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2017, pp. 4283–4290.
- [77] T. Liu, Z. Tan, C. Xu, H. Chen, and Z. Li, “Study on deep reinforcement learning techniques for building energy consumption forecasting,” *Energy and Buildings*, vol. 208, p. 109675, 2020.
- [78] J. Giraldo, E. Sarkar, A. A. Cardenas, M. Maniatakos, and M. Kantarcioglu, “Security and privacy in cyber-physical systems: A survey of surveys,” *IEEE Design & Test*, vol. 34, no. 4, pp. 7–17, 2017.
- [79] A. Kelarev, X. Yi, S. Badsha, X. Yang, L. Rylands, and J. Seberry, “A multi-stage protocol for aggregated queries in distributed cloud databases with privacy protection,” *Future Generation Computer Systems*, vol. 90, pp. 368–380, 2019.
- [80] M. Lecuyer, V. Atlidakis, R. Geambasu, D. Hsu, and S. Jana, “Certified robustness to adversarial examples with differential privacy,” in *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2019, pp. 656–672.
- [81] J. Cohen, E. Rosenfeld, and Z. Kolter, “Certified adversarial robustness via randomized smoothing,” in *International Conference on Machine Learning*. PMLR, 2019, pp. 1310–1320.
- [82] M. Pignati, M. Popovic, S. Barreto, R. Cherkaoui, G. D. Flores, J.-Y. Le Boudec, M. Mohiuddin, M. Paolone, P. Romano, S. Sarri *et al.*, “Real-time state estimation of the epfl-campus medium-voltage grid by using pmus,” in *2015 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, 2015, pp. 1–5.

- [83] Y. Yang, K. McLaughlin, T. Littler, S. Sezer, E. G. Im, Z. Yao, B. Pranggono, and H. Wang, “Man-in-the-middle attack test-bed investigating cyber-security vulnerabilities in smart grid scada systems,” 2012.
- [84] W. Yu, D. An, D. Griffith, Q. Yang, and G. Xu, “Towards statistical modeling and machine learning based energy usage forecasting in smart grid,” *ACM SIGAPP Applied Computing Review*, vol. 15, no. 1, pp. 6–16, 2015.
- [85] M. Du, K. Wang, Z. Xia, and Y. Zhang, “Differential privacy preserving of training model in wireless big data with edge computing,” *IEEE transactions on big data*, vol. 6, no. 2, pp. 283–295, 2018.
- [86] T. Wang, Y. Mei, W. Jia, X. Zheng, G. Wang, and M. Xie, “Edge-based differential privacy computing for sensor–cloud systems,” *Journal of Parallel and Distributed computing*, vol. 136, pp. 75–85, 2020.