7-2023

A Privacy Safeguarding Framework for the Smart Grid

Gaurav Shivaji Wagh
gauravwagh@mail.rit.edu

A Privacy Safeguarding Framework for the Smart Grid

by

Gaurav Shivaji Wagh

A dissertation submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in Computing and Information Sciences

B. Thomas Golisano College of Computing and
Information Sciences

Rochester Institute of Technology
Rochester, New York
July 2023

A Privacy Safeguarding Framework for the Smart Grid

by

Gaurav Shivaji Wagh

Committee Approval:

We, the undersigned committee members, certify that we have advised and/or supervised the candidate on the work described in this dissertation. We further certify that we have reviewed the dissertation manuscript and approve it in partial fulfillment of the requirements of the degree of Doctor of Philosophy in Computing and Information Sciences.

Dr. Sumita Mishra Date:
Dissertation Advisor

Dr. Stanisław Radziszowski Date:
Dissertation Committee Member

Dr. Anurag Agarwal Date:
Dissertation Committee Member

Dr. Andres Kwasinski Date:
Dissertation Committee Member

Dr. Jennifer Schneider Date:
Dissertation Defense Chairperson

Certified by:

Dr. Pengcheng Shi Date:
Ph.D. Program Director, Computing and Information Sciences

A Privacy Safeguarding Framework for the Smart Grid

by
Gaurav Shivaji Wagh

Submitted to the
B. Thomas Golisano College of Computing and Information Sciences Ph.D. Program in
Computing and Information Sciences
in partial fulfillment of the requirements for the
Doctor of Philosophy Degree
at the Rochester Institute of Technology

Abstract

The smart grid is an outcome of integrating communication technologies with traditional electrical systems. This enables the collection of granular metering data from the customer domain for providing grid and billing functionalities. However, the data collection process exposes the grid to various cyberattacks, posing a significant threat to customer privacy. This is a critical concern for the smart grid community and has hindered the global adoption of the smart grid technology. Although aggregation-based frameworks show promise for sharing metering data with the Electrical Service Provider while maintaining customer privacy, existing aggregation-based frameworks have several limitations. Some of these limitations include a high computational overhead on resource-constrained smart meters, susceptibility to single points of compromise due to dependency on a centralized entity, lack of support for dynamic billing functionality, and the absence of integrity verification capabilities for spatial and temporal metering data.

To address the aforementioned limitations, we propose a distributed privacy-preserving framework for the smart grid that utilizes secret sharing, commitments, and secure multiparty computation. The framework consists of smart meters employing secret sharing and commitments to outsource their data to multiple aggregating entities, known as Dedicated Aggregators. These Dedicated Aggregators utilize secure multiparty computation to perform spatial aggregation in a privacy-preserving manner and report the aggregated readings to the Electrical Service Provider. By offloading most computations to the Dedicated Aggregators, our framework ensures that it remains lightweight for the smart meters. The introduction of multiple Dedicated Aggregators also aids in mitigating concerns associated with single points of compromise. Additionally, we have adapted the framework to support temporal aggregation, enabling dynamic billing functionalities while preserving customer privacy. The temporal aggregation process is integrated with the spatial aggregation process, thus imposing no additional computational overhead on the smart meters. The framework is designed to cater to both semi-honest and malicious adversarial settings, and works even in the presence of a majority of dishonest Dedicated Aggregators. In the event that some Dedicated Aggregators deviate from the normal execution of computing spatial and/or temporal aggregation by making modifications to metering data, the Electrical Service Provider can detect and respond to such modifications in a privacy-preserving manner.

This dissertation presents our proposed framework and conducts a comprehensive analysis of it under different configurations. We develop a proof of concept to illustrate the practicality of implementing our framework in a real-world setting. We also compare its performance with other related works in the literature, evaluating the end-to-end delay for spatial aggregation. Additionally, we analyze the computational overhead on the smart meters in an embedded environment for various framework designs. The resilience of our proposed framework is analyzed against security and privacy threats. Finally, we identify future research directions to extend the capabilities of our framework.

Acknowledgments

I want to thank the following people who supported me during my research at the Rochester Institute of Technology.

Firstly, I would like to thank my advisor, Dr. Sumita Mishra, for her constant support and guidance. Her expertise, constructive feedback, valuable insights, patience, and empathy have shaped this research.

I would also like to thank my dissertation committee, Dr. Stanisław Radziszowski, Dr. Anurag Agarwal, and Dr. Andres Kwasinski. Their useful insights, opinions, and constructive feedback enhanced my research. I am thankful for their time and commitment to reviewing my research and providing the appropriate direction in problem-solving, writing, and presenting my research to a diverse audience.

I want to express my sincere gratitude to Jennifer Schneider for generously dedicating her valuable time to serve as my dissertation defense chair. I thank Dr. Pengcheng Shi (Director of Computing and Information Sciences) for supporting my research journey. I also want to express my special thanks to Min-Hong Fu, who helped me with the administrative tasks related to my Ph.D., allowing me to focus more on my research work.

Finally, I would like to thank my family and friends for their unwavering support and encouragement throughout my Ph.D. journey.

I dedicate this dissertation to my parents
Shivaji Dharma Wagh
&
Ujwala Shivaji Wagh

Contents

1 Introduction **1**

1.1 Background 1

1.2 Threats to Smart Grid 3

1.3 Motivation and Challenges of Smart Grid 4

1.4 Organizational Structure 8

1.5 Summary 9

2 Related Works **10**

2.1 Battery-based Frameworks 11

2.2 Distortion-based Frameworks 12

2.3 Aggregation-based Frameworks 13

2.3.1 In-Network Aggregation-based Frameworks 14

2.3.2 Centralized Aggregation-based Frameworks 16

2.3.3 Distributed Aggregation-based Frameworks 17

2.4 Research Gaps 19

2.5 Summary 20

3 Building Blocks 21

3.1 Secret Sharing Scheme 21

3.2 Commitment-based Schemes 25

3.3 Secure Multiparty Computation 28

3.4 Homomorphic Encryption-based Schemes 30

3.5 Summary 33

4 A Privacy Safeguarding Framework: Semi-Honest Setting 34

4.1 Introduction 34

4.2 System Model 37

4.2.1 Architectural Model 37

4.2.2 Threat Model 39

4.2.3 Billing Model 39

4.3 Cryptographic Prerequisites 41

4.3.1 Configuration of Shamir's Secret Sharing Scheme 41

4.3.2 Configuration of Secure Multiparty Computation 42

4.4 Proposed Framework 43

4.4.1 Initialization Phase 43

4.4.2 Share Creation Phase 45

4.4.3 Spatial Aggregation Phase 46

4.4.4 Temporal Aggregation Phase 46

4.4.5 Billing Phase 49

4.5 Overall Framework Design 50

4.6 Summary 54

5 A Privacy Safeguarding Framework: Malicious Setting 55

5.1 Introduction 55

5.2 System Model 59

5.2.1 Architectural Model 59

5.2.2 Threat Model 61

5.2.3 Billing Model 62

5.3 Cryptographic Prerequisites 62

5.3.1 Configuration of Commitment-based Scheme 62

5.3.2 Configuration of Shamir's Secret Sharing Scheme (SSS) with a constraint . . 63

5.3.3 Configuration of Secure Multiparty Computation 64

5.4 Proposed Framework 65

5.4.1 Initialization Phase 67

5.4.2 Commitment and Share Creation Phase 68

5.4.3 Spatial Aggregation Phase 69

5.4.4 Integrity Verification Phase: Spatial Aggregation 69

5.4.5 Temporal Aggregation Phase 70

5.4.6 Integrity Verification Phase: Temporal Aggregation 72

5.4.7 Billing Phase 73

5.5 Overall Framework Design 74

5.6 Summary 78

6 Results and Analysis 79

6.1 Introduction 79

6.2 Experimental Setup 80

6.3 Evaluation Metrics 81

6.3.1 Computational Overhead 81

6.3.2 End-to-End Delay 82

6.3.3 Resilience to Security Attacks 82

6.3.4 Resilience to Privacy Attacks 82

6.4 Performance Analysis 82

6.4.1 Computational Overhead on Smart Meters due to Spatial Aggregation in a Semi-Honest Setting 83

6.4.2 Computational Overhead on Dedicated Aggregators due to Spatial Aggregation in a Semi-Honest Setting 86

6.4.3 Overall Computational Overhead associated while reporting Spatially Aggregated Reading to the Electrical Service Provider in a Semi-Honest Setting . . 87

6.4.4 Computational Overhead on Smart Meters due to Temporal Aggregation in a Semi-Honest Setting 89

6.4.5 Computational Overhead on Dedicated Aggregators due to Temporal Aggregation in a Semi-Honest Setting 89

6.4.6 Overall Computational Overhead associated while reporting Temporally Aggregated Reading to the Electrical Service Provider in a Semi-Honest Setting 91

6.4.7 Computational Overhead on Smart Meters due to Spatial Aggregation in a Malicious Setting 92

6.4.8 Computational Overhead on Dedicated Aggregators due to Spatial Aggregation in a Malicious Setting . . . 93

6.4.9 Overall Computational Overhead associated while reporting Spatially Aggregated Reading to the Electrical Service Provider in a Malicious Setting . . . 94

6.4.10 Computational Overhead on Smart Meters due to Temporal Aggregation in a Malicious Setting . . . 96

6.4.11 Computational Overhead on Dedicated Aggregators due to Temporal Aggregation in a Malicious Setting . . . 96

6.4.12 Overall Computational Overhead associated while reporting Temporally Aggregated Reading to the Electrical Service Provider in a Malicious Setting . . 97

6.4.13 Comparative Analysis of Computational Overhead on Smart Meter due to Spatial Aggregation . . . 99

6.4.14 Comparative Analysis of Computational Overhead on Smart Meters due to Spatio-Temporal Aggregation . . . 100

6.4.15 Comparative Analysis of End-to-End Delay due to Spatial Aggregation . . . 101

6.5 Resilience Evaluation . . . 102

6.5.1 Resilience to Security Attacks . . . 102

6.5.2 Resilience to Privacy Attacks . . . 105

6.6 Qualitative Analysis . . . 105

6.7 Summary . . . 108

7 Conclusion and Future Work 109

7.1 Conclusion . . . 109

7.2 Future Work . . . 110

Appendices 123

A First Appendix 124

A.1 Graphical User Interface 124

A.2 UMass Dataset 126

List of Figures

1.1 NIST Reference Model 2

1.2 Electric Load Profile of a Customer 3

1.3 Our Proposed Framework 6

2.1 Overview of Privacy-Preserving Frameworks 10

2.2 Battery-based Framework 11

2.3 Distortion-based Framework [39] 12

2.4 Spatial and Temporal Aggregation 14

2.5 In-Network Aggregation-based Framework [54] 15

2.6 Centralized Aggregation-based Framework [11] 16

2.7 Distributed Aggregation-based Framework [86] 18

3.1 Example of Shamir's Secret Sharing Scheme 24

3.2 Example of Commitment-based Scheme 26

3.3 Example of Secure Multiparty Computation 28

3.4 Example of Homomorphic Encryption-based Scheme 30

4.1 Secure Multiparty Computation (Spatial Aggregation) 43

4.2 Flowchart of our Privacy Safeguarding Framework in a Semi-Honest Setting 51

4.3 A Detailed Workflow of our Privacy Safeguarding Framework in a Semi-Honest Setting 53

5.1 A Privacy Safeguarding Framework: Malicious Setting 59

5.2 Secure Multiparty Computation (Temporal Aggregation) 65

5.3 Flowchart of our Privacy Safeguarding Framework in a Malicious Setting 75

5.4 A Detailed Workflow of our Privacy Safeguarding Framework in Malicious Setting . 77

6.1 Computational Overhead on Smart Meters due to Spatial Aggregation in a Semi-Honest Setting 84

6.2 Comprehensive Analysis of Shamir's Secret Sharing Scheme for Smart Meters in a Semi-Honest Setting 85

6.3 Computational Overhead of updating Spatial Registers (Spatial Aggregation) associated with Dedicated Aggregators in a Semi-Honest Setting 86

6.4 Computational Overhead of reconstructing polynomial related to Spatial Aggregation in a Semi-Honest Setting 87

6.5 Overall Computational Overhead associated while reporting Spatially Aggregated Reading to the Electrical Service Provider in a Semi-Honest Setting 88

6.6 Computational Overhead of updating Temporal Registers (Temporal Aggregation) associated with Dedicated Aggregators in a Semi-Honest Setting 89

6.7 Computational Overhead of reconstructing polynomial related to Temporal Aggregation in a Semi-Honest Setting 90

6.8 Overall Computational Overhead associated while reporting Temporally Aggregated Reading to the Electrical Service Provider in a Semi-Honest Setting 91

6.9 Computational Overhead on a Smart Meter in a Malicious Setting 92

6.10 Computational Overhead of updating Spatial Registers and Spatial Commitment Registers (Spatial Aggregation) associated with Dedicated Aggregators in a Malicious Setting 93

6.11 Overall Computational Overhead associated while reporting Spatially Aggregated Reading to the Electrical Service Provider in a Malicious Setting 95

6.12 Computational Overhead on Dedicated Aggregators due to Temporal Aggregation in a Malicious Setting 96

6.13 Overall Computational Overhead associated while reporting Temporally Aggregated Reading to the Electrical Service Provider in a Malicious Setting 98

6.14 Comparative Analysis of Computational Overhead on Smart Meter due to Spatial Aggregation 99

6.15 Comparative Analysis of Computational Overhead on Smart Meters due to Spatio-Temporal Aggregation 100

6.16 Comparative Analysis of End-to-End Delay due to Spatial Aggregation 101

6.17 A Generalized Flowchart to demonstrate Resilience of our Proposed Framework to Security Attacks 104

A.1 Graphical User Interface of our Proposed Work 125

A.2 Visualization of missing data of UMass Dataset using Bar Plot 126

A.3 Visualization of null data of UMass Dataset using Heatmap 127

List of Tables

1.1 Traditional Electrical Grid versus Smart Grid 2

3.1 Table of notations for Shamir's Secret Sharing Scheme 22

3.2 Table of notations for Commitment-based Schemes 26

3.3 Table of notations for Paillier Encryption Scheme 32

4.1 Table of notations: A Privacy Safeguarding Framework in a Semi-Honest Setting . . 44

5.1 Table of notations: A Privacy Safeguarding Framework in a Malicious Setting . . . 66

6.1 Comparison of Aggregation-based Privacy-Preserving Frameworks 107

Chapter 1

Introduction

1.1 Background

The traditional or legacy electrical grid is being transformed into the smart grid by integrating state-of-the-art communication technologies [81, 99]. The varying demands of the 21^{st} century and the events such as the North East Blackout and Hurricane Sandy were the major motivating factors for this transformation [81, 99]. These events exposed the limitations of the traditional electric grid, such as a lack of resiliency and real-time monitoring capabilities. The evolution of the smart grid allows the Electrical Service Provider (ESP) to provide added grid functionalities, such as load balancing, demand forecasting, outage management, integration of renewable resources, billing functionalities, and enabling customers to have better control over consumption [19, 81].

The National Institute of Standards and Technology (NIST) has proposed a smart grid reference model to standardize the smart grid and ensure its widespread adoption [32,81]. The smart grid reference model (Fig. 1.1) consists of the following domains: Generation, Transmission, Distribution, Markets, Operations, Customers and Service Provider.

In the customer domain, the traditional (non-smart) meter is being replaced by the smart meter (SM) [32, 81]. The SM is responsible for collecting metering data from the customer domain and reporting it to the ESP, thereby eliminating the need to collect the metering data manually by the ESP. Based on the reporting frequency, the metering data can be categorized into high-frequency and low-frequency metering data. The high-frequency metering data is reported to the ESP every 15 minutes to provide grid functionalities. The low-frequency metering data is reported once a month for billing the customer associated with the SM for its corresponding consumption.

Table 1.1 compares the traditional electrical grid against the smart grid [27]. The main goal of the ESP is to maintain a balance between generation and consumption (demand). This is because if generation exceeds consumption, the excess electricity generated must be stored, which can be expensive. Conversely, if generation is less than consumption, it can lead to an outage.

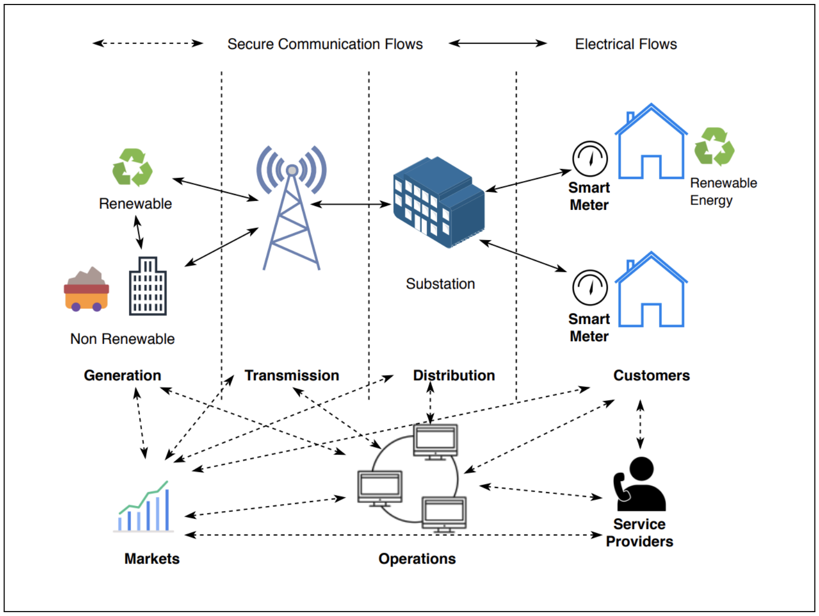


Figure 1.1: NIST Reference Model

Table 1.1: Traditional Electrical Grid versus Smart Grid

Parameters	Traditional Electric Grid	Smart Grid
Machinery	Electric	Digital
Communication	One-way	Two-way
Power Generation	Centralized	Distributed
Monitoring	Manual	Remote
Recovery	Manual	Automatic
Outage Management	Not supported	Adaptive and Islanded
Customized Tariffs	Not supported	Supported
Integration of Renewable Resources	Not supported	Supported

1.2 Threats to Smart Grid

According to the National Institute of Standards and Technology Interagency Report (NISTIR), the smart grid should be resilient against cyberattacks and natural disasters and address customer privacy concerns [32, 81]. Even though the transformation has several benefits, the integration of communication technologies exposes the smart grid to security and privacy attacks on high-frequency metering data [19, 21, 24, 33, 44, 45, 59, 67, 68, 82, 99]. Privacy concerns are a leading obstacle to the large-scale adoption of the smart grid. Studies have shown that customers' behavior patterns can be derived from high-frequency metering data collection. Based on the load profile, criminals can identify suitable times for carrying out nefarious activities. Marketers can identify appliances (as each appliance has a unique load signature) and send targeted advertisements to the customers (target marketing). Figure 1.2 represents the load profile of a specific customer based on the consumption data collected over 24 hours from a Reference Energy Disaggregation DataSet (REDD) using an open-source Non-Intrusive Load Monitoring (NILM)toolkit [9, 50]. The toolkit can disaggregate the energy consumption and helps derive which appliances are used in the household and their corresponding consumption. Privacy attacks can be quite targeted and specific, for example, determination of the Television channels being watched in a household [33]. Private investigators, spies, and reporters can also derive information regarding their potential suspects. The smart metering data reported to the ESP can be subjected to modification (attack on integrity) by malicious adversaries. This can disrupt the grid and billing functionalities, resulting in outages

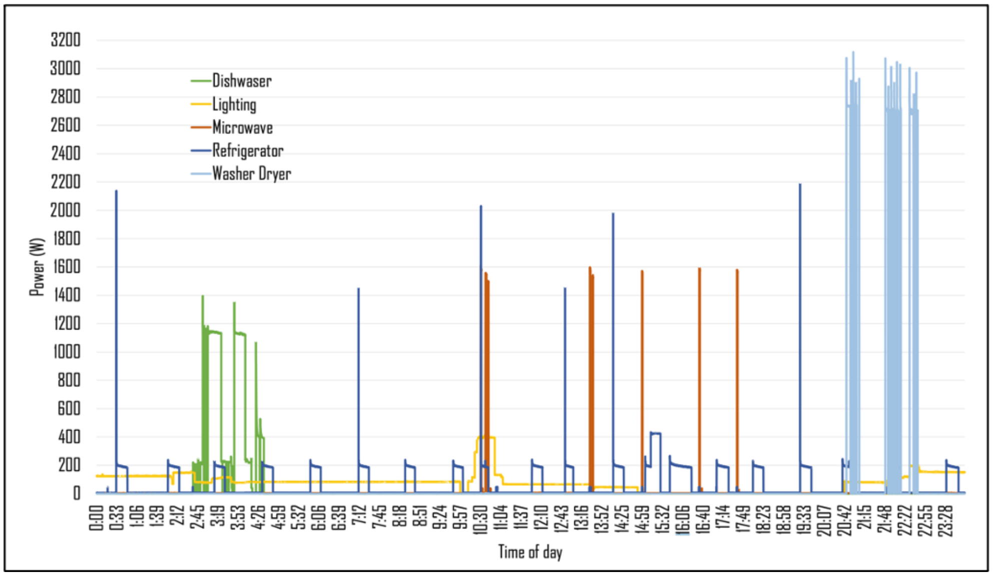


Figure 1.2: Electric Load Profile of a Customer

and eroding customer trust. In order to protect customer privacy and the integrity of metering data, it is important to forward the high-frequency metering data from the SM to the ESP in a secure and privacy-preserving manner.

1.3 Motivation and Challenges of Smart Grid

Various types of privacy-preserving frameworks have been proposed in the literature to address the privacy issues identified in the previous section (Section 1.2). The frameworks can be broadly categorized as: Battery-based Frameworks, Distortion-based Frameworks, and Aggregation-based Frameworks. In this dissertation, our focus will be on aggregation-based frameworks as they closely align with our research work. In aggregation-based frameworks, the aggregated reading is reported to the ESP. Since the reading is aggregated, the ESP cannot link the granular meter reading to a specific SM, thereby preserving the customer's privacy. However, the existing aggregation-based frameworks have at least one of the following limitations:

- High computational overhead on resource-constrained SMs
- Prone to single points of compromise due to dependency on a centralized entity
- Lack of support for dynamic billing integration while preserving customer privacy
- Lack of integrity verification of spatial and/or temporal metering data by considering a semi-honest threat model

In order to ensure the security and reliability of the smart grid, it is important to address the requirements provided by NIST [10, 19, 28, 32, 81, 83, 89, 96]. The requirements can be categorized into following types:

- Security Requirements: The security requirements can be described as follows:
  1. Confidentiality: The metering data should only be accessible to authorized entities.
  2. Integrity: The metering data should be accurate and consistent.
  3. Availability: The metering data is available to the ESP in a timely manner.
- Privacy Requirement: The privacy of the customers associated with smart grid must remain intact.

- Infrastructure Requirements: The infrastructure requirements can be described as follows:

  1. Backward compatibility: The new technology that is been incorporated in the smart grid must be able to support the legacy system components.
  2. Distributed Trust: The trust should be distributed across multiple entities so that the system is resilient to single points of compromise.
  3. Integrate Renewable Resources: The smart grid should be able to integrate and support renewable resources such as solar, wind etc.
  4. Accurate Billing: The smart grid must be able to accurately charge the customers based on their corresponding tariff and consumption.
  5. Lightweight: The framework should be lightweight. It should not have a high computational overhead on resource-constrained SMs.
  6. Resilient: The framework should be resilient to security and privacy attacks.
  7. Implementation Cost: The cost of implementing the framework and/or perform any upgrades should be as minimal as possible.
  8. Redundancy: The framework should have redundant communication and transmission lines so that it is resilient to power outages and/or loss of data.

- Regulatory Requirements: Given the complex nature of the smart grid, where various devices are interconnected, adherence to regulatory requirements is a must. These regulations serve as guidelines to ensure privacy regulations, operational and data protection standards.

In order to address the research gaps in the literature works and taking the requirements provided by NIST into consideration [32,81], we proposed a distributed aggregation-based privacy-preserving framework (Fig. 1.3). The distributed framework aims to focus and address the following NIST requirements: Security Requirements (1,2, and 3), Privacy Requirement and Infrastructure Requirements (2,4,5,6, and 8). The framework consists of three types of entities: SMs, Dedicated Aggregators (DAs), and the ESP. The SMs are responsible for reporting the high-frequency metering data to the ESP via the DAs in a privacy-preserving fashion. The DAs are responsible for performing spatio-temporal aggregation and reporting the result to the ESP. The ESP is responsible for initializing the SMs and the DAs and checking the integrity verification of spatio-temporal metering data in a privacy-preserving manner.

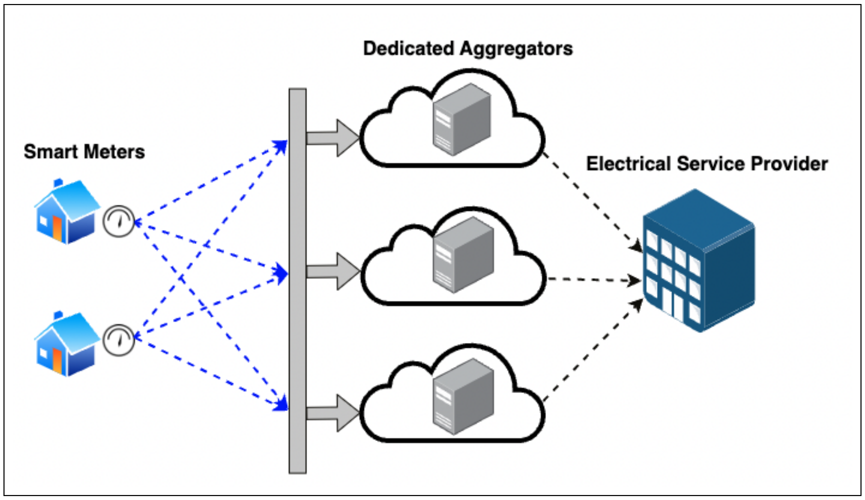


Figure 1.3: Our Proposed Framework

We have outlined the following research questions to achieve the above-mentioned objectives:

- **RQ1: How can smart meters (SMs) send spatial metering data to the Electrical Service Provider (ESP) in a way that protects privacy and ensures efficiency for the SMs themselves?**
  Since spatially aggregated metering data is required for providing grid functionalities, reporting the spatially aggregated reading in a privacy-preserving manner to the ESP is critical. This research question is addressed in **Chapter 4**, where a distributed framework consisting of multiple DAs is proposed. The DAs are responsible for performing spatial aggregation and reporting the spatially aggregated reading to the ESP. Since spatially aggregated reading is reported to the ESP, it cannot be linked to a specific SM, thereby preserving customers' privacy. This work [102], has been published in 2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT 2020). This research led to the next research question, which focused on temporal aggregation.

- **RQ2: How can smart meters (SMs) send temporal metering data to the Electrical Service Provider (ESP) in a way that protects privacy and ensures efficiency for the SMs themselves?**
  Since temporally aggregated metering data is required by the ESP for providing billing functionalities, the bill should be generated without violating customers' privacy. This research question is addressed in **Chapter 4**, where the same set of DAs that perform spatial aggregation are used to perform temporal aggregation. Since the computed bill is sent to the ESP (Flat Rate / Cumulative / Time of Use Billing) for the corresponding customer, deriving high-frequency reading from it is not feasible (desired privacy). This work [103], has been published in 2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm 2020).

- **RQ3: How can the Electrical Service Provider (ESP) incorporate integrity verification of spatially aggregated metering data without violating customer privacy?**
  RQ1 highlighted the importance of spatially aggregated data to the ESP for providing grid functionalities. However, the metering data can be subjected to modification by malicious adversaries. The resulting modification might disrupt the grid functionalities creating an outage. Therefore, it is important to identify such spatially aggregated modifications and take appropriate measures. We extended the previously proposed framework in a malicious setting (dishonest majority of DAs). We integrated a commitment-based scheme to ensure the integrity verification of metering data. The ESP carries out the integrity verification of spatially aggregated metering data in a privacy-preserving fashion. This work [105], has been published in 2022 IEEE International Symposium on Technologies for Homeland Security (HST 2022). This research led to the next research question, which focused on temporal aggregation in a malicious setting.

- **RQ4: How can the Electrical Service Provider (ESP) incorporate integrity verification of temporally aggregated metering data without violating customer privacy?**
  RQ2 highlighted the importance of temporally aggregated data for providing billing functionalities. However, the metering data can be subjected to modification by malicious adversaries. The resulting modification might cause an error in the billing computation, generating an incorrect bill for the customer, and also might erode the customer's trust. Therefore, it is important for the ESP to identify such modifications of temporally aggregated metering data and take appropriate measures. The same set of shares and commitments generated in RQ3 were utilized to provide integrity verification of metering data, resulting in no added computation for the resource-constrained SMs. This work [106], has been published in the 14th ACM International Conference on Future Energy Systems (e-Energy 2023).

- **RQ5: How to develop a practical proof of concept, that assesses computational overhead and end-to-end delay of our proposed framework in comparison to other state-of-the-art frameworks?**
  This research question is addressed in **Chapter 6**, where we implement a practical proof of concept to evaluate the performance of our framework against the relevant literature works. The proof of concept is developed in an embedded and cloud-based environment to evaluate parameters such as SM overhead and end-to-end delay for spatial aggregation. Additionally, we have incorporated smart meter traces from the UMass dataset [75] to mimic a more realistic scenario.

- **RQ6: How do we evaluate the resilience of our proposed framework against potential security attacks?**
  This research question is addressed in **Chapter 6**, in which we consider a malicious adversary (with dishonest majority of DAs) that is capable of launching active attacks, modifying metering data (spatial and/or temporal) with an aim to disrupt grid and billing functionalities. Our proposed framework is capable of detecting such modifications in a privacy-preserving manner.

- **RQ7: How do we evaluate the resilience of our proposed framework against potential privacy attacks?**
  This research question is addressed in **Chapter 6**, in which we consider an adversary that aims to breach the privacy of the customers by linking granular metering data to the corresponding identity of the SM. Our proposed framework is capable of preserving privacy of the customers even if one DA is honest. This work [104], has been published in 2021 IEEE International Conference on Communications (ICC 2021).

1.4 Organizational Structure

Chapter 2: Related Works: This chapter delves into a comprehensive literature survey of related works in the smart grid domain, identifying their contributions and research gaps.

Chapter 3: Building Blocks: This chapter describes the cryptographic preliminaries associated with our proposed framework, as well as those required for comparative frameworks.

Chapter 4: A Privacy Safeguarding Framework: Semi-Honest Setting: This chapter focuses on our proposed framework in a semi-honest setting, where adversaries aim to breach customers' privacy by linking the granular metering data to their identities. We discuss the potential strengths and highlight the limitations of our work in this chapter.

Chapter 5: A Privacy Safeguarding Framework: Malicious Setting: This chapter focuses on our proposed framework within a malicious setting, specifically dealing with a dishonest majority of DAs. In this context, the involved DAs can deviate from their normal execution and launch active attacks, such as modifying metering data, to disrupt grid and billing functionalities. These attacks have the potential to cause electric outages and undermine customers' trust associated with the smart grid.

Chapter 6: Results and Analysis: In this chapter, our aim is to provide insights into the effectiveness of our proposed framework compared to other relevant frameworks identified in Chapter 2 by conducting experiments. Additionally, we analyze our proposed framework against the NIST requirements discussed earlier.

Chapter 7: Conclusion and Future Work: This chapter concludes the dissertation by summarizing the findings, highlighting the impact of the research and discussing potential future directions in the smart grid domain.

1.5 Summary

This chapter provided a high-level overview of the smart grid, followed by the identification of its security and privacy issues. The motivation and challenges of the research were also emphasized, and seven research questions that aligned with the central theme of this dissertation were identified. A road map that outlines the organizational structure of the dissertation was also included.

Chapter 2

Related Works

In the previous chapter (Chapter 1), we introduced the smart grid and highlighted the associated security and privacy issues. In this chapter, we categorize the different privacy-preserving frameworks proposed by researchers to address the privacy issue, followed by identifying the associated research gaps.

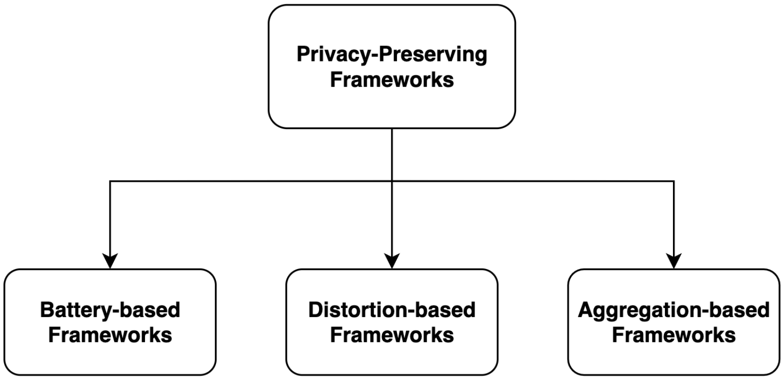

Privacy-Preserving Frameworks
Battery-based Frameworks
Distortion-based Frameworks
Aggregation-based Frameworks

Figure 2.1: Overview of Privacy-Preserving Frameworks

The privacy-preserving works in the literature can be categorized (Fig. 2.1) as follows:

- Battery-based Frameworks
- Distortion-based Frameworks
- Aggregation-based Frameworks

2.1 Battery-based Frameworks

The Battery-based Frameworks attempt to protect customers' privacy by masking the real consumption of the smart meter (SM) through the utilization of a rechargeable battery (Fig. 2.2) [6,29,30,46,47,71,97,98]. Since the Electrical Service Provider (ESP) and the rechargeable battery provide customers with electricity simultaneously, the SM reading does not directly represent the actual consumption, thereby preserving customers' privacy. The battery serves as a reliable source of electricity when other external sources are not functioning as expected. The batteries are also useful in areas with unstable power supply and/or during natural disasters. When the battery can meet the demands of the corresponding household completely without relying on the ESP, there is less metering data transmitted to the ESP. Additionally, the battery can integrate with other renewable resources such as solar, wind, etc.

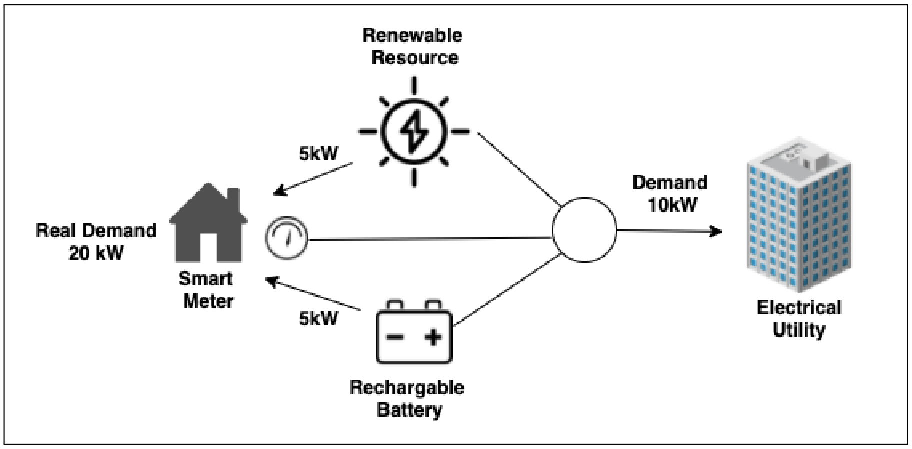


Figure 2.2: Battery-based Framework

In [46], the authors propose a Best Effort technique and a power mixing algorithm to mask the real consumption using a rechargeable battery. They also measure privacy by employing privacy metrics such as relative entropy, correlation/regression analysis, and cluster classification. The framework [46] enables grid and billing functionalities, but the customers' privacy could be breached due to battery capacity limitations. The authors do not provide a cost evaluation analysis regarding privacy and the corresponding battery capacity. As proposed in the paper [71], a Non-Intrusive Load Leveling technique aims to remove the most identifiable features from the high-frequency reading, making it difficult for Non-Intrusive Load Modelling algorithms to derive any useful information. In their study [47], the authors introduce a water-filling algorithm to mask the real consumption of a SM. They also provide a cost analysis for achieving the desired privacy level. In [6], the authors aim to preserve privacy by utilizing a battery and implementing suitable addition or removal of noise.

However, the generation of noise introduces additional load on the battery, significantly affecting its lifespan. To address this, the authors recommend installing an additional battery dedicated to generating noise, which leads to increased installation and maintenance costs. In [98], the authors introduce an Energy Management Unit (EMU) capable of drawing power from three main sources: a renewable resource, a Rechargeable Battery (RB), and the ESP. Whenever there is a request for a load from an electrical appliance in the household, the EMU can fulfill the request from any of the three connected sources. As a result, the SM reading does not directly represent the real household consumption, thereby preserving customers' privacy. In [30], a detailed mathematical analysis is presented, taking into account the physical constraints of the rechargeable battery, with respect to privacy. The battery capacity also plays a key role in protecting customers' privacy.

To summarize, Battery-based Frameworks aim to protect the privacy of the customers but have a high installation cost and maintenance overhead. While these frameworks enable customers to perform power management, there may be conflicts between the battery charging time and the dynamic billing functionality of the smart grid, especially concerning the ESP. The battery also has a limited lifespan, making this technique unsuitable for both customers and the ESP.

2.2 Distortion-based Frameworks

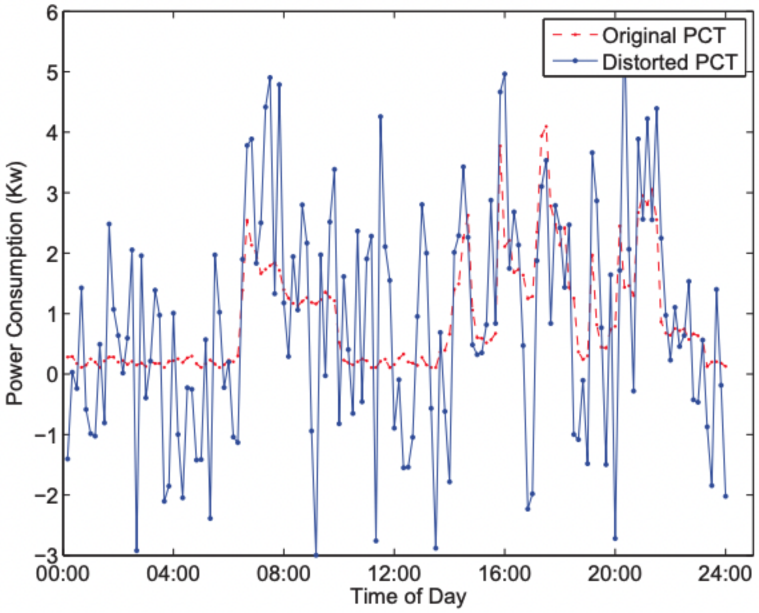


Figure 2.3: Distortion-based Framework [39]

In Distortion-based Frameworks, the SM reading is made obscure to adversaries by adding intentional noise [3, 13, 18, 35, 39, 62, 89, 92, 94]. In [13], the authors propose a solution to preserve customers' privacy by adding noise (normally distributed) to the granular consumption. The noise added by the SMs should have a mean of zero. Adding and removing SMs is easy and requires partial group management. In [3], Laplace's distribution is used to protect customers' privacy. The authors employ a hybrid approach by using noise and partial homomorphic encryption (Paillier) [77]. However, this expects SMs to be interconnected, which increases the computational overhead. In [94], the authors recommend utilizing the geometric distribution to mask the granular reading. The frameworks [3, 94] are prone to single points of compromise due to their dependency on a centralized entity. They assume all SMs operate honestly, as any deviation from their behavior makes overall consumption recovery impossible. In [39], the authors utilize Gaussian noise to obscure the reading from adversaries (Fig. 2.3). While it supports dynamic billing, additional modifications in the customers' domain, such as installing a hardware known as Privacy Component, are required. Similar to the previous frameworks, it is also prone to single points of compromise due to its dependency on a centralized entity. In some cases, prior knowledge of the threshold for the amount of noise to be added is required, as adding excessive noise could make reconstruction of the reading impossible. Hence, there is a trade-off between privacy and accuracy. The resulting reconstructed reading is not accurate, making the integration of billing functionality difficult [13, 39]. Customers' privacy could be breached through Long-term averaging attacks on the frameworks.

To summarize, Distortion-based Frameworks aim to protect customer privacy but may have one or more associated limitations as follows: implementation complexity due to the requirement of additional hardware being installed in the customers' domain, a high computational overhead on the SMs, lack of support for accurate billing functionality due to obscured readings, and vulnerability to single points of compromise.

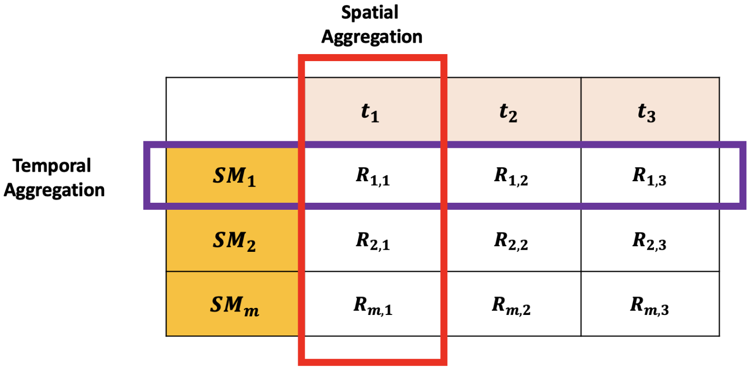
2.3 Aggregation-based Frameworks

In Aggregation-based Frameworks, the aggregated reading is reported to the ESP. Since the reading is aggregated, the ESP cannot link the granular metering data to a specific SM, thereby preserving customers' privacy. There are different levels of aggregation:

1. Home Level Aggregation: Aggregating metering data from various electronic appliances and generation sources [109].

2. Neighborhood Level Aggregation: For Aggregating metering data from different smart meters installed in the neighborhood) [102].

This dissertation will focus on Neighborhood Level Aggregation. Based on the nature of computation (Fig. 2.4), the aggregation-based frameworks can be categorized as follows: a) Spatial Aggregation and b) Temporal Aggregation. If aggregation is computed across a given set of SMs for a given time instance, it is known as spatial aggregation. It is required by the ESP to provide grid functionalities. When aggregation is performed for a specific SM for a given period, it is known as temporal aggregation. ESP requires it to provide billing functionalities. In Figure 2.4, $R_{m,3}$ represents instantaneous reading of m^{th} SM at third time instance ($t = 3$). The frameworks supporting both types of aggregation are known as spatio-temporal frameworks. We further categorize the aggregation-based frameworks on architecture, such as follows:



	t_1	t_2	t_3
SM_1	$R_{1,1}$	$R_{1,2}$	$R_{1,3}$
SM_2	$R_{2,1}$	$R_{2,2}$	$R_{2,3}$
SM_m	$R_{m,1}$	$R_{m,2}$	$R_{m,3}$

Figure 2.4: Spatial and Temporal Aggregation

- In-Network Aggregation-based Frameworks
- Centralized Aggregation-based Frameworks
- Distributed Aggregation-based Frameworks

2.3.1 In-Network Aggregation-based Frameworks

For In-Network Aggregation-based Frameworks, metering data is collected with the help of intermediate SMs within the communication network (Fig. 2.5) [8,14,17,20,25,40,43,53,54,95,100,101]. The In-Network Aggregation-based Frameworks leverage homomorphic techniques to perform the required spatially aggregated computation at different levels of the network hierarchy. While these frameworks significantly reduce the metering data transmitted within the network, they increase the computational overhead on resource-constrained SMs.

As spatially aggregated data is reported to the ESP, it cannot derive granular metering data with respect to a specific SM, thereby preserving privacy. After receiving the spatially aggregated data, the ESP can provide grid functionalities such as real-time monitoring, load balancing, optimization, and demand-response. Although the In-Network Aggregation-based Framework offers various advantages for the smart grid domain, it also has some potential drawbacks, such as the following:

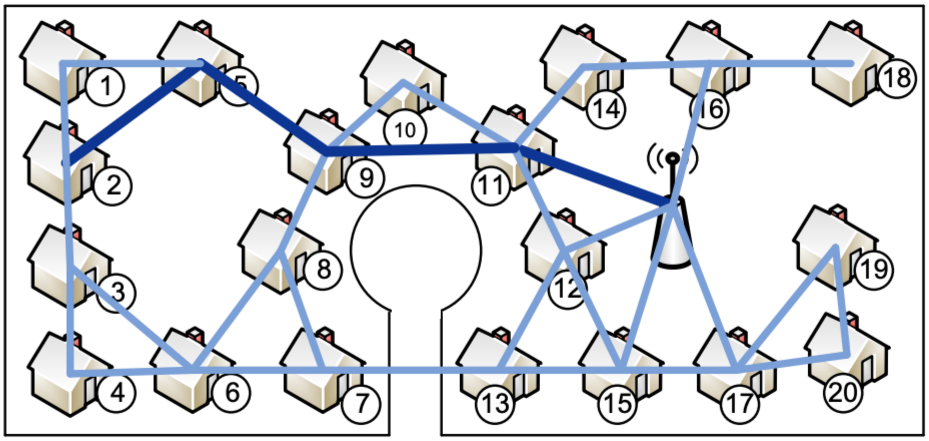


Figure 2.5: In-Network Aggregation-based Framework [54]

- Increased computational overhead: It increases the computational overhead on resource-constrained SMs as they are primarily responsible for processing the metering data.
- Loss of metering data granularity: As metering data is combined across intermediate SMs, there is a loss of granularity in the metering data required for billing purposes, particularly for Time of Use Billing Tariff.
- Increased latency: Aggregation occurring at different levels of the hierarchy introduces additional latency, which can impact the decision-making process.
- Privacy concerns: While in-network aggregation can provide privacy, there is still the possibility of privacy breaches as aggregated data from an initial SM may contain sensitive information.

- Scalability: The addition and/or removal of SMs poses a complex task, which can hinder the scalability of the network.

2.3.2 Centralized Aggregation-based Frameworks

In Centralized Aggregation-based Frameworks [1, 11, 13, 15, 22, 23, 24, 26, 31, 34, 36, 37, 41, 42, 48, 49, 51, 52, 55, 56, 60, 63, 64, 65, 66, 72, 73, 76, 80, 84, 85, 87, 88, 90, 93, 108, 110, 111], metering data is collected, processed, and aggregated in a centralized entity known as a dedicated aggregator (DA). This centralized entity is typically managed by a Trusted Third Party (TTP) that performs the computation and forwards the aggregated result to the ESP (Fig. 2.6). By reporting the aggregated result to the ESP, it prevents the derivation of instantaneous readings related to specific SMs, thereby preserving privacy.

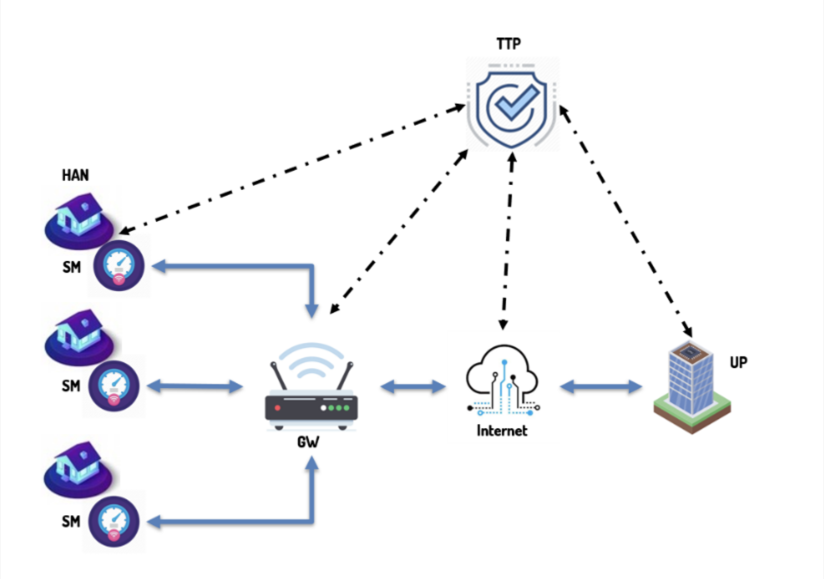


Figure 2.6: Centralized Aggregation-based Framework [11]

Centralized Aggregation-based Frameworks offer a variety of advantages, as follows:

- Efficient Metering Data Management: With the metering data reported and processed centrally, efficient meter data management is achieved.
- Improved Decision Making: Central processing of metering data, collected from multiple resources, enables the ESP to perform the required grid and billing functionalities for the corresponding set of SMs, leading to improved decision making.

- Scalability: The addition and/or removal of SMs is easier compared to In-Network Aggregation-based Frameworks, allowing for greater scalability.

Although there are various advantages associated with Centralized Aggregation-based Frameworks, there are also some drawbacks, as follows:

- Data Privacy and Security: The aggregating entity must have appropriate security and privacy measures in place to prevent unauthorized access and to ensure resilience against malicious attacks.
- Single Points of Compromise: Centralized Aggregation-based Frameworks rely on a single centralized aggregation point for the collection and processing of metering data. In the event of a failure or compromise of this centralized location, the entire framework becomes compromised, leading to disruptions in functionalities. This can potentially cause outages and erode customers' trust.
- Regulatory and Governance Control: Since metering data is collected at a centralized location managed by the Trusted Third Party (TTP), it is crucial for the TTP to ensure compliance with privacy and data-sharing regulations. Failure to meet these regulatory requirements can hinder the implementation and operation of Centralized Aggregation-based Frameworks.

2.3.3 Distributed Aggregation-based Frameworks

In Distributed Aggregation-based Frameworks [2, 58, 86, 102, 103, 104], the collection, processing, and computing of metering data are carried out by multiple entities in a distributed manner (Fig. 2.7). The primary objective of the Distributed Aggregation-based Framework is to provide spatio-temporal metering data to the ESP in a privacy-preserving manner, enabling the ESP to provide grid and billing functionalities. By reporting aggregated readings to the ESP, the framework ensures that the instantaneous reading related to a specific SM cannot be derived, thus preserving customers' privacy. The Distributed Aggregation-based Frameworks eliminate the need for a centralized aggregating entity for meter data processing. The Distributed Aggregation-based Framework has various advantages, as follows:

- Reduced computational overhead: Distributed Aggregation-based Frameworks are lightweight for resource-constrained devices as they offload the computational overhead to aggregating entities.

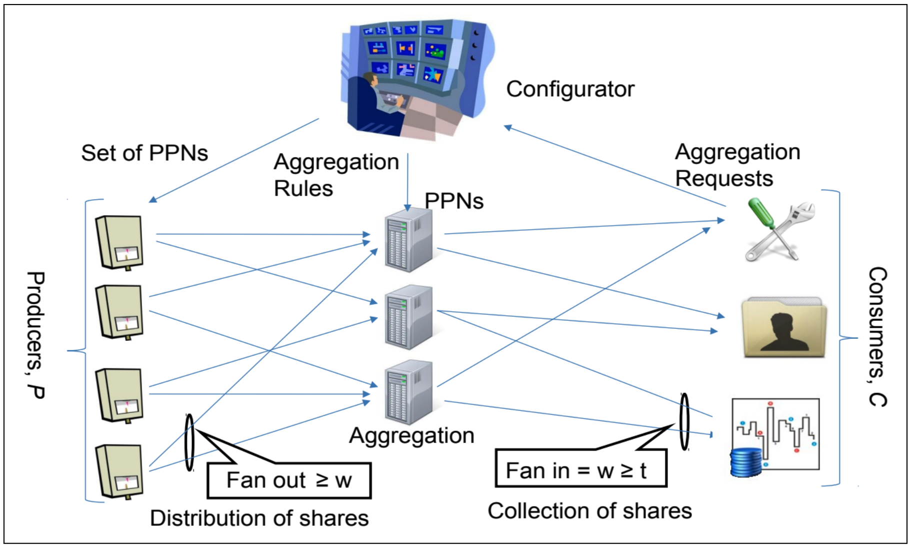


Figure 2.7: Distributed Aggregation-based Framework [86]

- Scalability: These frameworks can handle large-scale deployments in the smart grid domain by distributing computation to multiple aggregating entities participating in spatio-temporal aggregation. The distributed approach allows for the addition and/or removal of SM from the network without significant modifications.

- Fault Tolerance: Due to the distributed nature of aggregation computation, the framework can tolerate a certain threshold of aggregating entity failures compared to Centralized Aggregation-based Frameworks. This fault-tolerance feature enables the ESP to collect aggregating data promptly.

- Privacy: Customer privacy associated with the smart grid is preserved as SMs report shares of metering data to aggregating entities instead of instantaneous readings. The ESP receives aggregated readings for spatio-temporal metering data, thereby maintaining privacy from the ESP's perspective.

- Improved Decision Making: Processing metering data in a distributed fashion and reporting it to the ESP assists in performing the necessary grid and billing functionalities for the corresponding set of SM.

Although there are several advantages associated with existing Distributed Aggregation-based Frameworks, there are some drawbacks associated with them. Some of them are listed below:

- Data Integrity: Since metering data is outsourced from the SMs to the aggregating entities, it is crucial to ensure that the metering data is not subject to any modification. Ensuring consistency of metering data across all the aggregating entities poses a challenging task.

- Time Synchronization: Given that multiple dedicated aggregating entities collaborate to compute spatio-temporal aggregation, synchronization between them is essential to achieve reliable aggregation. Proper synchronization ensures that the metering data is collected from SMs and aggregated correctly.

- Cost of Deployments: Implementing a distributed aggregation-based framework can entail high computational and operational costs, including software updates, monitoring, and maintenance.

- Ownership Concerns: Since multiple aggregating entities participate in computing spatio-temporal aggregation in a distributed setting, it is essential to ensure that these entities comply with the regulatory issues regarding security and privacy. The Energy Service Provider (ESP) should implement proper access control methods to ensure a unique owner for each aggregating entity.

2.4 Research Gaps

As we observed in the previous section (Section 2.3), there are several challenges involved in developing a privacy-preserving framework for the smart grid domain. It is crucial to carefully consider the identified drawbacks and the requirements specified by the National Institute of Standards and Technology Interagency Report (NISTIR) during the development of such a framework [32,81]. The framework should be designed by taking into account the unique requirements and constraints of the associated systems. Ultimately, the developed framework must be practical to implement and resilient against cyberattacks, providing spatio-temporal metering data to the ESP while preserving customer privacy.

The focus of this dissertation is to address the following limitations of existing distributed aggregation-based privacy-preserving frameworks for the smart grid.

- High computational overhead on resource-constrained SMs

- Prone to single points of compromise due to dependency on a centralized entity

- Lack of support for dynamic billing integration while preserving customer privacy
- Lack of integrity verification of spatial and/or temporal metering data, considering a semi-honest threat model

To address the above-mentioned limitations, we propose a distributed aggregation-based privacy-preserving framework that can work in a semi-honest setting as well as malicious setting (dishonest majority of DAs) (Chapter 1, Fig. 1.3). The framework consists of SMs, DAs, and the ESP. The SMs are responsible for reporting the metering data to the ESP in a privacy-preserving manner via the DAs, using secret sharing and commitment-based scheme. The DAs perform spatio-temporal aggregation through secure multiparty computation. Lastly, the ESP is responsible for conducting integrity checks on spatio-temporal metering data while preserving privacy. A detailed explanation of the cryptographic preliminaries is covered in the next chapter.

2.5 Summary

In this chapter we categorized and described various frameworks that aimed to address the customer's privacy issues in the smart grid domain. As our research focused on Aggregation-based Frameworks, we highlighted the research gaps and derived useful insights for this dissertation.

Chapter 3

Building Blocks

This chapter delves into the essential cryptographic preliminaries required to understand our proposed framework and comparative frameworks. The following cryptographic preliminaries are covered in this chapter:

- Secret Sharing Scheme
- Commitment-based Schemes
- Secure Multiparty Computation
- Homomorphic Encryption-based Schemes

3.1 Secret Sharing Scheme

Secret Sharing Scheme [91] is a cryptographic technique employed to distribute a secret (S) among a group of n participants $(P_1, P_2, \ldots, P_n)$ in the form of shares $(share_1, share_2, ..., share_n)$ [5]. This ensures that the secret can only be reconstructed when a group of sufficient participants combine their corresponding shares. Secret sharing provides a robust way of distributing a secret while maintaining high level security. It finds applications in various cryptographic domains, such as password recovery, key management, and secure multiparty computation. Secret sharing guarantees that no individual participant possesses the knowledge of the real secret, but has a portion of it. Different types of secret sharing schemes exist with Shamir's Secret Sharing Scheme [91] being a widely used scheme in real-world scenarios. It is based on Lagrange Interpolation and is commonly known as the (t, n) threshold scheme where $1 < t \leq n$, indicating that at least t participants are required to reconstruct the secret. Lagrange Interpolation is a technique used to reconstruct a polynomial from a set of known points. Table 3.1 represents the notations used for Shamir's Secret Sharing Scheme.

Table 3.1: Table of notations for Shamir's Secret Sharing Scheme

Notation	**Meaning**
δ_j	Basis Polynomial
α_d	d^{th} Coefficient of Secret Sharing Polynomial
C	Combiner
c	Constant
D	Dealer
$t-1$	Degree of Secret Sharing Polynomial
$\mathbb{Z}_p$	Galois Field
P_j	j^{th} Participant
S	Secret
$F(x)$	Secret Sharing Polynomial
$share_j$	Share of j^{th} Participant
p	Prime Number
t	Threshold
n	Total Number of Participants

Shamir's Secret Sharing Scheme consists of following phases:

Setup Phase: In this phase, the Dealer (an entity responsible for generating the shares of a given secret) determines the following:

- A prime number (p) to define the finite field. This value is public to all the entities (Dealer and Combiner) and participants.
- The secret that needs to be shared with the Combiner (an entity responsible for reconstructing the given secret from the shares).
- A threshold (t) to determine the minimum number of shares required to reconstruct the secret.

Share Creation and Distribution Phase: In this phase, the Dealer performs the following operations:

- It randomly selects a polynomial $F(x)$ (eq. 3.1) over a finite field $\mathbb{Z}_p$ with a degree $(t-1)$, where t represents the threshold decided in the previous phase. The coefficients of the polynomial (α_d) are selected randomly from the finite field, and the constant term of the polynomial represents the secret.

$$F(x) = \sum_{d=1}^{t-1} \alpha_d \; x^d + \; S \;\; (mod \;\; p) \tag{3.1}$$

- The Dealer generates the shares (eq. 3.2) by using the identities $(j \in \{1, 2, \ldots, n\})$ of the n participants $(P_1, P_2, \ldots, P_n)$.

$$share_j = F(j) \tag{3.2}$$

- The computed shares $(share_j)$ where $(j \in \{1, 2, \ldots, n\})$ are distributed to the corresponding participants $(P_1, P_2, \ldots, P_n)$.

Reconstruction Phase: In this phase, the Combiner performs the following operations:

- It fetches the shares from the corresponding participants $(P_1, P_2, \ldots, P_n)$.
- It employs Lagrange Interpolation technique over finite fields.
- It computes the reconstructed polynomial and fetches the secret by accessing the constant (eq. 3.3).

$$F(x) = \sum_{j=1}^{n} (share_j) \; (\delta_j(x)) = \sum_{d=0}^{t-1} \alpha_d \; x^d \;\; (mod \;\; p) \tag{3.3}$$

Figure 3.1 represents a 3-5 threshold scheme, where the secret (S) is broken into five shares $(share_1, share_2, \ldots, share_5)$, such that any three shares can reconstruct the secret (S).

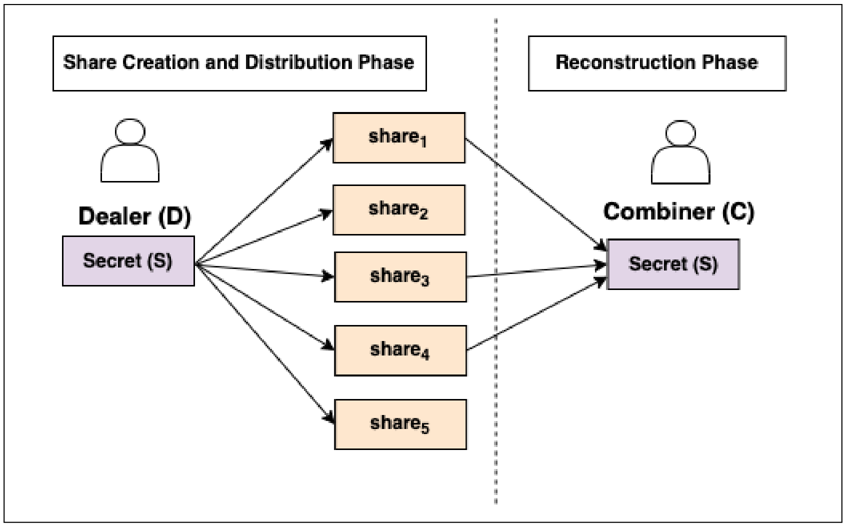


Figure 3.1: Example of Shamir's Secret Sharing Scheme

Properties of Shamir's Secret Sharing Scheme:

- Individual shares do not provide any information about the secret.
- Any set of t shares can recover the secret.
- Any set of $t-1$ shares cannot recover any information about the secret.
- The corresponding shares of the secrets S_1 and S_2 can be added across different participants if the degree of secret sharing polynomial is the same. Thus, the resultant share obtained represents the share of the total secrets $(S_1 + S_2)$.
- The corresponding shares of the secret (S) can be multiplied by a constant (c) such that the resultant secret obtained represents the constant times the secret $(c \times S)$.
- Even if some shares are unavailable or lost, the secret can still be reconstructed until the threshold number of shares are available to the Combiner.
- The Dealer can select the threshold based on the requirement of the given use case.
- The secret sharing scheme provides confidentiality since only the authorized entity (Combiner in our case) can reconstruct the secret.
- An adversary has to compromise up to t participants in order to reconstruct the secret, a task which is more difficult than compromising a centralized location (entity).
- An adversary without having the required number of shares cannot reconstruct the given secret even if it has infinite computing and time capacity, thereby implying information-theoretic security.

- The Dealer can dynamically create new shares for new participants without affecting the existing set of shares for a given set of secret.
- Security can be enhanced by increasing the degree of the polynomial, thereby increasing the number of threshold limit to reconstruct the secret (Note: The constant term remains the same, that is, the secret).

Although Shamir's Secret Sharing Scheme has various advantages, it has the following limitations:

- During the share reconstruction process, the Combiner needs a way to verify the correctness of each share provided by the participants. Verifiable secret sharing [8] has been studied by the smart grid community to determine the correctness of the shares. However, the scheme focuses on sharing encryption keys used by smart meters (SMs) as the secret, rather than the corresponding granular meter reading of the SMs.
- Single point of reconstruction: The secret exists with the Dealer, who is the owner of the secret and the Combiner, who is the recipient of the secret. In this setup, the secret is held by both entities. It is advisable to have a backup in scenarios where a single points of failure can occur.

3.2 Commitment-based Schemes

Commitment-based scheme [79] is a cryptographic technique that enables the Prover (P) to commit to a specific value while preserving its secrecy and integrity, and later reveal it to the Verifier (V). The Verifier is the entity that receives commitment from the Prover. The Verifier's role is to check the integrity of the commitment. One notable application of the Commitment-based Schemes is in electronic voting systems, where it ensures that voters cannot modify their decision after committing, thus safeguarding the integrity of the electronic voting process (Fig. 3.2). Table 3.2 represents the notations used for Commitment-based Scheme.

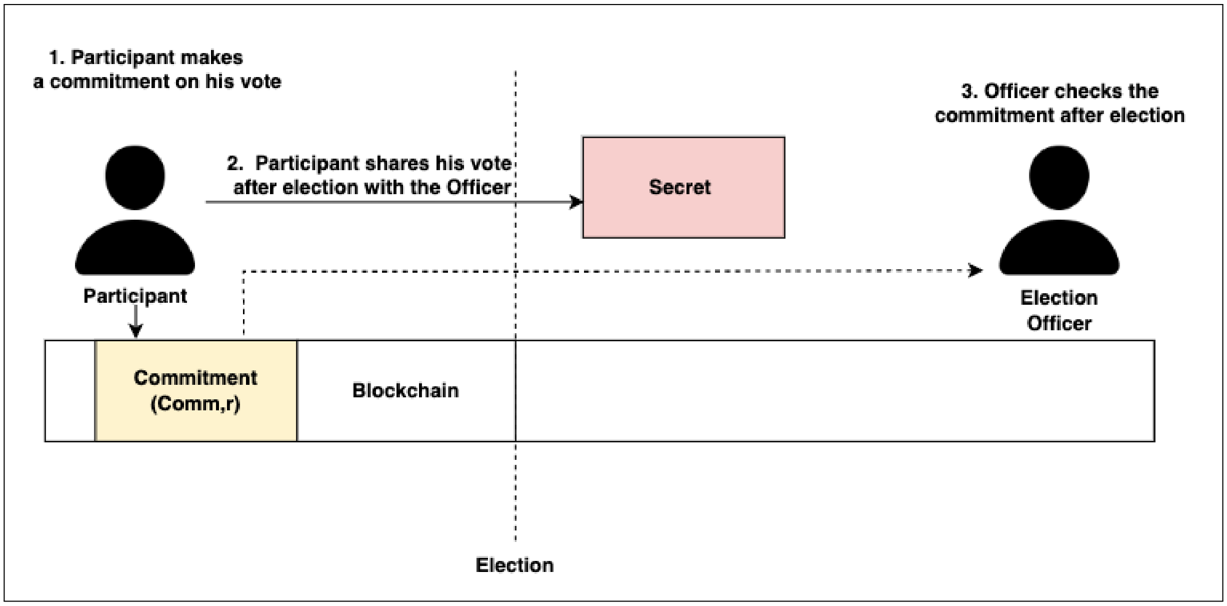

1. Participant makes a commitment on his vote
2. Participant shares his vote after election with the Officer
Secret
3. Officer checks the commitment after election
Participant
Election Officer
Commitment (Comm,r)
Blockchain
Election

Figure 3.2: Example of Commitment-based Scheme

Table 3.2: Table of notations for Commitment-based Schemes

Notation	Meaning
g, h	Commitment Parameters
r	Decommitment Value
C	Generated Commitment
p	Prime Number
P	Prover
S	Secret
V	Verifier

The Commitment-based Scheme consists of following phases:

Commit Phase: During this phase, the Prover creates a commitment (C) (eq. 3.4), using a decommitment value (r) for the corresponding secret (S), and shares it with the Verifier. The commitment parameters g, h are public and are also known to the Verifier.

$$C = g^s \; h^r \; (mod \; p) \tag{3.4}$$

Reveal Phase: During this phase, the Prover reveals the secret, and the decommitment value to the Verifier. The Verifier computes the commitment using the values received from the Prover and verifies it against the received commitment. If the commitments match, it implies that the secret was not altered after the commitment was distributed to the Verifier.

Properties of Commitment-based Schemes:

- Security and Privacy: It allows participants to commit to a value while keeping it hidden until the Prover discloses it. This property is also known as the hiding property.
- Binding property: It ensures that the Prover cannot change the committed value after generating and distributing the commitment to the Verifier, thereby preventing dishonest behaviour.
- Integrity and Immutability: The Commitment-based Schemes binds a Prover to a specific committed value, ensuring integrity and immutability. Once a commitment is generated, it becomes difficult to alter or modify.
- Non-malleability property: This property guarantees that no participant can derive the original committed value from the commitment by modifying or altering the commitment.
- Verifiability and Auditing: It provides the Verifier with the ability verify and audit a commitment. This encourages trust, accountability, and transparency among participants (Prover and Verifier).
- Non-Repudiation: By employing a commitment scheme, the participants cannot deny their commitment to a specific value. A commitment made and later revealed with necessary proof servers as digital evidence that the Prover cannot contradict. This property is particularly useful in legal or contractual scenarios where participants must adhere to certain terms and/or conditions.

Following are some limitations associated with the Commitment-based Schemes:

- Irreversibility: The commitment scheme is designed to be irreeversible in nature, meaning that once a commitment is generated by the Prover it cannot be changed or modified.
- Trust assumption: The commitment scheme relies on the trust between the Prover and the Verifier. If the trust is compromised, it can lead to attacks on the commitment scheme.
- Computational overhead: The scheme can introduce computational overhead depending on the construction of function, verification, and associated cryptographic operations. It might hamper the scalability and efficiency of the system in use.
- Communication overhead: Since the commitment scheme requires multiple communication between the Prover and Verifier during the commit and reveal phases, it can introduce communication overhead, thereby introducing latency in the distributed systems involving multiple

participants. Therefore, it is important to understand the requirements of the system and use the cryptographic Commitment-based Schemes accordingly.

3.3 Secure Multiparty Computation

Secure multiparty computation (SMPC) [16,57] (Fig. 3.3) is a method that allows multiple participants $(P_1, P_2, \ldots, P_n)$ to collaboratively compute a given function based on their inputs, without revealing their inputs $(x_1, x_2, \ldots. x_n)$ to each other. This technique is useful in scenarios where the participants possess sensitive data and do not want to disclose it to other participants, yet still need to compute a given function that involves data from all the participants.

The general SMPC process can be subdivided into three phases:

Input preparation: Each participant prepares their input for the computation, without revealing their original inputs to the other participants.

Computation: The participants jointly compute the given function on the prepared inputs.

Output reconstruction: The participants reconstruct the output of the computation.

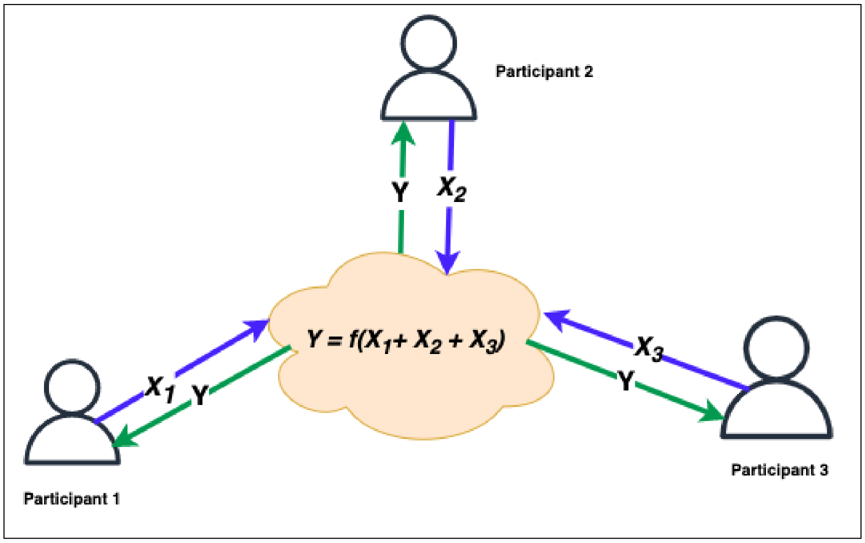


Figure 3.3: Example of Secure Multiparty Computation

Example: Three participants (P_1, P_2, P_3) want to calculate the average of their (x_1, x_2, x_3) salaries without revealing the salaries to each other.

- Input preparation: Each participant encrypts their salary using a given cryptographic technique (Note: The original input, that is individual salaries, is hidden from each participant).
- Computation: The participants compute the summation function by directly working on the ciphertext without converting it into plaintext.
- Output reconstruction: The encrypted ciphertext is decrypted collaboratively by the participants without learning the individual salaries. Finally, the decrypted sum of salaries is divided by the total number of participants (in our case, three) to compute the average salary.

At the end of the process, the participants ($P_1, P_2, and\ P_3$) get the average salary without revealing their salaries to each other. SMPC ensures no participants gain additional information beyond the expected output, which is the average of their salaries.

Properties of Secure Multiparty Computation (SMPC):

- Privacy: SMPC allows multiple participants to collaboratively compute a given function on their private inputs without disclosing their inputs. This property ensures that the given inputs of the participants remain private.
- Correctness: SMPC is designed to compute the given function correctly, regardless of the value of inputs from the corresponding participants.
- Fairness: SMPC ensures no participant can gain any additional advantage by holding on to their inputs and/or deviating from the normal execution.
- Robustness: SMPC can operate correctly even within a subset of malicious participants trying to deviate from the normal execution, making it robust.
- Verifiability: It can check the correctness of the given function to ensure the accuracy.

Although secure multiparty computation has various advantages, it has the following drawbacks:

- Computational and Communication Overhead: The participants require significant computation and communication resources to perform SMPC.
- Scalability: The computational and communication overhead increases as the number of participants grows. Therefore, designing, implementing, and deploying SMPC can be a complex task.

3.4 Homomorphic Encryption-based Schemes

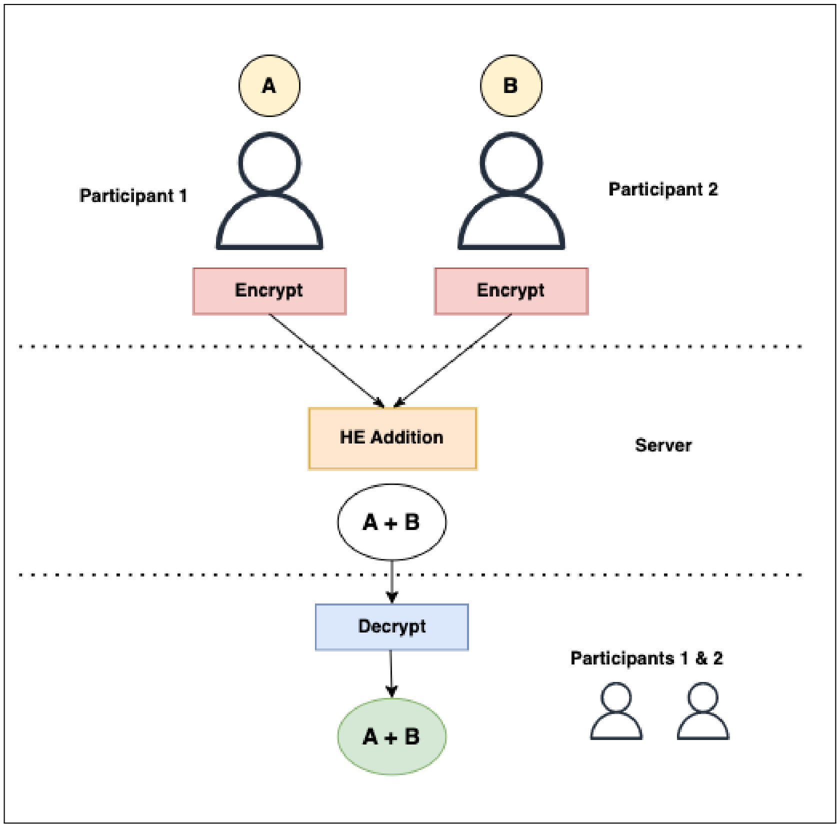


Figure 3.4: Example of Homomorphic Encryption-based Scheme

A Homomorphic Encryption-based Scheme (Fig. 3.4) allows participants to perform operations on encrypted data without decrypting it to plaintext, making the scheme applicable in scenarios where privacy preservation is required. The Homomorphic Encryption-based Schemes can be categorized as follows:

- Partial Homomorphic Encryption-based Scheme (PHE) enables limited computations on encrypted data. PHE supports either the addition or multiplication operation on encrypted data without requiring decryption into plaintext. However, PHE does not support both operations simultaneously.
- Somewhat Homomorphic Encryption-based Scheme (SHE) enables computations on encrypted data, specifically supporting the addition and/or multiplication operation with a constant.

However, SHE has limitations on the number of computations it can support, which makes it less powerful compared to the Fully Homomorphic Encryption-based Scheme.

- Fully Homomorphic Encryption-based Scheme (FHE) supports a broader range of operations compared to the previous two schemes. However, it has a high computational overhead associated with it.

For our comparative analysis, we will be focusing on the Paillier Encryption Scheme [77] introduced by Pascal Paillier, a type of Partial Homomorphic Encryption-based Scheme. This scheme has been widely studied for smart grid applications. It is an additive homomorphic encryption scheme, which means it supports addition operations on the encrypted ciphertext.

The Paillier Encryption Scheme can be generalized as follows:

Keys:

- Public Key (*public*): (n, g)
- Private Key (*private*): (λ, μ)

Encryption:

- Select a secret message (S) such that $0 \leq S < n$
- Select a random number (r) such that $0 \leq r < n$ and $gcd(r, n) = 1$
- Compute ciphertext $(C) = Enc_{public}(S, r) = g^S \cdot r^n mod\, n^2$

Decryption:

- Compute the plaintext $(S) = Dec_{private}(C) = L(C^\lambda\ mod\, n^2) \cdot \mu\ mod\, n$
  Note: Here L is a function defined as $L(x) = (x - 1)/n$

Properties of Paillier Encryption Scheme:

- Homomorphic Addition: Given two ciphertexts C_1 and C_2 of corresponding plaintext S_1 and S_2, the product of two ciphertexts when decrypted results in sum of the plaintexts.

- Scalar Multiplication property: Given a ciphertext (S) and a scalar multiple (c) in plaintext, the ciphertext raised to the power of the scalar constant gets decrypted to the product of the plaintext and the scalar multiple $(S \times c)$. The homomorphic addition and scalar multiplication property makes the Paillier scheme useful in aggregation scenarios dealing with sensitive data.
- Non-Malleability: The Paillier Encryption Scheme is non-malleable, meaning that even if an adversary modifies the ciphertext, the decrypted plaintext remains unaffected. This property provides integrity and prevents unauthorized manipulation of the encrypted data.
- Efficiency: The Paillier Encryption Scheme is computationally efficient compared to Fully Homomorphic Encryption-based Schemes.

Table 3.3: Table of notations for Paillier Encryption Scheme

Notation	**Meaning**
C	Ciphertext
L	Function
(λ, μ)	Private Key
(n, g)	Public Key
r	Random Number
c	Scalar Constant
S	Secret

Limitations of Paillier Encryption Scheme are as follows:

- Homomorphic Operation Limitations: Paillier Encryption Scheme supports only homomorphic addition and scalar multiplication properties. However, it does not support other general multiplication, division, and exponential operations on the encrypted data. This limits the Paillier Encryption Scheme's utilization on various computations.
- Computational Overhead: The Paillier Encryption Scheme is more efficient than the Fully Homomorphic Encryption-based Scheme. However, it still has a substantial computational overhead on resource-constrained devices due to modular and exponential operations associated with it. Hence, the parameter selection and management associated with Paillier Encryption Scheme requires careful attention as it ensures security.

3.5 Summary

This chapter provided an overview of the cryptographic preliminaries that formed the basis of our proposed framework, including the Secret Sharing Schemes, Commitment-based Schemes, and Secure Multiparty Computation. It also discussed the Homomorphic Encryption-based Schemes, which were associated with the comparative frameworks.

Chapter 4

A Privacy Safeguarding Framework: Semi-Honest Setting

4.1 Introduction

In this chapter, we will be focusing on the following research questions (RQs):

- **RQ1: How can smart meters (SMs) send spatial metering data to the Electrical Service Provider (ESP) in a way that protects privacy and ensures efficiency for the smart meters themselves?**
- **RQ2: How can smart meters (SMs) send temporal metering data to the Electrical Service Provider (ESP) in a way that protects privacy and ensures efficiency for the smart meters themselves?**

In Chapter 1, we emphasized the importance of spatial and temporal metering data required by the ESP for providing grid and billing functionalities. However, studies have indicated that accessing high-frequency metering data can breach customer privacy [19, 21, 25, 33, 44, 45, 59, 67, 68, 82, 99]. In Chapter 2, we highlighted the significance of Aggregation-based Frameworks in addressing the privacy issues in the smart grid. By reporting the aggregated metering data to the ESP, it becomes difficult for the ESP to establish the link between the high-frequency metering data and the corresponding SM, thus preserving customer privacy. In this chapter, we conduct a critical literature review of the existing Aggregation-based Frameworks for the smart grid and propose our distributed privacy-preserving framework to address their limitations.

The Aggregation-based Frameworks are categorized based on the architectural model as follows [8, 14,15,25,40,54,86,87,100,102]: In-Network Aggregation-based frameworks, Centralized Aggregation-based Frameworks and, Distributed Aggregation-based Frameworks.

In-Network Aggregation-based Frameworks: These frameworks expect the metering data to be collected with the help of intermediate SMs within the communication network. We critically analyze the most closely related In-Network Aggregation-based Frameworks.

In [8], the authors propose an In-Network Aggregation-based Framework that enables SMs to securely communicate high-frequency metering data to the ESP while preserving customer privacy. This is achieved through the use of secure multiparty computation (SMPC) and verifiable secret sharing techniques. Notably, the proposed solution eliminates the need for a dedicated aggregator (DA) and instead relies on in-network aggregation utilizing the SMs themselves. Verifiable secret sharing is employed to safeguard the keys used for encrypting the high-frequency metering data. Each SM generates shares of its private keys and distributes them among neighboring SMs. Meanwhile, the SMs directly report the low-frequency metering data to the ESP. Upon receiving the aggregated value of the keys, the ESP employs it to decrypt the total encrypted reading. This decryption process, involving the aggregation of encrypted readings, becomes possible due to the homomorphic properties of the cryptographic schemes utilized. By aggregating the metering data over a given period, the framework effectively prevents the ESP from deriving granular metering data, thereby preserving privacy. Additionally, the framework extends support for Flat Rate Billing. In [14], the authors propose an In-Network Aggregation-based Framework that supports spatio-temporal aggregation. It utilizes the Pedersen Commitment Scheme [79] for verifiability of the bill generated and, the Paillier scheme [77] for computing spatio-temporal metering data. However, the computational overhead is high due to the interconnection of SMs. This framework supports both Flat Rate and Dynamic Billing. Authors in [100] explore a framework that combines a Fully Homomorphic Encryption-based (FHE) scheme with SMPC to report spatially aggregated data to the ESP within an acceptable timeframe. However, this framework does not support temporal aggregation and exhibits a high computational overhead on the SMs. Complete group management is necessary for SM addition and/or removal in this framework. In [54], the authors present an In-Network Aggregation-based Framework that supports spatial aggregation using the Paillier Encryption Scheme. However, this framework does not include billing functionality and imposes a high computational overhead on resource-constrained SMs. On the other hand the framework proposed in [40] is capable of tolerating a threshold of SM crashes during the aggregation process.

It employs the Shamir's Secret Sharing Scheme to preserve customer privacy and operates under an honest-but-curious model. However, the requirement for interconnection among SMs increases the architectural complexity.

Centralized Aggregation-based Frameworks: These frameworks expect the metering data to be collected, processed, and aggregated by a centralized entity known as a dedicated aggregator (DA). This centralized entity is typically managed by a Trusted Third Party (TTP) that performs the computation and forwards the aggregated result to the ESP. We critically analyze the most closely related Centralized Aggregation-based Frameworks.

In [15], the authors propose a framework that offers fault tolerance by utilizing the Paillier Encryption Scheme. The framework supports spatial and temporal aggregation and outperforms the framework presented in [25]. However, it is susceptible to single points of compromise. Additionally, the framework only supports Flat Rate Billing and requires partial group management for SM addition and/or removal. The framework presented in [87] supports spatial aggregation while preserving customer privacy. It expects each SM to encrypt its reading using homomorphic encryption-based scheme and report it to the aggregator. Through SMPC the aggregator computes the spatial aggregation without violating customer privacy. However, the framework is prone to single points of compromise and lacks the support for dynamic billing.

Distributed Aggregation-based Frameworks: These frameworks utilize multiple dedicated entities to perform collection, processing and aggregation of metering data. We critically analyze the most closely related Distributed Aggregation-based Frameworks.

In [86], the authors address the limitations of [87] by introducing privacy-preserving nodes (PPNs) between the SMs and ESP. The framework employs Shamir's Secret Sharing Scheme to collect metering data from the SMs and leverages its homomorphic properties to support spatial and temporal aggregation. However, the framework lacks support for dynamic billing and is prone to single points of compromise, as the ESP is responsible for the centralized reconstruction after fetching data from the PPNs. To address the limitations of [86], we propose a framework [102] that incorporates Shamir's Secret Sharing Scheme for SMs and utilizes SMPC in conjunction with the DAs employing and the ESP. However, the framework does not support dynamic billing such as Time of Use Billing Tariff.

With the integration of dynamic billing functionalities in the smart grid domain, the computation of accurate bills necessitates granular high-frequency metering data. However, the accessibility of this data poses potential privacy risks for the customers. Therefore, integration of dynamic billing functionality while preserving customer privacy is a difficult task. As the smart grid aims to provide grid and billing functionalities, employing a thoughtfully designed spatio-temporal aggregation framework can serve the dual purpose of safeguarding customer privacy and facilitating dynamic billing. Such a framework enables spatial and temporal aggregation functionalities within the same underlying system.

To summarize the aforementioned Aggregation-based Frameworks, the existing solutions have one or more of the following limitations:

- Prone to single points of compromise due to dependency on a centralized entity for aggregation [15, 87]
- High computational overhead on resource-constrained SMs [8, 14, 15, 40, 87, 100]
- Lack of support for dynamic billing integration while preserving customer privacy [8, 40, 54, 86, 87, 100, 102]

We aim to address these limitations through our distributed framework in a semi-honest setting, which can report spatio-temporal metering data while preserving customer privacy. Since the majority of the computing load is shifted to the DAs, the framework is lightweight in terms for the resource-constrained SMs.

4.2 System Model

This section describes the architectural, threat, and billing model associated with the privacy safeguarding framework in a semi-honest setting.

4.2.1 Architectural Model

Smart Meters (SMs):

A smart meter is a resource-constrained device that is installed in the customer's domain. It is capable of reporting both high-frequency and low-frequency metering data to the ESP through the DAs. The SMs can perform cryptographic operations, such as generating shares for corresponding instantaneous readings, utilizing the parameters provided by the ESP. We assume that each customer has one SM installed, and the instantaneous readings (consumption) are independent of each other. Each SM is assigned a unique identifier denoted by i, where $i \in \{1, 2, \ldots, m\}$; $m =$ represents the total number of SMs. Note: The SMs do not interact with each other, their interactions are solely with the DAs.

Dedicated Aggregators (DAs):

A dedicated aggregator is an intermediate device positioned between the SMs and the ESP. Unlike the SMs, it possesses high computational capabilities. The DAs receive their initialization parameters from the ESP, similar to the SMs. The main responsibility of the DAs is to perform spatio-temporal aggregation employing SMPC. Each DA is assigned a unique identifier, denoted by j, where $j \in \{1, 2, \ldots, n\}$; n represents the total number of DAs participating in the spatial and temporal aggregation. Each DA maintains the following types of memory registers:

- Spatial Register ($SR_{j,t}$): This register stores the aggregated shares received by DA_j from m SMs for a given time instance (t).
- Temporal Register ($TR_{i,j}$): This register stores the aggregated shares received by DA_j from SM_i over a time period (T).

Each DA is responsible for reporting the output of the SMPC (the reconstructed polynomial) to the ESP. Our research aligns with the NIST guidelines that emphasize the use of a redundancy strategy to enhance the resilience of the smart grid (Section 1.3) [32, 81]. DAs can be deployed on-premises, cloud-based, and/or in a hybrid manner (both on-premises and cloud) [7, 12, 34, 61, 69, 78].

Electrical Service Provider (ESP):

The ESP is responsible for communicating the initialization parameters to the SMs and DAs. The Initialization Phase is assumed to be secure against active-passive attacks. The ESP is also responsible for setting the billing tariffs including Flat Rate, Cumulative, and Time of Use Billing Tariffs. The ESP is accountable for providing grid and billing functionalities.

4.2.2 Threat Model

In this chapter, we focus on passive threats to privacy (semi-honest setting). The SMs are assumed to be tamper-resistant but are trusted for their readings. The entities (SM, DA, and ESP) are considered semi-honest. The DAs and ESP follow the protocol but may attempt to breach customers' privacy by colluding with each other, utilizing their received inputs and associated registers. Their objective is to link the high-frequency meter readings to the identities of the SMs. The SMs can try to infer high-frequency metering data of other SMs. We assume that the exchanges between the entities (SM-DA, DA-DA and, DA-ESP) are protected from active adversaries through secure connections. Additionally, the internal clocks are in sync with each other, ensuring synchronization within the framework.

4.2.3 Billing Model

Various types of billing tariffs are utilized in the smart grid community [10, 28] to promote energy conservation to manage the supply-demand curve. Each billing tariff serves a specific purpose and aims to achieve a particular goal. The ESP determines the selection of a billing tariff during the Initialization Phase (Section 4.4.1). The choice of a specific tariff depends on the ESP's goals and the customer's preferences. This subsection describes the different types of billing tariffs supported by our proposed framework.

Flat Rate Tariff

A Flat Rate Tariff applies a uniform price per unit ($price_{unit}$) to the entire meter reading (kWh), regardless of the time and duration of consumption. While a Flat Rate Tariff is easier for consumers to understand, it may not provide a strong incentive for energy conservation. This type of billing tariff is also supported in legacy electrical grids.

Example: A customer's electricity consumption is 770 kWh in a given month, and the charge per unit kWh is \$0.10 (determined by the ESP). In this case, the total bill generated will be \$77 (770 kWh × \$0.10 per kWh). The total bill generated for the customer is directly proportional to the electricity consumed during the given month.

Cumulative Tariff

In this type of tariff, a different pricing structure is applied to the total electricity consumption over a given period, usually a month. The overall consumption is divided into intervals, with each interval mapped to a special rate per consumption unit. This tariff encourages electricity conservation and reflects the increase in the marginal cost of providing the service as consumption increases. The ESP might have a Cumulative Tariff that charges a lower rate for the initial 200 kWh of electrical consumption in a month and a higher rate for any consumption above 200 kWh. This motivates customers to consume less electricity as the cost of the electric bill increases with additional consumption beyond the limit. However, the Cumulative Billing Tariff may be more challenging for customers to understand than a flat rate billing tariff.

Example: A customer with a total electricity consumption of 770 kWh over a month, and the ESP has the following pricing structure: For the first 200 kWh consumed, the price is $0.10 per kWh, and for any consumption above 200 kWh, the price is $0.20 per kWh. In this case, the total bill generated would be $134. The first 200 kWh would be charged at $0.10 per kWh, totaling $20, and the remaining 570 kWh would be charged at $0.20 per kWh, totaling $114.

Time of Use Tariff

The Time of Use (TOU) Tariff is a type of dynamic billing tariff. In TOU, the pricing structure for electricity varies based on the time of consumption. The TOU Tariff is determined by the actual cost of generating and distributing electricity at different intervals. This type of billing tariff encourages the customers to shift their consumption to non-peak hours when electricity demand is lower, and the price is cheaper. Typically, a TOU tariff divides electricity into peak, shoulder, and non-peak periods, with the highest prices during peak periods, followed by shoulder periods and non-peak periods. TOU Tariffs can be used with SM, which record instantaneous consumption, enabling more accurate billing and increased granularity in pricing visibility. However, TOU tariffs are more complex for customers to understand than Flat Rate and Cumulative Tariffs.

Example: The ESP implements the following TOU tariff structure:

- Peak periods from Monday to Wednesday, 08:00 AM to 09:00 PM, with a price of $0.30 per kWh.
- Shoulder periods are all day on weekends and Friday with a price of $0.20 per kWh.

- Non-peak periods from Monday to Wednesday, 09:00 PM to 8:00 AM with a price of $0.10 per kWh.

If a customer consumes 570 kWh during peak periods in a month, the total bill would be $171 (570 kWh × $0.30 per kWh). If the customer consumes 570 kWh during shoulder periods, the total bill would be $114 (570 kWh × $0.20 per kWh). If the customer consumes 570 kWh during non-peak periods, the total bill would be $57 (570 kWh × $0.10 per kWh). In this example, the customer would benefit economically by shifting their energy consumption to non-peak periods when electricity prices are lowest. Therefore, TOU Billing Tariffs can effectively manage the supply-demand curve and motivate energy conservation among customers.

4.3 Cryptographic Prerequisites

4.3.1 Configuration of Shamir's Secret Sharing Scheme

Shamir's Secret Sharing Scheme [91] is a threshold secret sharing scheme used by resource-constrained SMs. It enables the SM to divide its instantaneous reading ($R_{i,t}$) into n number of shares, which are then distributed to the DAs for spatial and temporal aggregation by employing SMPC. Here n represents the total number of DAs. The shares are divided in a manner that allows the DAs to reconstruct the instantaneous reading with a subset of k shares. Therefore, the scheme is also known as a (k, n) threshold scheme (k and n are positive integers) where k represents the minimum number of shares required to reconstruct instantaneous reading ($R_{i,t}$). However, in order to provide the ESP with spatial and temporal aggregated data for grid and billing functionalities, we leverage the homomorphic properties of Shamir's Secret Sharing Scheme as follows:

- Additive Property: The corresponding shares of the instantaneous readings ($R_{1,t}$) and ($R_{2,t}$) can be added across different SMs (SM_1 and SM_2) if the degree of secret sharing polynomial is the same. Thus, the resultant share obtained represents the share of the total readings ($R_{1,t} + R_{2,t}$). This additive property of the Shamir's Secret Sharing Scheme helps the DAs to compute spatially aggregated reading in a privacy-preserving manner.
- The corresponding shares of the instantaneous reading ($R_{i,t}$) can be multiplied by a constant (c) such that the resultant secret obtained represents the constant times the instantaneous reading ($c \times (R_{i,t})$). This property is useful for incorporating dynamic billing (Time of Use Tariff) in a distributed and privacy-preserving manner.

- Shamir's Secret Sharing Scheme is information-theoretically secure, meaning that an adversary without the required number of threshold shares cannot reconstruct the secret, even with infinite time and computing capacity.

A more detail explanation about how Shamir's Secret Sharing Scheme is employed in the proposed framework is covered in Section 4.4.

4.3.2 Configuration of Secure Multiparty Computation

SMPC [16, 57] enables a group of DAs in our proposed framework to compute a function (spatial aggregation as well as temporal aggregation) without disclosing their individual inputs (shares received from the SMs). In SMPC, all the participating DAs jointly compute a function over their inputs such that they are only aware of the result (spatially aggregated reading and /or temporally aggregated reading) and their inputs stored in spatial ($SR_{j,t}$) and temporal registers ($TR_{i,j}$) .

Since only aggregated data is shared across the DAs in the form of registers (spatial register and temporal register), the DAs cannot reconstruct the instantaneous reading with a given smart meter thereby preserving customer privacy. Additionally, the DAs share the aggregated result of the SMPC with the ESP for providing grid and billing functionalities. Even in that case, ESP also cannot breach the customer privacy as it receives computed spatially aggregated reading and/or temporally aggregated reading from the DAs.

The SMPC phases can be defined as follows:

- **Input preparation:** Each DA prepares its input by aggregating the granular shares received from the SMs in spatial and temporal registers. This ensures that their inputs are ready to be utilized in the computation of spatio-temporal aggregation without revealing the input (granular shares) to the other dedicated aggregators.

- **Computation:** The DAs jointly compute spatially aggregated reading over a given set of SMs for a given instance of time (utilizing spatial registers) (Fig. 4.1). Additionally, the DAs compute temporally aggregated reading for a given set of SMs over a given period of time (utilizing temporal registers).

- **Output reconstruction:** The DAs reconstruct the output of the computation (spatially aggregated reading and temporally aggregated reading) without revealing their inputs (granular shares) to each other.

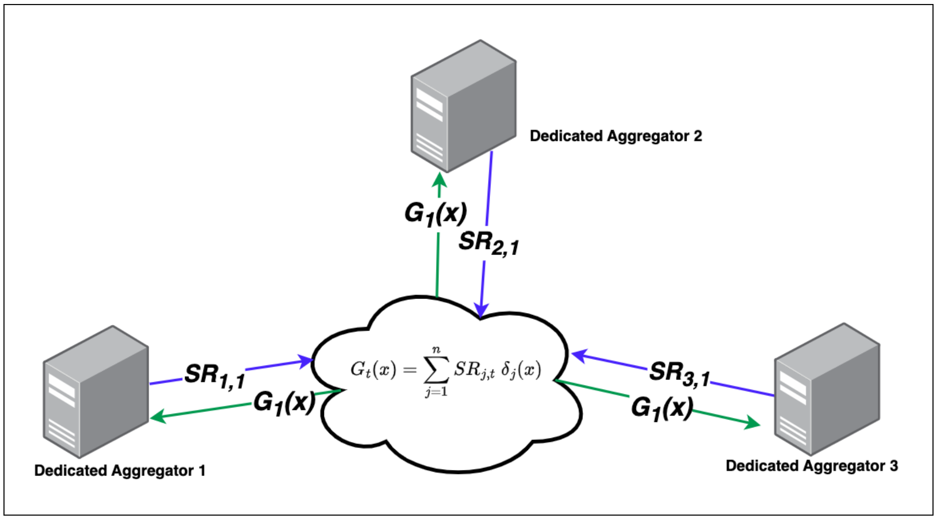


Figure 4.1: Secure Multiparty Computation (Spatial Aggregation)

Our proposed framework is lightweight in terms of computational overhead on SMs, as the majority of the work is performed by the DAs. By employing Shamir's Secret Sharing Scheme, the shares generated by each SM are distributed across the DAs. This allows for the computation of spatially or temporally aggregated readings when at least k or more DAs exchange their respective registers. Therefore, SMPC helps the framework address single points of compromise. A more detailed explanation of how SMPC is employed in our proposed framework is covered in Section 4.4.

4.4 Proposed Framework

This section describes the phases associated with our privacy safeguarding framework in a semi-honest setting. The framework employs Shamir's Secret Sharing Scheme and SMPC to provide spatio-temporal metering data to the ESP while preserving consumer privacy. Table 4.1 represents the notations used for our privacy safeguarding framework in a semi-honest setting. The proposed framework consists of the following phases:

4.4.1 Initialization Phase

The ESP communicates the initializing parameters to the SMs and DAs.

The SMs receive the following:

- Degree of secret sharing polynomial $(k-1)$
- Prime number (p)

Table 4.1: Table of notations: A Privacy Safeguarding Framework in a Semi-Honest Setting

Notation	**Meaning**
$\delta_j(x)$	Basis Polynomial
$Bill_i$	Bill generated for i^{th} Smart Meter
$\beta_{d,t}$	d^{th} Coefficient of Reconstructed Polynomial at time t (Spatial Aggregation)
$\gamma_{d,i,t}$	d^{th} Coefficient of Reconstructed Polynomial for i^{th} Smart Meter at time t (Temporal Aggregation)
$\alpha_{i,d,t}$	d^{th} Coefficient of Secret Sharing Polynomial of i^{th} Smart Meter at time instance t
Con_{int}	Consumption for given interval (Cumulative Tariff)
$k-1$	Degree of Secret Sharing Polynomial
ESP	Electrical Service Provider
$Tariff$	Flat Rate / Cumulative / Time of Use (TOU)
SM_i	i^{th} Smart Meter
$R_{i,t}$	Instantaneous reading of i^{th} Smart Meter at time instance t
t	Instantaneous time
DA_j	j^{th} Dedicated Aggregator
DA_{List}	List of Dedicated Aggregators
SM_{List}	List of Smart Meters
n	Number of Dedicated Aggregators
m	Number of Smart Meters
$price_{int}$	Price for given interval (Cumulative Tariff)
$price_{max}$	Price for last interval (Cumulative Tariff)
$price_t$	Price for that given time instance (Time of Use Tariff)
$price_{unit}$	Price per unit consumption (Flat Rate Tariff)
p	Prime number
$G_t(x)$	Reconstructed Polynomial at time instance t (Spatial Aggregation)
$H_{i,T}(x)$	Reconstructed Polynomial of i^{th} Smart Meter at time T (Temporal Aggregation)
$F_{i,t}(x)$	Secret Sharing Polynomial of i^{th} Smart Meter at time instance t
$(share_j)_{i,t}$	Share generated by i^{th} Smart Meter for j^{th} Dedicated Aggregator at time instance t
$SR_{j,t}$	Spatial Register of j^{th} Dedicated Aggregator
$TR_{i,j}$	Temporal Register of i^{th} Smart Meter at j^{th} Dedicated Aggregator
T	Total time

- List of DAs participating in spatio-temporal aggregation (DA_{List})

The DAs receive the following:

- Degree of secret sharing polynomial ($k-1$)
- Prime number (p)
- List of Dedicated Aggregators participating in spatio-temporal aggregation (DA_{List})
- List of Smart Meters (SM_{List})
- Billing Tariff Type (Flat rate/Cumulative/ Time Of Use)

In addition, each DA computes the basis polynomials ($\delta_j(x)$) (eq. 4.1) by utilizing the DA_{List} as follows:

$$\delta_j(x) = \prod_{\substack{l=1 \\ l \neq j}}^{k} \frac{x-l}{j-l} \tag{4.1}$$

4.4.2 Share Creation Phase

In order to generate shares for its corresponding instantaneous reading ($R_{i,t}$), each SM (SM_i) selects a random polynomial $F_{i,t}(x)$ (eq. 4.2) of degree ($k-1$) such that the constant ($\alpha_{i,0,t}$) represents the instantaneous meter reading ($R_{i,t}$) at that instance. For $d \geq 1$, the SM_i selects the coefficients $\alpha_{i,d,t}$ of the polynomial randomly from $\mathbb{Z}_p \setminus \{0\}$. The instantaneous reading and the selected polynomial by a specific SM are independent of other SMs. Each SM_i creates a share for DA_j by utilizing its identity from the DA_{List} (eq. 4.3).

$$F_{i,t}(x) = \sum_{d=1}^{k-1} \alpha_{i,d,t}\ x^d +\ R_{i,t}\ \ (mod\ p) \tag{4.2}$$

$$(share_j)_{i,t} = F_{i,t}(j) \tag{4.3}$$

Each Smart Meter (SM_i) distributes the shares ($(share_j)_{i,t}$) to the corresponding DAs participating in the spatio-temporal aggregation.

4.4.3 Spatial Aggregation Phase

In this phase, the DAs in the DA_{List} collaboratively compute the spatially aggregated reading across SMs in the SM_{List} for a given instance of time (t). Each DA is responsible for aggregating the received shares in spatial registers ($SR_{j,t}$) (eq. 4.4). The spatially aggregated reading for a given time instance across the given set of SMs (SM_{List}) must belong to the field $\mathbb{Z}_p$ due to requirement of Shamir's Secret Sharing Scheme.

$$SR_{j,t} = \sum_{i=1}^{m} (share_j)_{i,t} \tag{4.4}$$

After all the shares are received from the given set of SMs in the SM_{List}, the DAs employ SMPC in order to computed spatially aggregated reading across SMs. Note: Only spatial registers are exchanged between the DAs. The DAs reconstruct a polynomial of degree $(k-1)$ (eq. 4.5) by utilizing same set of basis polynomials (eq. 4.1). The reconstructed polynomial is same across all DAs as they work on same set on input parameters (spatial registers (eq. 4.4) and basis polynomials (eq. 4.1)). The reconstructed polynomial is represented in the following equation:

$$G_t(x) = \sum_{j=1}^{n} SR_{j,t}\ \delta_j(x) = \sum_{d=0}^{k-1} \beta_{d,t} x^d \ \ (mod \ \ p) \tag{4.5}$$

The reconstructed polynomial ($G_t(x)$) is reported to the ESP. Post reporting, the spatial register (SR_j) is reinitialized to zero (since shares are added across SMs). The spatially aggregated reading across a given set of SMs is derived by ESP (eq. 4.6) by solving the reconstructed polynomial for $x = 0$.

$$G_t(0) = \beta_{0,t} = \sum_{i=1}^{m} R_{i,t} \tag{4.6}$$

4.4.4 Temporal Aggregation Phase

The phase can be sub-classified into the following three categorizes based on the type of billing tariff applied:

1. Time of Use Billing Tariff

2. Flat Rate Billing Tariff

3. Cumulative Billing Tariff

Note: The temporal aggregation is performed by utilizing the same shares that arrive for spatial aggregation. Hence the spatial and temporal aggregation frameworks can be implemented in parallel without adding overhead on the SMs.

Time of Use Billing Tariff

In this phase, the DAs in the DA_{List} collaboratively compute the aggregated bill for the SMs in the SM_{List} for a given period of time (T). Each DA is responsible for aggregating the received shares in the temporal registers (eq. 4.7) respectively. Here $price_t$ represents price at that given instance (Time of Use Billing Tariff).

$$TR_{i,j} = \sum_{t=1}^{T} (share_j)_{i,t} \times price_t \tag{4.7}$$

The reconstruction of polynomial is similar to (eq. 4.5) but utilizes the temporal registers (eq. 4.8). As the shares have been already multiplied with the price for the given instance ($price_t$), the computed result (eq. 4.9) represents the bill ($Bill_i$) for the given SM_i.

$$H_{i,T}(x) = \sum_{j=1}^{n} TR_{i,j}\ \delta_j(x) = \sum_{d=0}^{k-1} \gamma_{d,i,T}\ x^d\ \ (mod\ \ p) \tag{4.8}$$

$$Bill_i = H_{i,T}(0) \tag{4.9}$$

The computed bill is reported to the ESP. Post reporting, the temporal registers are reinitialized to zero (since shares are added for a specific SM over a period of time).

Flat Rate Billing Tariff

In this phase, the DAs in the DA_{List} collaboratively compute the total aggregated reading across SMs in the SM_{List} for a given period of time (T). Each DA is responsible for storing the received shares in the temporal registers (eq. 4.10).

$$TR_{i,j} = \sum_{t=1}^{T} (share_j)_{i,t} \tag{4.10}$$

The reconstruction of polynomial is similar to (eq. 4.5) but utilizes the temporal registers (eq. 4.11).

$$H_{i,T}(x) = \sum_{j=1}^{n} TR_{i,j}\ \delta_j(x) = \sum_{d=0}^{k-1} \gamma_{d,i,T}\ x^d\ \ (mod\ \ p) \tag{4.11}$$

Each DA computes total consumption for SM_i for a given period of time (T) by solving the reconstructed polynomial for $x = 0$ (eq. 4.12).

$$H_{i,T}(0) = \gamma_{0,i,T} = \sum_{t=1}^{T} R_{i,t} \tag{4.12}$$

The DAs multiply $price_{unit}$ received from the ESP and, the bill ($Bill_i$) is computed for the corresponding smart meter.

$$Bill_i = H_{i,T}(0) \times price_{unit} \tag{4.13}$$

The computed bill is reported to the ESP. Post reporting, the temporal registers are reinitialized to zero (since shares are added for a specific SM over a period of time).

Cumulative Billing Tariff

In this phase, the DAs in the DA_{List} collaboratively compute the total aggregated reading across SMs in the SM_{List} for a given period of time (T). Each DA is responsible for storing the received shares in the temporal registers (eq. 4.14).

$$TR_{i,j} = \sum_{t=1}^{T} (share_j)_{i,t} \tag{4.14}$$

The reconstruction of polynomial is similar to (eq. 4.5) but utilizes the temporal registers (eq. 4.15).

$$H_{i,T}(x) = \sum_{j=1}^{n} TR_{i,j}\ \delta_j(x) = \sum_{d=0}^{k-1} \gamma_{d,i,T}\ x^d \tag{4.15}$$

Each DA computes total consumption for SM_i for a given period of time (T) by solving the reconstructed polynomial for $x = 0$ (eq. 4.16).

$$H_{i,T}(0) = \gamma_{0,i,T} = \sum_{t=1}^{T} R_{i,t}\ \ (mod\ \ p) \tag{4.16}$$

The computed bill for the SM_i is represented in equation 4.17, where max represents the last interval corresponding to the total consumption and the $price_{max}$ is the price associated with the last interval. Whereas, Con_{int} represents the consumption interval and $price_{int}$ represents the corresponding price associated to it.

$$\begin{aligned} Bill_i = & \left(\sum_{int=1}^{max-1} Con_{int} \times price_{int} \right) \\ & + \left(\left(H_{i,T}(0) - \sum_{int=1}^{max-1} Con_{int} \right) \times price_{max} \right) \end{aligned} \tag{4.17}$$

The computed bill is reported to the ESP. Post reporting, the temporal registers are reinitialized to zero (since shares are added for a specific SM over a period of time).

4.4.5 Billing Phase

As the SMs have reported the metering data to the DAs over a period of time (T), the DAs process the data to determine the bill for each SM and later forward it to the ESP. In the Billing Phase, the ESP reports the bill to the corresponding customer associated with the SM (SM_i) in the SM_{List} (typically done at the end of a month). Since the bill is computed over a period of time, the granularity of the instantaneous reading is protected from both the DAs and the ESP.

4.5 Overall Framework Design

The flowchart in Figure 4.2 illustrates our proposed privacy safeguarding framework, showcasing the spatio-temporal aggregation flow over a specific time period (T). Additionally, it highlights the entities involved in each phase. At $t = 1$, the ESP communicates the initialization parameters to the SMs and the DAs. In the Share Creation Phase, the SMs utilize the initialization parameters to create shares and distribute them to the corresponding DAs. In the Spatial Aggregation Phase, the DAs employ SMPC to compute the spatially aggregated reading across the given set of SMs. This reading is reported to the ESP so that it can provide grid functionalities. The Temporal Aggregation Phase also takes place on the same set of shares for a given time instance t. If the instantaneous time is equal to the time period (T), it reports the computed bill to the customer for the corresponding pre-selected tariff. Otherwise, the subsequent set of shares are captured from the SMs for the next time instance.

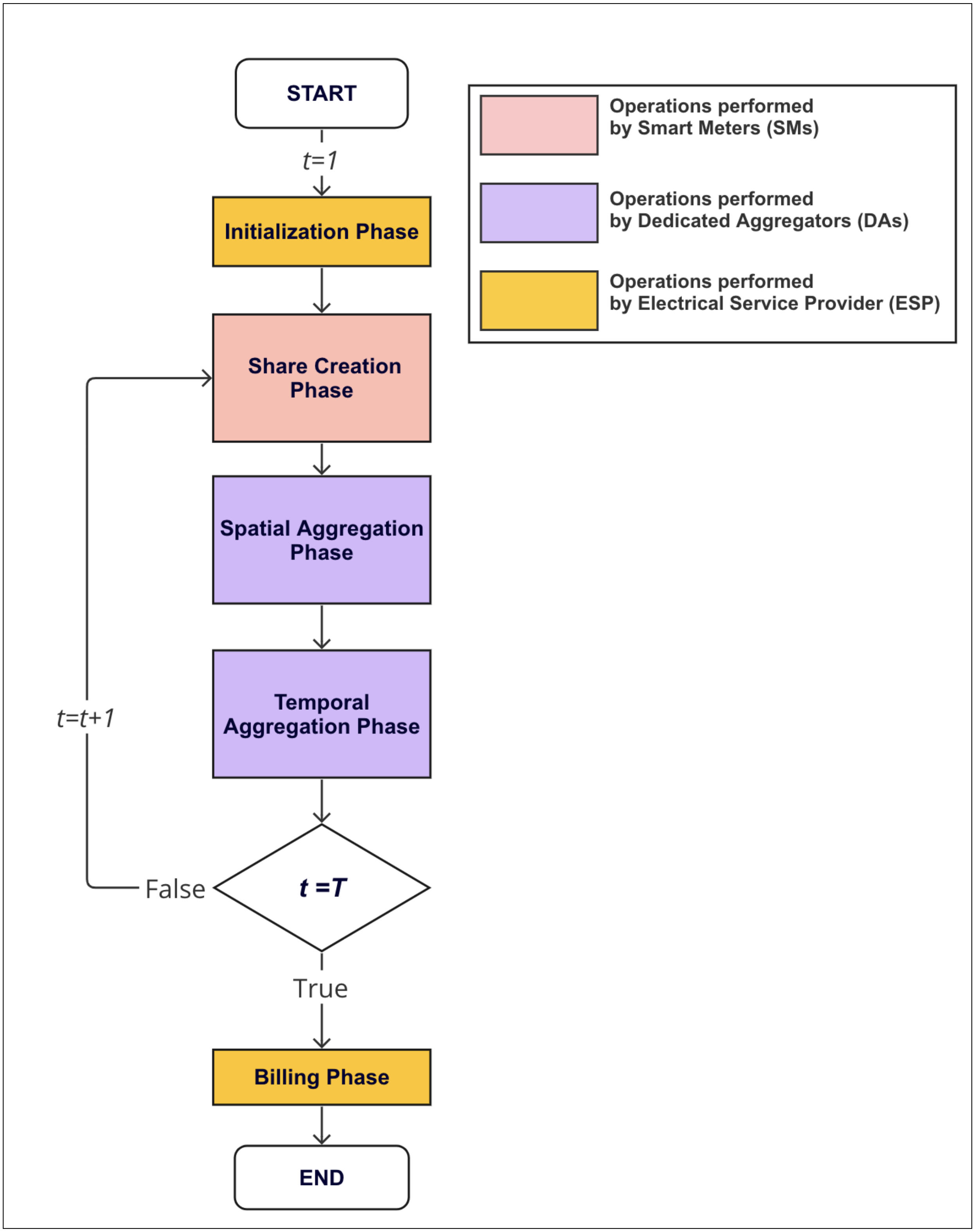

START
t=1
Initialization Phase
Share Creation Phase
Spatial Aggregation Phase
Temporal Aggregation Phase
t =T
False
t=t+1
True
Billing Phase
END
Operations performed by Smart Meters (SMs)
Operations performed by Dedicated Aggregators (DAs)
Operations performed by Electrical Service Provider (ESP)

Figure 4.2: Flowchart of our Privacy Safeguarding Framework in a Semi-Honest Setting

Figure 4.3 provides a detailed overview of the step-by-step process involved in each phase of our privacy safeguarding framework for a semi-honest setting over a given period of $T = 2$. It highlights the granular steps associated with the framework and outlines the various tasks involved. Furthermore, it describes the sequential interactions among entities involved in each phase of our proposed framework.

In the first phase, the ESP initiates the SMs and DAs by transmitting the initialization parameters. During the Share Creation Phase, the SMs utilize these initialization parameters to generate shares representing their instantaneous meter readings. These shares are then distributed to the participating DAs for spatio-temporal aggregation.

During the Spatial Aggregation Phase, each DA receives shares for a specific time instance from the SMs and aggregates them in the spatial register. Simultaneously, the shares are updated in the temporal registers associated with each SM. The DAs utilize SMPC to reconstruct the polynomial for spatial aggregation. The reconstructed polynomial for the corresponding time instance ($t = 1$) is then reported to the ESP. Once reported, the spatial registers are reset to zero. The ESP uses the reconstructed polynomial to obtain the spatially aggregated reading. This process continues until the current time instance reaches the total time period ($t = T$).

When $t = T$, temporal aggregation takes place among the same set of DAs using SMPC. The DAs reconstruct the polynomial for temporal aggregation and calculate the bill based on the pre-selected tariff. Subsequently, the bill is transmitted to the ESP. After reporting, the temporal registers are reset to zero. In the Billing Phase, the ESP communicates the computed bill to the customer associated with the respective SM.

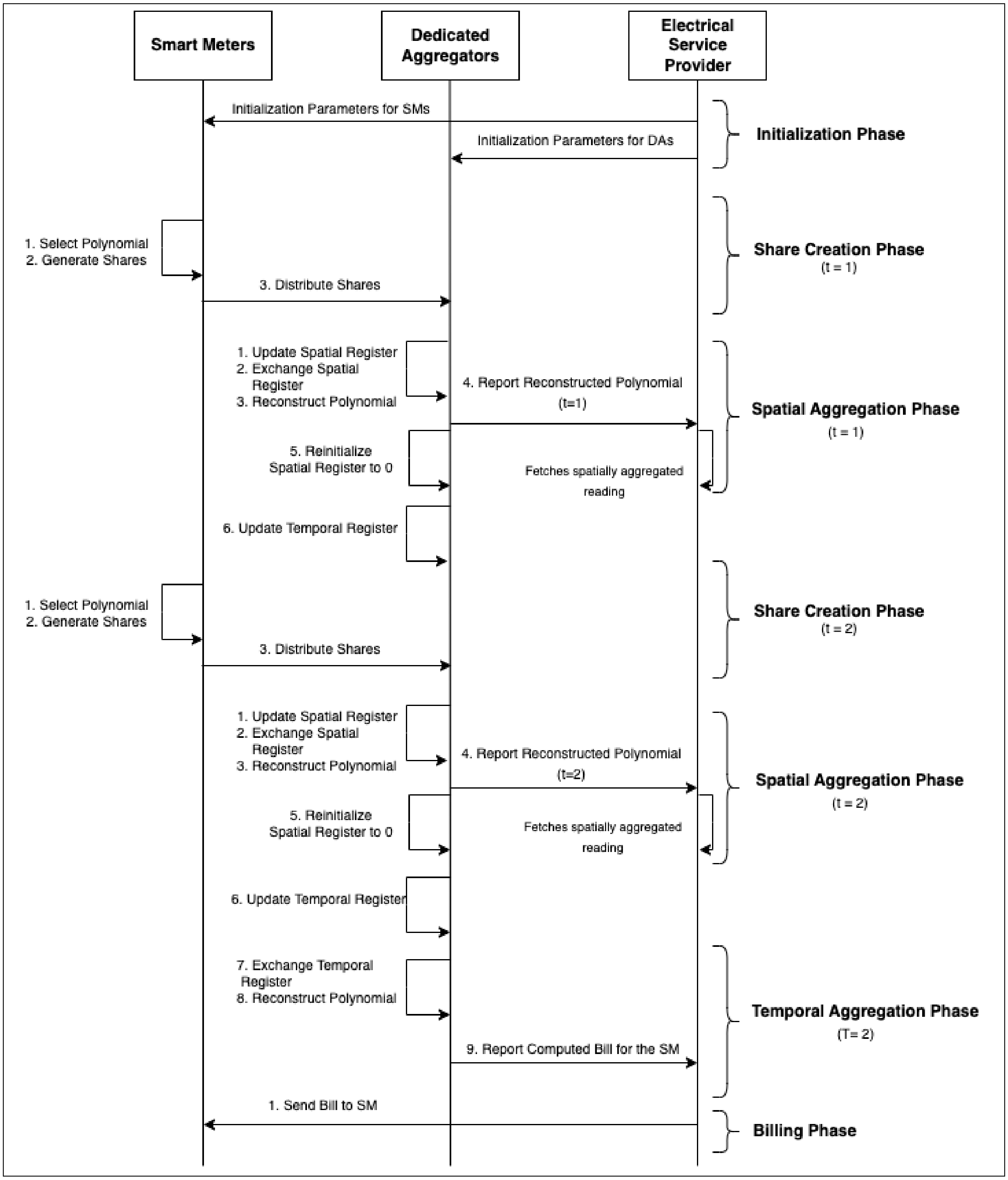

Smart Meters
Dedicated Aggregators
Electrical Service Provider
Initialization Parameters for SMs
Initialization Parameters for DAs
Initialization Phase
1. Select Polynomial
2. Generate Shares
3. Distribute Shares
Share Creation Phase
(t = 1)
1. Update Spatial Register
2. Exchange Spatial Register
3. Reconstruct Polynomial
4. Report Reconstructed Polynomial (t=1)
5. Reinitialize Spatial Register to 0
Fetches spatially aggregated reading
Spatial Aggregation Phase
(t = 1)
6. Update Temporal Register
1. Select Polynomial
2. Generate Shares
3. Distribute Shares
Share Creation Phase
(t = 2)
1. Update Spatial Register
2. Exchange Spatial Register
3. Reconstruct Polynomial
4. Report Reconstructed Polynomial (t=2)
5. Reinitialize Spatial Register to 0
Fetches spatially aggregated reading
Spatial Aggregation Phase
(t = 2)
6. Update Temporal Register
7. Exchange Temporal Register
8. Reconstruct Polynomial
9. Report Computed Bill for the SM
Temporal Aggregation Phase
(T= 2)
1. Send Bill to SM
Billing Phase

Figure 4.3: A Detailed Workflow of our Privacy Safeguarding Framework in a Semi-Honest Setting

4.6 Summary

In this chapter, we proposed a distributed framework in a semi-honest setting to address two identified research questions. The framework enabled the sharing of spatio-temporal metering data with the ESP via DAs in a privacy-preserving manner while keeping the computational overhead low for the smart meters. However, the framework does not consider integrity attacks on metering data, given that it is being outsourced to aggregating entities owned by trusted third parties. In our next chapter, we extend our proposed framework's threat model to a malicious setting with a dishonest majority of aggregating entities.

Chapter 5

A Privacy Safeguarding Framework: Malicious Setting

5.1 Introduction

In this chapter, we will be focusing on the following research questions (RQs):

- **RQ3: How can the Electrical Service Provider (ESP) incorporate integrity verification of spatially aggregated metering data without violating customer privacy?**
- **RQ4: How can the Electrical Service Provider (ESP) incorporate integrity verification of temporally aggregated metering data without violating customer privacy?**

In chapter 4, we proposed a distributed framework that enables the reporting of smart metering data from the smart meters (SMs) to the ESP through dedicated aggregators (DAs) in a privacy-preserving manner. The framework utilizes Shamir's Secret Sharing Scheme and secure multiparty computation (SMPC) to compute the spatio-temporal metering data. However, the threat model considered was semi-honest, which is not a realistic assumption given that the metering data is outsourced to aggregating entities operated by trusted third parties. In order to enhance the applicability of our framework in real-world scenarios, we extend our previous framework to accommodate both semi-honest and malicious threat models. We conduct a comprehensive analysis of existing literature works focusing on Aggregation-based Frameworks, taking into account the integrity of metering data.

The Aggregation-based Frameworks are categorized based on the architectural model as follows [11, 13, 14, 22, 23, 25, 26, 31, 37, 40, 48, 53, 55, 56, 86, 100, 102, 103, 104]: In-Network Aggregation-based Frameworks, Centralized Aggregation-based Frameworks, and, Distributed Aggregation-based Frameworks.

In-Network Aggregation-based Frameworks: These frameworks expect the metering data to be collected with the help of intermediate SMs within the communication network. We critically analyze the most closely related In-Network Aggregation-based Frameworks [14, 25, 40, 53, 100].

In [14], an In-Network Aggregation-based Framework is introduced, supporting spatio-temporal aggregation. It utilizes the Pedersen Commitment Scheme [79] for bill verifiability and the Paillier Encryption Scheme [77] for computing spatio-temporal metering data. Despite its advantages, the interconnection of SMs in this framework leads to a significant computational overhead. Additionally, it provides support for both Flat Rate and Dynamic Billing, assuming a semi-honest threat model. In [25], an alternative In-Network Aggregation-based Framework employs secure multiparty computation (SMPC) and Homomorphic Encryption-based Schemes for collecting spatial and temporal meter data. However, this framework imposes a high computational overhead on resource-constrained SMs, lacks support for dynamic billing, and assumes a semi-honest threat model. To address the issue of SM crashes during the aggregation process, [40] proposes a framework capable of tolerating a certain threshold of SM failures. It ensures customer privacy by utilizing Shamir's Secret Sharing Scheme and operates under an honest-but-curious model. Nevertheless, the requirement for interconnecting SMs increases the architectural complexity of the system. In [53], a framework is introduced to ensure the integrity of spatial metering data using a homomorphic signature scheme with batch verification. However, this framework lacks integrity verification for temporal metering data and incurs a high computational overhead on resource-constrained SMs. Researchers in [100] explore a framework that combines a Fully Homomorphic Encryption-based (FHE) Scheme with SMPC to report spatially aggregated data to the ESP within an acceptable timeframe. This framework, while offering valuable capabilities, does not support temporal aggregation and exhibits a high computational overhead on the SMs. Additionally, complete group management is required for SM addition and/or removal. While the framework ensures integrity verification for spatial metering data, it does not address integrity verification for temporal metering data.

Centralized Aggregation-based Frameworks: These frameworks expect the metering data to be collected, processed, and aggregated by a centralized entity known as a dedicated aggregator

(DA). This centralized entity is typically managed by a Trusted Third Party (TTP) that performs the computation and forwards the aggregated result to the ESP. We critically analyze the most closely related Centralized Aggregation-based Frameworks [11, 13, 22, 23, 26, 31, 37, 48, 55, 56].

In [13], the authors presented a Centralized Aggregation-based Frameworks that utilizes anonymization techniques to eliminate identifiable information from the instantaneous reading. The anonymization process is carried out by a TTP. This framework operates under a semi-honest setting and does not incorporate billing functionality. The authors of [22] introduce a Centralized Aggregation-based Scheme that prioritizes dynamic billing and addresses customer privacy concerns. They propose a Chameleon hash function to ensure integrity verification of spatial and temporal metering data. Despite these advancements, the framework is susceptible to potential single points of compromise. In their investigation presented in [23], the authors explore an anonymization-based technique aimed at preserving customer privacy. The framework relies on a DA to anonymize the data before transmitting it to the ESP. However, the framework lacks mechanisms for verifying the integrity of metering data and does not offer support for billing functionality. In their work presented in [26], the authors introduce a privacy-enhanced aggregation-based framework designed to withstand internal attacks. The SM utilize blinding factors to conceal their detailed metering data and enable batch verification of spatial metering data. It is important to note that the framework does not incorporate integrity verification for temporal metering data. In [31], the authors propose a framework that supports spatio-temporal aggregation by employing hash and executive OR operations. They also provide a solution for secure billing, but the framework fails to support dynamic billing functionalities. The paper [37] presents a Centralized Aggregation-based Framework that utilizes blockchain technology to safeguard customer privacy and maintain the integrity of metering data. Additionally, the authors incorporate the use of Bloom filters for authentication purposes. In [48], the authors propose a fog-enabled privacy-preserving aggregation-based framework that supports fault tolerance. The framework employs Boneh-Goh-Nissam (BGN) cryptosystem and employs Elliptic Curve Digital Signature Algorithm (ECDSA) for authentication. In [55], the authors propose a Centralized Aggregation-based Framework that employs Homomorphic Encryption to support spatial aggregation and incorporates adaptive key evolution technique to ensure user session keys are secure. The research conducted in [56] introduces a novel approach using dual blockchain technology to safeguard customer privacy and facilitate the secure sharing of metering data with the ESP. The framework utilizes a private blockchain to link the actual identity of the SM with a pseudonym. Furthermore, secure signature mechanisms and an identity-based proxy re-encryption scheme are implemented to ensure authentication and aggregation capabilities, respectively In the study presented in [11], the author puts forward a Centralized Aggregation-based Framework that facilitates the aggregation of spatio-temporal metering data while simultaneously preserving customer privacy. The framework incorporates a Paillier Encryption-based Scheme in conjunction with the Elliptic Curve Diffie Hellman Ephermeral (ECDHE) algorithm. The framework, operates on the assumption of a semi-honest threat model and lacks support for dynamic billing functionality.

Distributed Aggregation-based Frameworks: These frameworks utilize multiple dedicated entities to perform collection, processing and aggregation of metering data. We critically analyze the most closely related Distributed Aggregation-based Frameworks [86, 102, 103, 104].

In their work presented in [86], the authors aim to overcome the limitations of [87] by introducing privacy-preserving nodes (PPNs) between the SMs and the ESP. They propose a framework that utilizes Shamir's Secret Sharing Scheme to collect metering data from the SMs and leverages the homomorphic properties of the scheme to enable spatial and temporal aggregation. The responsibility of centralized data reconstruction from the PPNs lies with the ESP, which makes the framework vulnerable to single points of compromise. To address the limitations of [86], we propose a framework [102] that incorporates Shamir's Secret Sharing Scheme for SMs and utilizes SMPC in conjunction with the DAs employing and the ESP. The framework is capable of providing spatio-temporal metering data in a privacy-preserving fashion but lacks to provide support for Dynamic Billing such as Time of Use Billing Tariff. To integrate the dynamic billing functionality, we proposed a framework in a semi-honest setting [103]. We utilized the same set of shares to support the temporal aggregation of a metering data in a privacy preserving manner, thus imposing minimal additional computational overhead on the SMs.

The existing Distributed Aggregation-based Frameworks in the literature assume a semi-honest threat model and do not consider metering data integrity in the design. This is not a realistic assumption, given that the DAs may belong to third parties that are not under the purview of the ESP at all times. Although malicious adversarial models for the smart grid have been studied in the literature [22, 26, 31, 48, 56], but they adopt a centralized design. Additionally, most of the frameworks do not not support dynamic billing functionality.

To summarize the aforementioned aggregation-based frameworks, the existing solutions have one or more of the following limitations:

- Prone to single points of compromise due to dependency on a centralized entity for aggregation [11, 13, 22, 23, 26, 31, 37, 48, 55, 56]
- High computational overhead on resource-constrained SMs [11, 13, 14, 25, 37, 40, 53, 55, 56, 100]

- Lack of support for dynamic billing integration while preserving customer privacy [11, 13, 23, 25, 26, 31, 37, 40, 48, 54, 55, 56, 100]
- Lack of integrity verification of spatial and temporal metering data [11, 13, 23, 25, 26, 31, 37, 40, 48, 54, 55, 56, 100]

We aim to address these limitations through our distributed framework in a malicious setting that can check integrity verification of spatio-temporal metering data while preserving customer privacy. Since the majority of the computing load is shifted to the DAs, the framework is lightweight in terms for the resource-constrained SMs.

5.2 System Model

This section describes the architectural model (Fig. 5.1), threat, and billing model associated with our proposed framework.

5.2.1 Architectural Model

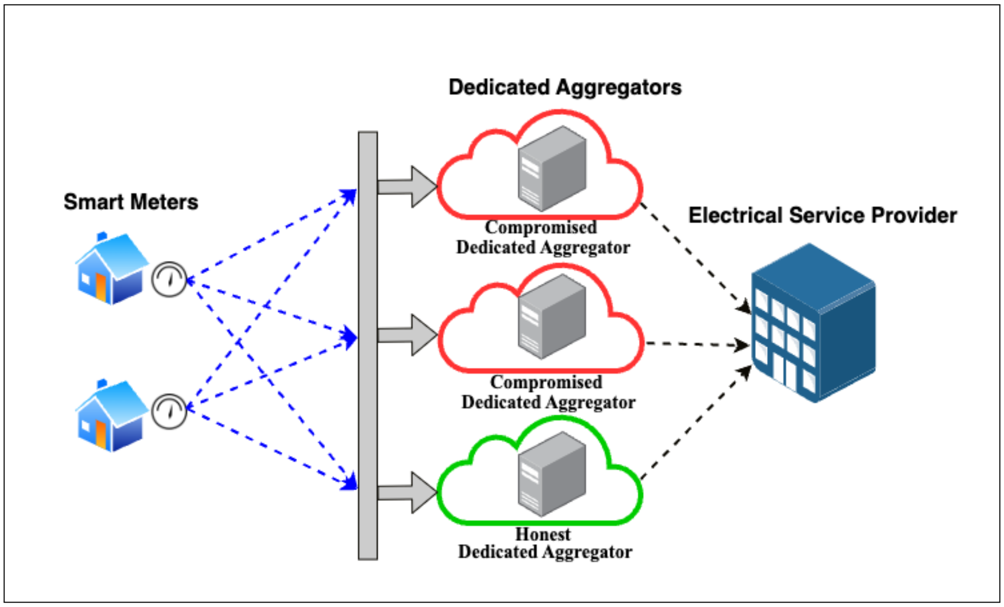


Figure 5.1: A Privacy Safeguarding Framework: Malicious Setting

Smart Meters (SMs):

A smart meter is a resource-constrained device that is installed in the customer's domain. It is capable of reporting both high-frequency and low-frequency metering data to the ESP through the DAs. The SMs can perform cryptographic operations, such as generating shares for corresponding instantaneous readings, utilizing the parameters provided by the ESP. We assume that each customer has one SM installed, and the instantaneous readings (consumption) are independent of each other. Each SM is assigned a unique identifier denoted by i, where $i \in \{1, 2, \ldots, m\}$; $m =$ represents the total number of SMs. Note: The SMs do not interact with each other, their interactions are solely with the DAs.

Dedicated Aggregators (DAs):

A dedicated aggregator is an intermediate device positioned between the SMs and the ESP. Unlike the SMs, it possesses high computational capabilities. The DAs receive their initialization parameters from the ESP, similar to the SMs. The main responsibility of the DAs is to perform spatio-temporal aggregation employing SMPC. Each DA is assigned a unique identifier, denoted by j, where $j \in \{1, 2, \ldots, n\}$; n represents the total number of DAs participating in the spatial and temporal aggregation. Each DA maintains the following types of memory registers:

- Spatial Register ($SR_{j,t}$): This register stores the aggregated shares received by DA_j from m SMs for a given time instance (t).
- Spatial Commitment Register (SCR_t): Stores the aggregated commitments from m SMs received by DA_j for a given time instance (t)
- Temporal Register ($TR_{i,j}$): This register stores the aggregated shares received by DA_j from SM_i over a time period (T).
- Temporal Commitment Register (TCR_i) Stores the aggregated commitments received by DA_j from SM_i over a time period (T).

Each DA is responsible for reporting the output of the SMPC (the reconstructed polynomial) to the ESP. Our research aligns with the NIST guidelines that emphasize the use of a redundancy strategy to enhance the resilience of the smart grid (Section 1.3) [32, 81]. DAs can be deployed on-premises, cloud-based, and/or in a hybrid manner (both on-premises and cloud) [7, 12, 34, 61, 69, 78].

Electrical Service Provider (ESP):

The responsibility of communicating the initialization parameters to the SMs and DAs lies with the ESP. The Initialization Phase is assumed to be secure against active-passive attacks. Additionally, the ESP is assumed to possess historic spatio-temporal metering data, which allows for approximate estimations in the event of malicious modifications to the metering data. The ESP is also responsible for configuring the billing tariffs, including Flat Rate, Cumulative, and Time of Use Billing Tariffs, as well as verifying the integrity of the spatio-temporal metering data. The primary focus of the framework is the integrity verification of spatio-temporal data in a malicious setting, specifically addressing scenarios involving the compromise of entities, with a particular emphasis on cases where the majority of DAs are dishonest.

5.2.2 Threat Model

As we have seen in the previous chapter (Chapter 4), a distributed aggregation-based privacy-preserving framework was proposed in a semi-honest setting. However, this assumption is not realistic given that the metering data is outsourced to aggregation entities owned by Trusted Third Parties. In this chapter, we extend the framework to a malicious setting, which is more realistic (Fig. 5.1). The malicious adversary can compromise up to *(n-1)-out-of-n* DA(s) (theoretical assumption enforced) and has access to granular shares, commitments, corresponding memory registers, as well as initialization parameters from the ESP after compromising the DA(s). The goals of the malicious adversary are as follows:

- Disrupt grid and billing functionalities
- Breach privacy of the customers

Disrupting grid and billing functionality can be achieved by modifying metering data and/or commitments. The adversary can modify shares, commitments, or registers associated with compromised DAs before SMPC occurs. The adversary can also modify reconstructed polynomials and commitments related to a given set of DAs after SMPC. These modifications would affect the ESP's decision-making process and may result in an outage. The adversary can also violate the customers' privacy by linking the instantaneous reading with the identity of the SM. The remaining entities, SMs and ESP, are considered semi-honest. The SMs can try to infer high-frequency metering data of other SMs. We assume that the exchanges between the entities (SM-DA, DA-DA and, DA-ESP) are protected from active adversaries through secure connections. Additionally, the internal clocks are in sync with each other, ensuring synchronization within the framework. The main aim of our proposed framework is to detect modifications to the metering data for each time instance while preserving customer privacy.

5.2.3 Billing Model

The proposed framework is capable of supporting different types of billing tariffs, such as Flat Rate, Cumulative, and Time Of Use Billing Tariff. The details of these different types of billing tariffs are highlighted in Section 4.2.3.

5.3 Cryptographic Prerequisites

5.3.1 Configuration of Commitment-based Scheme

As we have seen in Section 3.2, a Commitment-based Scheme enables the Prover (P) to commit (eq. 5.1) to its corresponding secret (S) and later reveal the secret to the Verifier (V). We extend the 2-party commitment method explained in Section 3.2 to a distributed setting in our proposed framework, where SMs generate the commitments (proving entities) and later send them to the ESP (verifier) via DAs (intermediate entities).

$$C_{i,t} = g^{R_{i,t}} \;\; h^{r_{i,t}} \; (mod\, p) \tag{5.1}$$

- The DA(s) and/or the ESP cannot derive any information about the instantaneous reading ($R_{i,t}$) related to a SM (SM_i) from the commitment ($C_{i,t}$).
- The commitment ($C_{i,t}$) generated from a given secret ($R_{i,t}$) cannot reveal a different secret ($R'_{i,t}$).
- When two different commitments ($C_{1,t}, C_{2,t}$) generated from two different readings ($R_{1,t}$, $R_{2,t}$) are multiplied, the corresponding commitment (C_{add}) is the commitment for the sum of two readings ($R_{1,t}$)+, ($R_{2,t}$). This property of commitments makes it uniquely suitable for our scheme, unlike other integrity verification schemes [76, 100] based on hashes and signatures.

- When a commitment ($C_{i,t}$) generated from a given secret ($R_{i,t}$) is raised to a constant power (c), then the corresponding commitment (C_{mul}) is equal to the commitment value multiplied by itself that many number of times ($c \times R_{i,t}$). This property of commitments makes it uniquely suitable for our scheme in terms of Time of Use (TOU) billing tariff.

The details of the commitment scheme is terms of spatial and temporal aggregation is presented in Section 5.4.

5.3.2 Configuration of Shamir's Secret Sharing Scheme (SSS) with a constraint

The Shamir's Secret Sharing Scheme [91] is utilized by each resource-constrained SM to break its instantaneous reading ($R_{i,t}$) into n number of shares. These shares are later distributed to the DAs to compute spatial and temporal aggregation by employing SMPC. The shares are divided in such a way that given a subset of k shares, the DAs are able to reconstruct the instantaneous reading. Hence the scheme is also known as a (k, n) threshold scheme (k and n are positive integers) where k represents the minimum number of shares required to reconstruct instantaneous reading ($R_{i,t}$). However, by doing so, the DAs can learn about the granular instantaneous reading of each SM, thereby breaching the privacy of the customers associated with the share. To protect the privacy of the customers, we leverage the homomorphic properties of Shamir's Secret Sharing Scheme such as follows:

- Additive Property: The corresponding shares of the instantaneous readings ($R_{1,t}$) and ($R_{2,t}$) can be added across different SMs (SM_1 and SM_2) if the degree of secret sharing polynomial is the same. Thus, the resultant share obtained represents the share of the total readings ($R_{1,t} + R_{2,t}$). This additive property of the Shamir's Secret Sharing Scheme helps the DAs to compute spatially aggregated reading in a privacy-preserving fashion.

- The corresponding shares of the instantaneous reading ($R_{i,t}$) can be multiplied by a constant (c) such that the resultant secret obtained represents the constant times the instantaneous reading ($c \ \times (R_{i,t})$). This property is useful for incorporating Dynamic Billing (Time of Use Billing Tariff) in a distributed and privacy-preserving fashion.

- Shamir's Secret Sharing Scheme is information theoretic secure: An adversary without given number of threshold shares cannot reconstruct the secret even with infinite time and computing capacity.

In our proposed framework, we impose a constraint ($r_{i,t}$) via ESP on the first coefficient ($\alpha_{i,1,t}$) of the secret sharing polynomial ($F_{i,t}(x)$) to include integrity verification with the secret sharing process (eq. 5.2). In the proposed framework, the SM is responsible for creating the shares for their corresponding consumption and reporting to the DAs.

$$F_{i,t}(x) = \sum_{d=2}^{k-1} \alpha_{i,d,t}\; x^{d} + (r_{i,t})\; x + \; R_{i,t} \;\; (mod\; p) \tag{5.2}$$

A more detailed explanation about how the constraint is imposed on Shamir's Secret Scheme is covered in Section 5.4.

5.3.3 Configuration of Secure Multiparty Computation

SMPC [16, 57] enables a group of DAs in our proposed framework to compute a function (spatial aggregation as well as temporal aggregation) without disclosing their individual inputs (shares received from the SMs). In SMPC, all the participating DAs jointly compute a function over their inputs such that they are only aware of the result (spatially aggregated reading and /or temporally aggregated reading) and their inputs stored in spatial ($SR_{j,t}$) and temporal ($TR_{i,j}$) registers. Since only aggregated data is shared across the DAs in the form of registers (spatial register and temporal register), the DAs cannot reconstruct the instantaneous reading with a given SM thereby preserving customer privacy. Additionally, the DAs share the aggregated result of the SMPC with the ESP for providing grid and billing functionalities. Even in that case, ESP also cannot breach the customer privacy as it receives computed spatially aggregated reading and/or temporally aggregated reading from the DAs.

The SMPC phases can be defined as follows:

- **Input preparation:** Each DA prepares its input by aggregating the granular shares received from the SMs in spatial and temporal registers. This ensures that their inputs are ready to be utilized in the computation of spatio-temporal aggregation without revealing the input (granular shares) to the other DAs.
- **Computation:** The DAs jointly compute spatially aggregated reading over a given set of SMs for a given instance of time (utilizing spatial registers). Additionally, the DAs compute temporally aggregated reading for a given set of SMs over a given period of time (utilizing temporal registers)(Fig. 5.2).

- **Output reconstruction:** The DAs reconstruct the output of the computation (spatially aggregated reading and temporally aggregated reading) without revealing their inputs (granular shares) to each other.

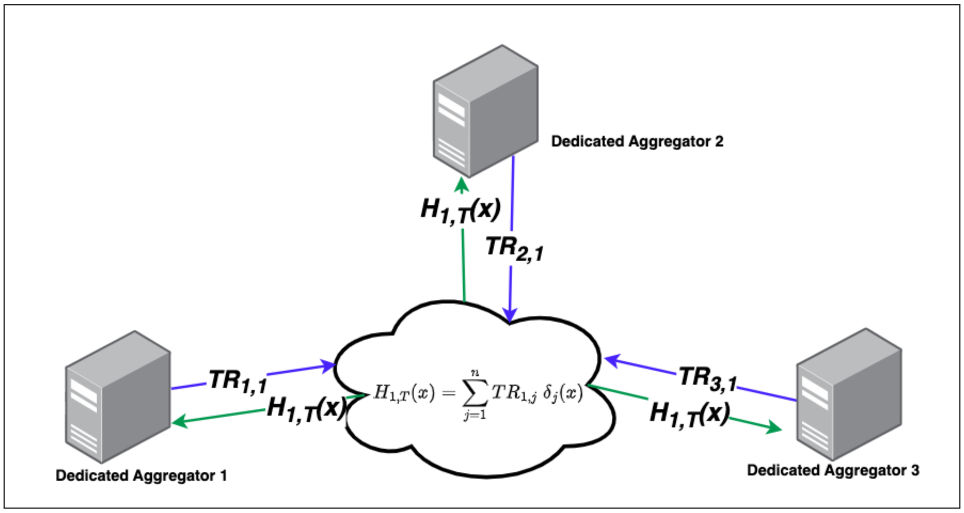


Figure 5.2: Secure Multiparty Computation (Temporal Aggregation)

Our proposed framework is lightweight in terms of computational overhead on SMs, as the majority of the work is performed by the DAs. By employing Shamir's Secret Sharing Scheme, the shares generated by each SM are distributed across the DAs. This allows for the computation of spatially or temporally aggregated readings when at least k or more DAs exchange their respective registers. Therefore, SMPC helps the framework address single points of compromise. A more detailed explanation of how SMPC is employed in our proposed framework is covered in Section 5.4.

5.4 Proposed Framework

In this section, we describe the phases associated with our proposed framework in detail. The framework employs Shamir's Secret Sharing Scheme, Commitment-based Scheme and, SMPC to provide integrity verification of spatio-temporal metering data while preserving consumer privacy. Table 5.1 represents the notations used for our privacy safeguarding framework in a malicious setting. The proposed framework consists of the following phases:

Table 5.1: Table of notations: A Privacy Safeguarding Framework in a Malicious Setting

Notation	Meaning
$\delta_j(x)$	Basis Polynomial
$Bill_i$	Bill generated for i^{th} Smart Meter
$\beta_{d,t}$	d^{th} Coefficient of Reconstructed Polynomial at time t (Spatial Aggregation)
$\gamma_{d,i,t}$	d^{th} Coefficient of Reconstructed Polynomial for i^{th} Smart Meter at time t (Temporal Aggregation)
$\alpha_{i,d,t}$	d^{th} Coefficient of Secret Sharing Polynomial of i^{th} Smart Meter at time instance t
g, h	Commitment Creation parameters
$C_{i,t}$	Commitment generated by i^{th} Smart Meter at time instance t
$C_{computed,t}$	Computed Commitment by the ESP (Spatial Aggregation)
$C_{computed,i,T}$	Computed Commitment by the ESP (Temporal Aggregation) for i^{th} Smart Meter
Con_{int}	Consumption for given interval (Cumulative Tariff)
$k-1$	Degree of Secret Sharing Polynomial
ESP	Electrical Service Provider
SM_i	i^{th} Smart Meter
$R_{i,t}$	Instantaneous reading of i^{th} Smart Meter at time instance t
t	Instantaneous time
DA_j	j^{th} Dedicated Aggregator
DA_{List}	List of Dedicated Aggregators
SM_{List}	List of Smart Meters
n	Number of Dedicated Aggregators
m	Number of Smart Meters
$price_{int}$	Price for given interval (Cumulative Tariff)
$price_{max}$	Price for last interval (Cumulative Tariff)
$price_t$	Price for that given time instance (Time of Use Tariff)
$price_{unit}$	Price per unit consumption (Flat Rate Tariff)
p	Prime Number
$C_{received,t}$	Received Commitment by the ESP (Spatial Aggregation)
$C_{received,i,T}$	Received Commitment by the ESP (Temporal Aggregation) for i^{th} Smart Meter
$G_t(x)$	Reconstructed Polynomial at time instance t (Spatial Aggregation)
$H_{i,T}(x)$	Reconstructed Polynomial of i^{th} Smart Meter (Temporal Aggregation)
$F_{i,t}(x)$	Secret Sharing Polynomial of i^{th} Smart Meter at time instance t
$r_{\text{List},i}$	Seed List of i^{th} Smart Meter
$r_{i,t}$	Seed Value of i^{th} Smart Meter at time instance t
$(share_j)_{i,t}$	Share generated by i^{th} Smart Meter for j^{th} Dedicated Aggregator at time instance t
SCR_t	Spatial Commitment Register of j^{th} Dedicated Aggregator
$SR_{j,t}$	Spatial Register of j^{th} Dedicated Aggregator
TCR_i	Temporal Commitment Register of j^{th} Dedicated Aggregator for i^{th} Smart Meter
$TR_{i,j}$	Temporal Register of j^{th} Dedicated Aggregator for i^{th} Smart Meter
T	Total time

5.4.1 Initialization Phase

The ESP communicates the initializing parameters to the SMs and DAs.

The SMs receive the following:

- Degree of secret sharing polynomial $(k-1)$
- Prime number (p)
- Parameters for commitment creation (g, h)
- List of DAs participating in spatio-temporal aggregation (DA_{List})
- Seed list $(r_{\text{List},i} = \{r_{i,1}, r_{i,2}, \ldots, r_{i,T}\})$

Note: The seed list is different for each SM and T represents the total time interval.

The DAs receive the following:

- Degree of secret sharing polynomial $(k-1)$
- Prime number (p)
- List of Dedicated Aggregators participating in spatio-temporal aggregation (DA_{List})
- List of Smart Meters (SM_{List})
- Billing Tariff Type (Flat Rate/Cumulative/ Time Of Use)

Note: The tariff is communicated to the DAs only for Time of Use (Dynamic) Billing scenario. No tariff is communicated to the DAs for Flat Rate Billing and Cumulative billing. The ESP is responsible for applying the Flat Rate or Cumulative Billing Tariff on the aggregated temporal metering data from each SM in the SM_{List}.

In addition, each DA computes the basis polynomials $(\delta_j(x))$ (eq. 5.3) by utilizing the DA_{List} as follows:

$$\delta_j(x) = \prod_{\substack{l=1 \\ l \neq j}}^{k} \frac{x-l}{j-l} \tag{5.3}$$

5.4.2 Commitment and Share Creation Phase

In this phase each SM (SM_i) where $i \in \{1, 2, \ldots, m\}$ creates a commitment and shares (one for each DA) for its corresponding instantaneous reading ($R_{i,t}$) for a given instance of time (t). The commitments and shares are created by employing Pedersen Commitment Scheme [79] and Shamir's Secret Sharing Scheme [91] respectively.

Commitment Creation

In order to generate a commitment for its corresponding instantaneous reading ($R_{i,t}$) each SM (SM_i) utilizes the initialization parameters received from the ESP. The seed value ($r_{i,t}$) from the seed list ($r_{\text{List},i}$) at that given instance (t) is used as a decommitment value for commitment creation. Equation 5.4 represents creation of commitment by SM_i for a given instance of time (t).

$$C_{i,t} = g^{R_{i,t}} \;\; h^{r_{i,t}} \;\; (mod\; p) \tag{5.4}$$

Share Creation

In order to generate shares for its corresponding instantaneous reading ($R_{i,t}$), each SM (SM_i) selects a random polynomial $F_{i,t}(x)$ (eq. 5.5) of degree ($k-1$) such that the constant ($\alpha_{i,0,t}$) represents the instantaneous meter reading ($R_{i,t}$) and the first coefficient ($\alpha_{i,1,t}$) represents the seed value ($r_{i,t}$) at that instance. For $d \geq 2$, the SM_i selects the coefficients $\alpha_{i,d,t}$ of the polynomial randomly from $\mathbb{Z}_p \setminus \{0\}$. Each SM_i creates a share for DA_j by utilizing its identity from the DA_{List} (eq. 5.6).

$$F_{i,t}(x) = \sum_{d=2}^{k-1} \alpha_{i,d,t}\; x^d + (r_{i,t})\; x + \; R_{i,t} \;\; (mod\; p) \tag{5.5}$$

$$(share_j)_{i,t} = F_{i,t}(j) \tag{5.6}$$

Each SM (SM_i) distributes the commitment ($C_{i,t}$) and shares ($(share_j)_{i,t}$) to the corresponding DAs participating in the spatio-temporal aggregation.

5.4.3 Spatial Aggregation Phase

In this phase, the DAs in the DA_{List} collaboratively compute the spatially aggregated reading across SMs in the SM_{List} for a given instance of time (t). Each DA is responsible for aggregating the received commitments and shares in the spatial commitment registers (SCR_t) (eq. 5.7) and spatial registers ($SR_{j,t}$) (eq. 5.8) respectively. The spatially aggregated reading for a given time instance across the given set of SMs (SM_{List}) must belong to the field $\mathbb{Z}_p$ due to requirement of Shamir's Secret Sharing Scheme.

$$SCR_t = \prod_{i=1}^{m} C_{i,t} \tag{5.7}$$

$$SR_{j,t} = \sum_{i=1}^{m} (share_j)_{i,t} \tag{5.8}$$

After all the commitments and shares are received from the given set of SMs in the SM_{List}, the DAs employ SMPC in order to computed spatially aggregated reading across SMs. Note: Only spatial registers are exchanged between the DAs. The DAs reconstruct a polynomial of degree $(k-1)$ by utilizing same set of basis polynomials (eq. 5.3). The reconstructed polynomial (eq. 5.9) is same across all DAs as they work on same set on input parameters (spatial registers (eq. 5.8) and basis polynomials (eq. 5.3)). The reconstructed polynomial is represented in the following equation:

$$G_t(x) = \sum_{j=1}^{n} SR_{j,t}\ \delta_j(x) = \sum_{d=0}^{k-1} \beta_{d,t} x^d \ \ (mod\ p) \tag{5.9}$$

The reconstructed polynomial ($G_t(x)$) and spatial commitment register (SCR_t) are reported to the ESP. Post reporting, the spatial commitment registers (SCR_t) are reinitialized to one (since the commitments are multiplied across SMs) and, spatial register ($SR_{j,t}$) are reinitialized to zero (since shares are added across SMs).

5.4.4 Integrity Verification Phase: Spatial Aggregation

In this phase, the ESP checks the integrity verification of spatially aggregated reading by utilizing the reconstructed polynomial ($G_t(x)$) and aggregated commitments ($C_{received,t}$) received from the

DAs for a given instance of time (t). The spatially aggregated reading and aggregated seed list is derived by the ESP from the reconstructed polynomial by accessing the constant ($\beta_{0,t}$) and the first coefficient ($\beta_{1,t}$) respectively. The ESP computes the following commitment (eq. 5.10) in order to check the integrity verification of spatially aggregated reading as follows:

$$C_{computed,t} = g^{\beta_{0,t}}\ h^{\beta_{1,t}}\ (mod\ p) \tag{5.10}$$

Following are the two possible cases described below:

- **Case 1:** $C_{computed,t} = C_{received,t}$ **:** When the computed commitment matches the received commitment value from the DAs, it indicates that no modification of share(s) or spatial register(s) has occurred for the given time instance (t). The constant of the reconstructed polynomial represents the spatially aggregated value across SMs for that given instance (t). With the derived information, the ESP can provide grid functionalities (real-time monitoring and load balancing).
- **Case 2:** $C_{computed,t} \neq C_{received,t}$ **:** When the computed commitment value does not match the value received from the DAs, it indicates that share(s) or spatial register(s) have been modified/tampered during the given time instance (t). As a result, the ESP disregards the spatially aggregated reading for that time instance and alternatively estimates the total energy consumption using average values from previously validated readings.

5.4.5 Temporal Aggregation Phase

The phase can be sub-classified into the following two categorizes based on the type of billing tariff applied: 1) Time of Use Billing Tariff and 2) Flat Rate and Cumulative Billing Tariff.

Time of Use Billing Tariff

In this phase, the DAs in the DA_{List} collaboratively compute the aggregated bill across SMs in the SM_{List} for a given period of time (T). Each DA is responsible for aggregating the received commitments and shares in the temporal commitment registers (eq. 5.11) and temporal registers (eq. 5.12) respectively. Here $price_t$ represents price at that given instance (Time of Use Billing Tariff).

$$TCR_i = \prod_{t=1}^{T} C_{i,t}^{price_t} \tag{5.11}$$

$$TR_{i,j} = \sum_{t=1}^{T} (share_j)_{i,t} \times price_t \tag{5.12}$$

Flat Rate Billing Tariff and Cumulative Billing Tariff

In this phase, the DAs in the DA_{List} collaboratively compute the total aggregated reading across SMs in the SM_{List} for a given period of time (T). Each DA is responsible for storing the received commitments and shares in the temporal commitment registers (eq. 5.13) and temporal registers (eq. 5.14) respectively.

$$TCR_i = \prod_{t=1}^{T} C_{i,t} \tag{5.13}$$

$$TR_{i,j} = \sum_{t=1}^{T} (share_j)_{i,t} \tag{5.14}$$

Note: The computed bill for each SM SM_i in the SM_{List} must belong to the field $\mathbb{Z}_p$ due to requirement of Shamir's Secret Sharing Scheme.

After all the commitments and shares are received from the given set of SMs in the SM_{List}, the DAs employ SMPC in order to computed bill across SMs. Note: Only temporal registers are exchanged between the DAs. The DAs reconstruct a polynomial of degree $(k-1)$ by utilizing same set of basis polynomials (eq. 5.3). The reconstructed polynomial is same across all DAs for a given SM as they work on same set on input parameters (temporal registers and basis polynomials (eq. 5.3)). The reconstructed polynomials is represented in the following equation:

$$H_{i,T}(x) = \sum_{j=1}^{n} TR_{i,j}\ \delta_j(x) = \sum_{d=0}^{k-1} \gamma_{d,i,T}\ x^d \ \ (mod\, p) \tag{5.15}$$

5.4.6 Integrity Verification Phase: Temporal Aggregation

In case of temporal aggregation, the ESP will perform a similar procedure (as compared to integrity verification for spatial aggregation) but with polynomial $H_{i,T}(x)$ for a given SM_i. The ESP derives the total bill in case of Time of Use Billing Tariff $(\gamma_{0,i,T})$ or total metering data consumed $(\gamma_{0,i,T})$ in case of Flat Rate and Cumulative Billing Tariff, along with the aggregated seed list $(\gamma_{1,i,T})$ for a given SM (SM_i) in the SM_{List} from the reconstructed polynomial $(H_{i,T}(x))$. The ESP computes the overall commitment by utilizing the following equation:

$$C_{computed,i,T} = g^{\gamma_{0,i,T}}\ h^{\gamma_{1,i,T}}\ (mod\ p) \tag{5.16}$$

The two possible cases are described as follows:

- **Case 1:** $C_{computed,i,T} = C_{received,i,T}$ **:** If the computed commitment matches the commitment value received from the DAs, then no modification of share(s) / temporal register(s) has occurred across the given time period (T) for SM_i. Depending on the tariff selection, following computation will be performed:

  Time of Use Billing Tariff

  For the Time of Use Billing Tariff, since tariff was informed to the DAs by the ESP in the Initialization Phase, and has been incorporated in the Temporal Aggregation Phase (Time of Use Billing (Section 5.4.5)), the constant of the reconstructed polynomial represents the bill computed for a given Time of Use Billing Tariff (eq. 5.17).

  $$Bill_i = H_{i,T}(0) \tag{5.17}$$

  Flat Rate Billing and Cumulative Billing Tariff

  For Flat Rate Billing Tariff and Cumulative Billing tariff the constant of the reconstructed polynomial represents the total consumption for SM_i for that given time period T (Section 5.2.3). Since the temporally aggregated metering data for a given SM_i is verified, ESP will compute the corresponding bill by employing Flat Rate Tariff (eq. 5.18) or Cumulative Billing Tariff (eq. 5.19). Note: Here $price_{unit}$ represents price per unit. Whereas, Con_{int} represents the consumption interval and $price_{int}$ represents the corresponding price associated to it.

Additionally, max represents the last interval corresponding to the total reading and the $price_{max}$ is corresponding price associated to it.

$$Bill_i = H_{i,T}(0) \times price_{unit} \tag{5.18}$$

$$\begin{aligned} Bill_i = &\left(\sum_{int=1}^{max-1} Con_{int} \times price_{int} \right) \\ &+ \left(\left(H_{i,T}(0) - \sum_{int=1}^{max-1} Con_{int} \right) \times price_{max} \right) \end{aligned} \tag{5.19}$$

- **Case 2:** $C_{computed,i,T} \neq C_{received,i,T}$**:** If the computed commitment does not match with the commitment value received from the DAs, then it implies modification of share(s) / Temporal Register(s) has occurred during the given time period (T). The ESP discards the temporally aggregated reading or the computed bill (in case of Time of Use Billing Tariff) and uses alternative methods to generate customer billing (e.g., estimated billing [10]).

5.4.7 Billing Phase

As the SMs have reported the metering data to the DAs over a period of time (T), the DAs process the data to determine the bill (in case of Time of Use Billing) or total consumption (in case of Flat Rate or Cumulative Biling) for each SM and later forward it to the ESP. In the Billing Phase, the ESP reports the bill to the corresponding customer associated with the SM (SM_i) in the SM_{List} (typically done at the end of a month) for Time of Use Billing. Whereas, for Flat Rate and Cumulative Tariff, ESP computes the billing after receiving the total consumption from the DAs. Since the bill is computed over a period of time, the granularity of the instantaneous reading is protected from both the DAs and the ESP.

5.5 Overall Framework Design

The depicted flowchart in Figure 5.3 presents our proposed privacy-preserving framework in a malicious setting. It showcases the process of spatio-temporal aggregation and the corresponding integrity verification flow across a specific time period (T). Furthermore, it provides an overview of the entities involved in each phase. At $t = 1$, the ESP communicates the initialization parameters to the SMs and the DAs. In the Commitment and Share Creation Phase, the SMs utilize the initialization parameters to create shares and commitments and distribute them to the corresponding DAs. In the Spatial Aggregation Phase, the DAs employ SMPC to compute the reconstructed polynomial related to spatial aggregation. The reconstructed polynomial from each DA is reported to the ESP along with the spatial commitment registers so that it can check the integrity of the spatially aggregated metering data and provide grid functionalities accordingly (Integrity Verification Phase: Spatial Aggregation). The Temporal Aggregation Phase also takes place on the same set of shares and commitments for a given time instance t. If the instantaneous time is equal to the time period (T), it reports the reconstructed polynomial related to temporal aggregation (by employing SMPC) along with the temporal registers for each SM to the ESP so that it can check the integrity verification of temporally aggregated data (Integrity Verification Phase: Temporal Aggregation) and report the corresponding bill for the customer (Billing Phase). Otherwise, the subsequent set of shares are captured from the SMs for the next time instance.

Figure 5.4 presents a detailed workflow that illustrates the step-by-step process of our privacy safeguarding framework in a malicious setting for a given time period of $T = 2$. It highlights the different tasks involved in our framework and outlines the sequential interactions among entities in each phase. The first phase involves the ESP initializing the SMs and the DAs by communicating the initialization parameters. In the Commitment and Share Creation Phase, the SMs utilize these parameters to generate commitments and shares for their instantaneous meter readings. These shares and commitments are then distributed to the DAs participating in the spatio-temporal aggregation computation. During the Spatial Aggregation Phase, each DA aggregates the received shares and commitments for a specific time instance in the spatial register and spatial commitment register, respectively. Simultaneously, the shares and commitments are updated in the temporal registers and temporal commitment registers associated with each SM. The DAs employ SMPC to reconstruct the polynomial for spatial aggregation. The reconstructed polynomial, along with the spatial commitment register, is reported to the ESP for the corresponding time instance ($t = 1$). Subsequently, the spatial registers and spatial commitment registers are reset to zero and one, respectively. In the Integrity Verification Phase (Spatial Aggregation), the ESP verifies the integrity

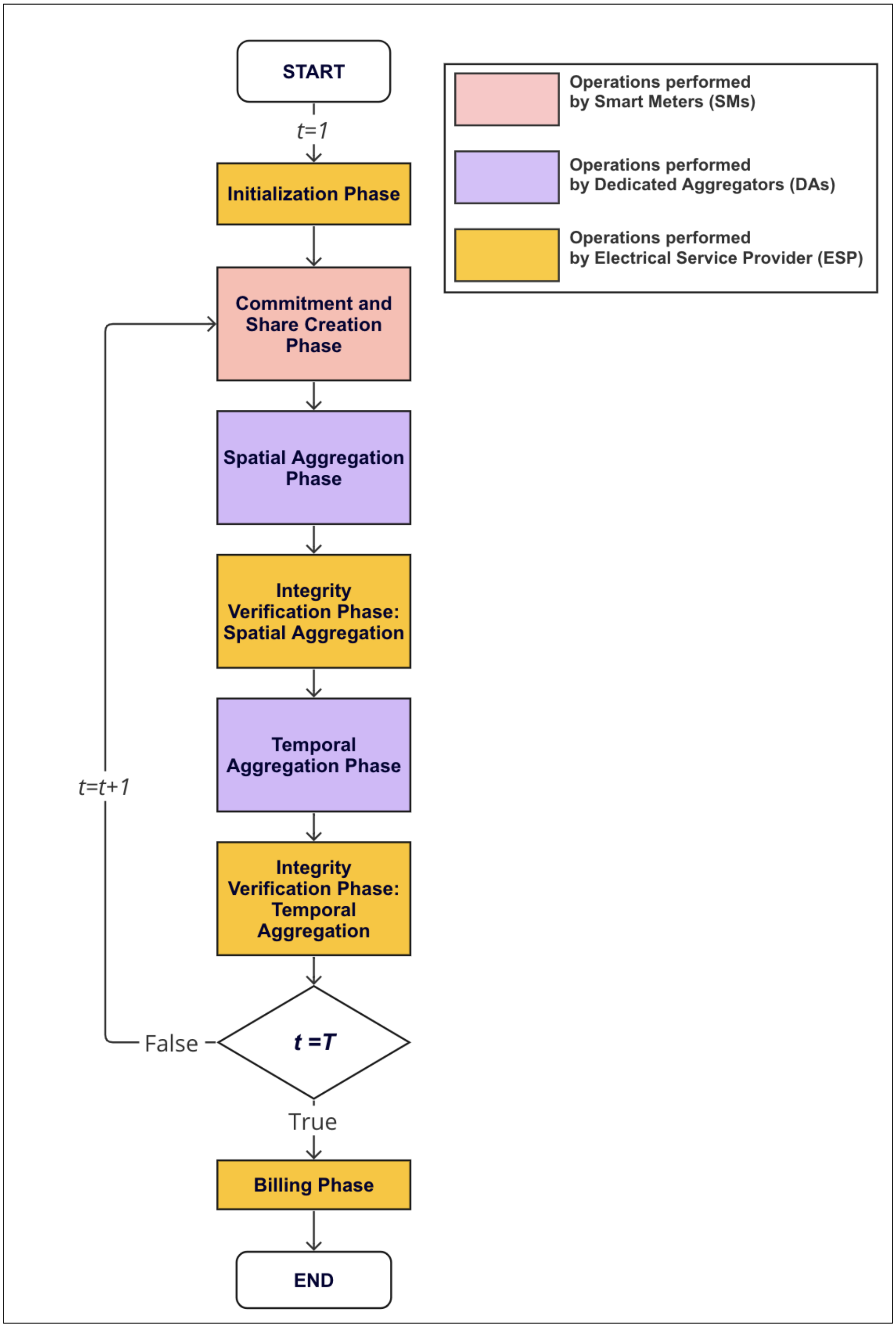

START
t=1
Initialization Phase
Commitment and Share Creation Phase
Spatial Aggregation Phase
Integrity Verification Phase: Spatial Aggregation
Temporal Aggregation Phase
Integrity Verification Phase: Temporal Aggregation
t =T
False
t=t+1
True
Billing Phase
END
Operations performed by Smart Meters (SMs)
Operations performed by Dedicated Aggregators (DAs)
Operations performed by Electrical Service Provider (ESP)

Figure 5.3: Flowchart of our Privacy Safeguarding Framework in a Malicious Setting

of the spatially aggregated metering data by computing commitments using the parameters derived from the reconstructed polynomial and comparing them with the received commitments from the DAs. A match indicates that no modifications have been made to the spatial metering data, while a mismatch prompts the ESP to estimate values based on historical data. This process continues until the current time instance is not equal to the total time period ($t = T$). When $t = T$, the temporal aggregation takes place among the same set of DAs using SMPC. The DAs reconstruct the polynomial for temporal aggregation and transmit it to the ESP, along with the temporal commitment registers, for the Integrity Verification Phase: Temporal Aggregation. Once reported to the ESP, the temporal registers and temporal commitment registers are reset to zero and one, respectively. The ESP computes commitments and compares them with the received commitments from the DAs. A match indicates the absence of modifications to the temporal metering data, while a mismatch triggers the ESP to estimate values based on historical data for billing purposes. In the final phase, the Billing Phase, the ESP reports the computed bill to the customer.

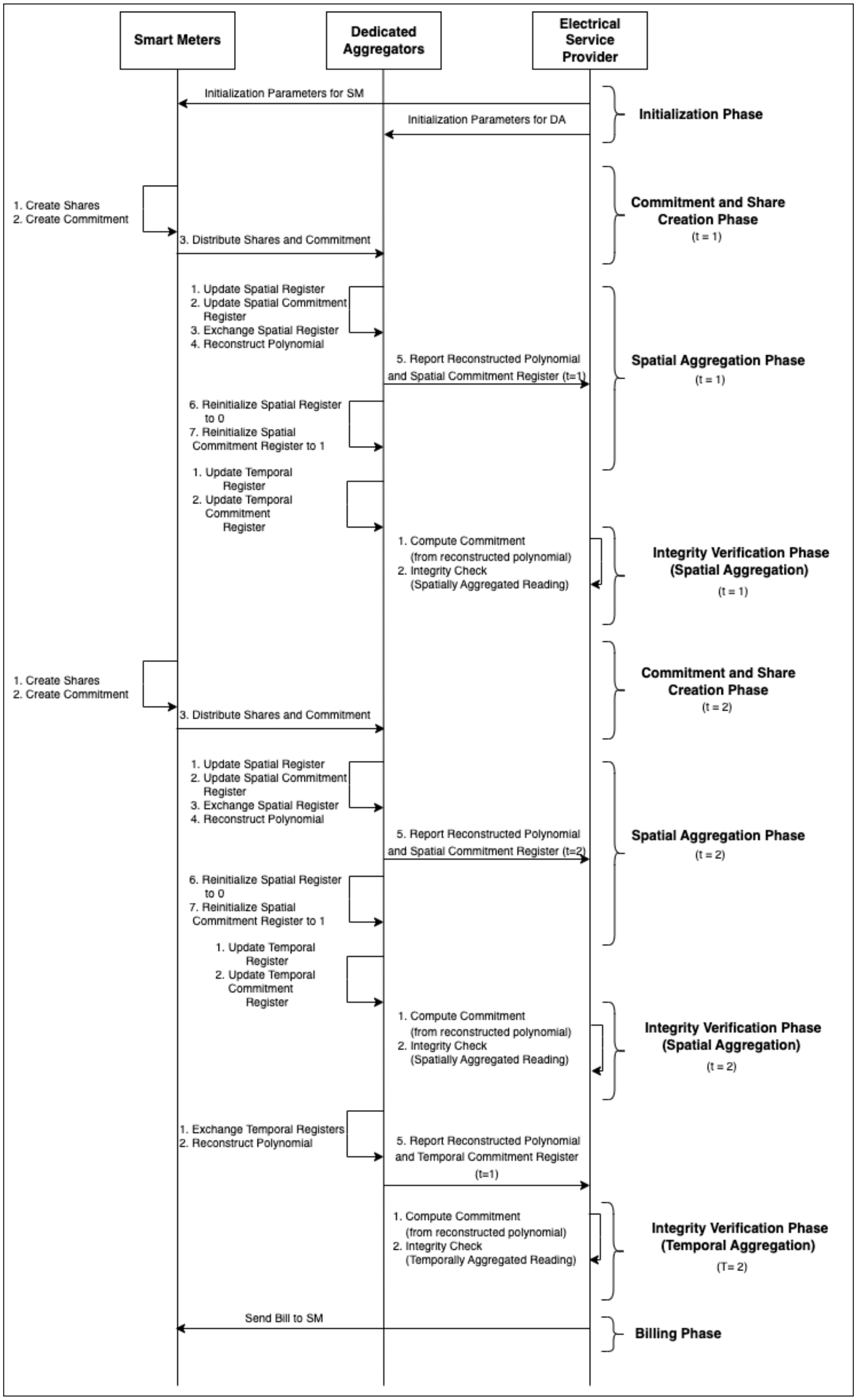

Smart Meters
Dedicated Aggregators
Electrical Service Provider
Initialization Parameters for SM
Initialization Parameters for DA
Initialization Phase
1. Create Shares
2. Create Commitment
3. Distribute Shares and Commitment
Commitment and Share Creation Phase
(t = 1)
1. Update Spatial Register
2. Update Spatial Commitment Register
3. Exchange Spatial Register
4. Reconstruct Polynomial
5. Report Reconstructed Polynomial and Spatial Commitment Register (t=1)
Spatial Aggregation Phase
(t = 1)
6. Reinitialize Spatial Register to 0
7. Reinitialize Spatial Commitment Register to 1
1. Update Temporal Register
2. Update Temporal Commitment Register
1. Compute Commitment (from reconstructed polynomial)
2. Integrity Check (Spatially Aggregated Reading)
Integrity Verification Phase (Spatial Aggregation)
(t = 1)
1. Create Shares
2. Create Commitment
3. Distribute Shares and Commitment
Commitment and Share Creation Phase
(t = 2)
1. Update Spatial Register
2. Update Spatial Commitment Register
3. Exchange Spatial Register
4. Reconstruct Polynomial
5. Report Reconstructed Polynomial and Spatial Commitment Register (t=2)
Spatial Aggregation Phase
(t = 2)
6. Reinitialize Spatial Register to 0
7. Reinitialize Spatial Commitment Register to 1
1. Update Temporal Register
2. Update Temporal Commitment Register
1. Compute Commitment (from reconstructed polynomial)
2. Integrity Check (Spatially Aggregated Reading)
Integrity Verification Phase (Spatial Aggregation)
(t = 2)
1. Exchange Temporal Registers
2. Reconstruct Polynomial
5. Report Reconstructed Polynomial and Temporal Commitment Register (t=1)
1. Compute Commitment (from reconstructed polynomial)
2. Integrity Check (Temporally Aggregated Reading)
Integrity Verification Phase (Temporal Aggregation)
(T= 2)
Send Bill to SM
Billing Phase

Figure 5.4: A Detailed Workflow of our Privacy Safeguarding Framework in Malicious Setting

5.6 Summary

In this chapter, we introduced a distributed privacy-preserving framework in a malicious setting and tackled two research questions. The framework we propose has the ability to securely report spatio-temporal data to the ESP through the use of DAs, ensuring privacy. Furthermore, it can maintain the integrity of spatio-temporal metering data while preserving customer privacy. In the upcoming chapter, we will assess the feasibility of developing our proposed framework and evaluate its performance by comparing it to other relevant literature works, based on identified parameters.

Chapter 6

Results and Analysis

6.1 Introduction

In this chapter, we will be addressing the following three research questions (RQs):

- **RQ5: How to develop a practical proof of concept, that assesses computational overhead and end-to-end delay of our proposed framework in comparison to other state-of-the-art frameworks?**
- **RQ6: How do we evaluate the resilience of our proposed framework against potential security attacks?**
- **RQ7: How do we evaluate the resilience of our proposed framework against potential privacy attacks?**

To address RQ5, we have developed a practical proof of concept that utilizes both an embedded and a cloud-based environment. Additionally, we leverage this proof of concept to evaluate the performance of the framework across various configurations and conduct a comparative analysis with respect to relevant frameworks in the literature [11,14]. RQ6 considers the proposed framework in a malicious setting, where the adversary aims to disrupt the grid and billing functionalities by modifying the metering data, commitments, and/or billing tariffs. As the malicious adversary possesses the capability to execute not only active attacks but also passive attacks, in RQ7 we consider the proposed framework in a malicious setting, where the adversary aims to breach the privacy of the customer(s) by linking the granular reading with the associated identity of the smart meter (SM). In this chapter, we describe the experimental setup for the proof of concept that we have developed to evaluate the performance of our framework and compare it with other relevant works in the literature that share similar security and privacy goals. Additionally, we assess the framework's resilience against potential security and privacy attacks. Finally, we provide a qualitative comparison of Aggregation-based Frameworks based on the identified metrics.

In our proposed framework, the responsibility of creating shares and commitments for the corresponding meter readings lies with the SMs. These shares and commitments are then shared with the Dedicated Aggregators (DAs). The DAs perform spatio-temporal aggregation and forward the aggregated data to the Electrical Service Provider (ESP) for integrity verification. Among the comparative frameworks, the framework proposed by Borges *et* al. is an In-Network Aggregation-based Framework that provides verifiable spatio-temporal aggregation [14]. It utilizes a Paillier Encryption-based Scheme [77] along with a Commitment-based Scheme [79] to achieve the desired functionality. Notably at the time of the publication of this dissertation, this framework is the only one that supports verifiable dynamic billing functionality. For our experiments, we have designed the In-Network Aggregation-based Framework up to level 1, where the leaf nodes (SMs) report to the parent SM that is responsible for forwarding the aggregated result to the ESP. If additional levels are constructed (increased depth) for the In-Network Aggregation-based Framework, the end-to-end delay will be further affected. In contrast, the second comparative framework is a Centralized Aggregation-based Framework that supports spatio-temporal aggregation [11]. It employs Paillier Encryption-based Scheme [77] and the Elliptic Curve Diffie Hellman Ephemeral (ECDHE) Scheme [4]. The ECDHE Scheme relies on a trust key to further secure the encrypted meter readings. The associated entities (SM, DA, and ESP) generate the trust key for each given time instance.

6.2 Experimental Setup

In this section, we will describe the experimental setup we used to develop the proof of concept for our proposed framework [107], as well as the competing frameworks [11, 14]. SMs typically have several crucial tasks and often have limited computing power available for deployment of privacy-preserving technologies. Therefore, a scheme that imposes minimal overhead to the SMs is desired. To assess the computational overhead of the aforementioned frameworks, we utilize the Raspberry Pi 3 Model B as a representative SM. Raspberry Pis are commonly used in the literature to simulate SMs [70,100]. However, it is important to note that the actual overhead may differ when implementing these schemes on real-world SMs due to other factors. Nonetheless, our objective is to conduct a comparative analysis of our framework with other frameworks. Assuming other factors are equal, we anticipate a similar performance trend in actual deployments. Additionally, we have configured the Raspberry Pi with real-world traces of meter readings obtained from the UMass dataset [75].

The DAs are deployed in a cloud environment. However, the DAs can be deployed in various environments, including on-premise, cloud, or hybrid configurations (both on-premise and cloud) [7, 12, 61, 69, 78]. For the deployment of our proof of concept, we have opted for Amazon Web Services (AWS) as our cloud service provider to host our DAs, enabling them to perform spatio-temporal aggregation. The ESP, responsible for initializing SMs and DAs, is also hosted on AWS as part of our cloud deployment. The ESP plays a crucial role in initializing the SMs and DAs. Setting the degree of the polynomial to $(k - 1 = n - 1)$ in our proposed framework enhances its resilience against cyberattacks. This configuration ensures that the adversary would need to acquire all n shares to successfully reconstruct the high-frequency meter reading, significantly increasing the difficulty for them to compromise the framework's security. In addition to its initialization role, the ESP is also responsible for providing grid and billing functionalities to the customers associated with the SMs. To implement our proof of concept, we have utilized a 512-bit Paillier key size for both the In-Network Aggregation-based Framework and the Centralized Aggregation-based Framework. In our proposed framework, we specify the size of the prime number p as 512 bits. Furthermore, we employ the prime number (p) to generate commitments associated with the In-Network Aggregation-based Framework and our proposed framework.

6.3 Evaluation Metrics

6.3.1 Computational Overhead

The computational overhead is a critical metric for resource-constrained SMs. SMs are responsible for reporting high-frequency metering data at frequent intervals of time (e.g., every 15 minutes), making computational efficiency a crucial requirement. A high computational overhead on an SM can result in slower overall performance of the smart grid framework. Therefore, minimizing the computational overhead is essential to ensure faster transmission of metering data to the ESP for efficient grid functionalities.

6.3.2 End-to-End Delay

The end-to-end delay is another critical metric that measures the time required for metering data to travel from the source (SMs) to the destination (ESP). In the smart grid, it is essential to transmit the metering data, especially high-frequency metering data, to the ESP in real-time to facilitate prompt decision-making and to balance the supply-demand curve. Any delay in transmitting the metering data may lead to inaccurate supply-demand forecasting, resulting in potential outages. By minimizing the end-to-end delay, the smart grid can swiftly respond to fluctuating energy demands, identify modifications, and take appropriate actions. This capability is a crucial feature sought after by ESP.

6.3.3 Resilience to Security Attacks

This aspect covers the comprehensive resilience of the framework against modification attacks on high-frequency data. A strong framework should have the capability to promptly and effectively detect and respond to cyberattacks, reducing the impact of the attack on the system and safeguarding its core functionality and performance. This capability is a crucial feature desired by ESP.

6.3.4 Resilience to Privacy Attacks

This aspect encompasses the privacy considerations pertaining to the customers associated with the smart grid. Privacy in the smart grid involves safeguarding energy consumption data (especially high-frequency metering data), billing information, and personal identification data. Unauthorized access to high-frequency metering data can enable attackers to infer individuals' presence or absence, facilitate targeted marketing efforts, and potentially lead to malicious activities based on metering data or load profiles. Therefore, preserving the privacy of metering data is of utmost importance from both the customer's and ESP's perspectives.

6.4 Performance Analysis

Our goal in this section is to evaluate our proposed framework with different sets of configurations. We also aim to examine and contrast our proposed framework with other relevant frameworks identified in Section 6.1. The analysis is conducted using the proof of concept we developed, as described in Section 6.2, and takes into account the metrics identified in Section 6.3. By conducting this thorough analysis, we aim to offer valuable insights into the effectiveness of our proposed framework compared to existing frameworks in the literature [11, 14].

6.4.1 Computational Overhead on Smart Meters due to Spatial Aggregation in a Semi-Honest Setting

Figure 6.1 represents the computational overhead associated with a given SM due to spatial aggregation in a semi-honest setting. The overhead on the SM is due to the Shamir's Secret Sharing Scheme, that utilizes the initialization parameters from the ESP (such as degree of secret sharing polynomial, list of DAs, and the prime number) along with its instantaneous reading (from the UMass dataset) to create the corresponding shares for the DAs. For our work, we set the degree of the polynomial to $(n-1)$ as more number of shares imply higher privacy and resilience against cyberattacks. This setting ensures that all the n shares are required to successfully recover the instantaneous reading associated with a given SM. As seen in Figure 6.1, the computational overhead increases with an increase in the number of DAs.

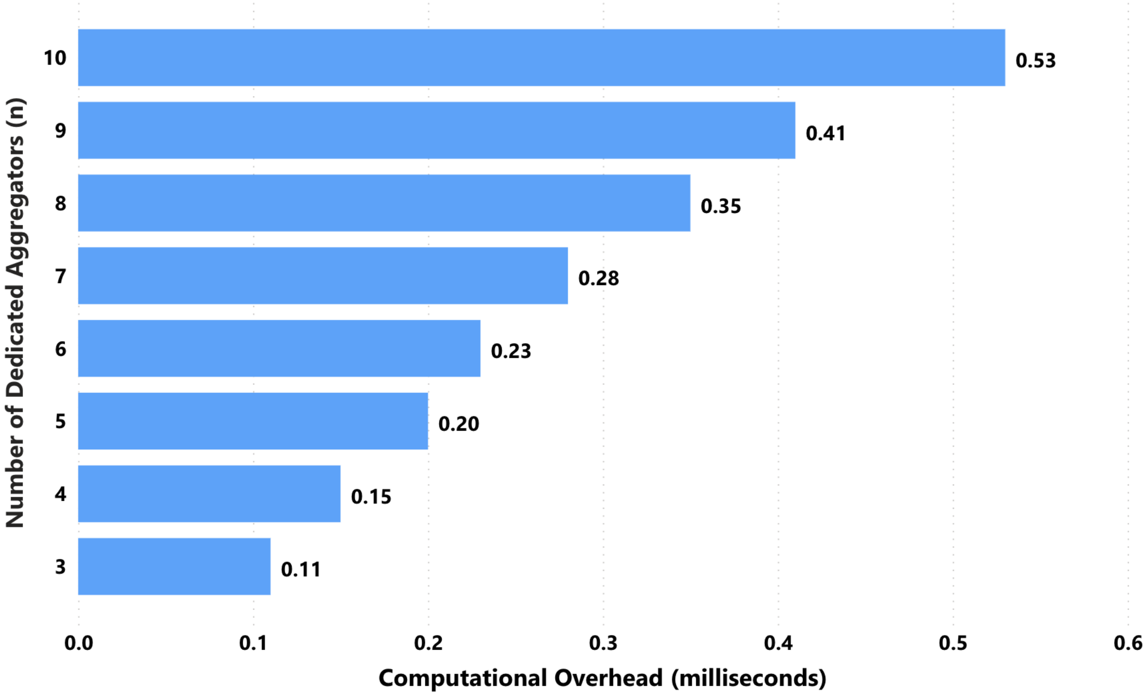


Figure 6.1: Computational Overhead on Smart Meters due to Spatial Aggregation in a Semi-Honest Setting

In Figure 6.2, we further conduct a comprehensive examination of the increasing computational overhead associated with Shamir's Secret Sharing Scheme as the number of DAs increases in a semi-honest setting (Fig. 6.1). Our investigation focuses on analyzing two distinct phases within the scheme: the Polynomial Selection Phase and the Share Creation Phase, as depicted in Figure 6.2. Through our analysis, we reveal that the Polynomial Selection Phase has a relatively minimal impact on the overall computational overhead of the Shamir's Secret Sharing Scheme, compared to the Share Creation Phase. The process of Share Creation Phase requires the SM to tackle the computational challenge of constructing a brand new share for the new DA by solving a polynomial equation (with modulus operation) using the updated DA_{list}. This places a significant computational overhead on the SM.

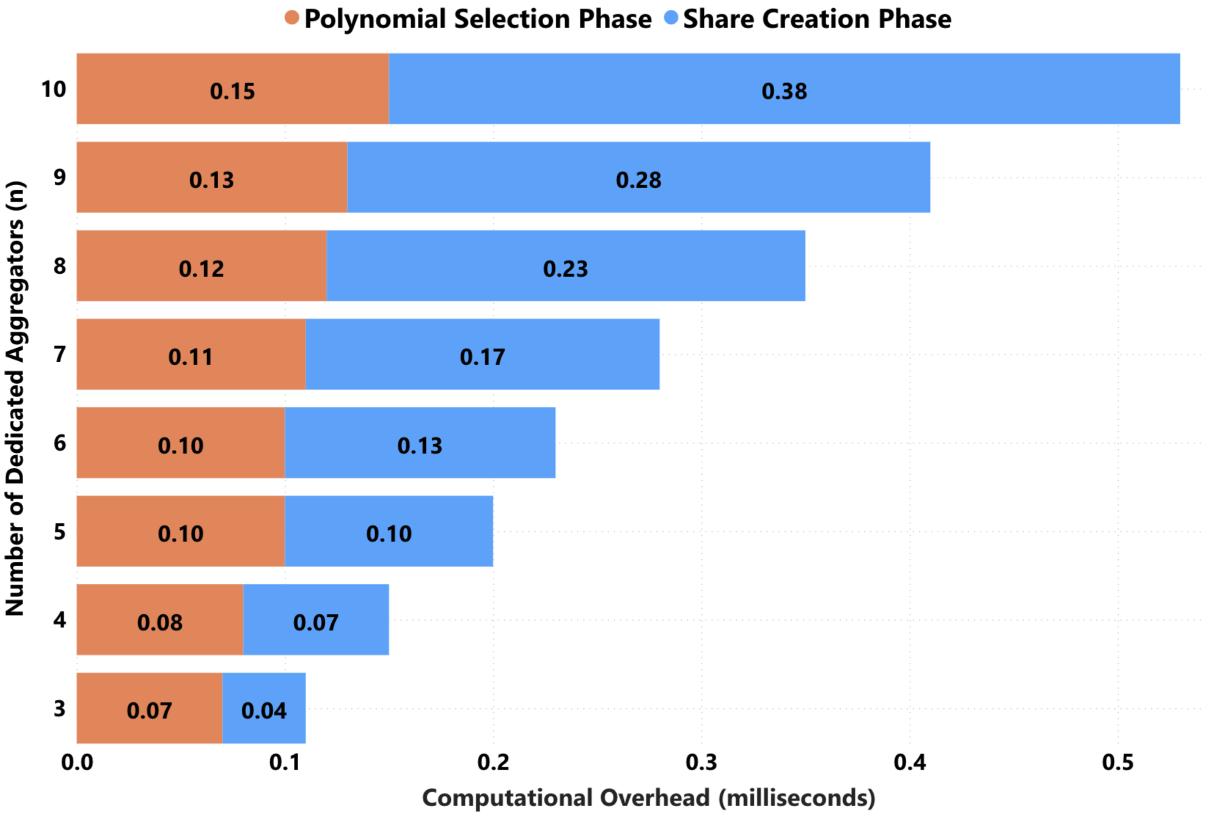


Figure 6.2: Comprehensive Analysis of Shamir's Secret Sharing Scheme for Smart Meters in a Semi-Honest Setting

6.4.2 Computational Overhead on Dedicated Aggregators due to Spatial Aggregation in a Semi-Honest Setting

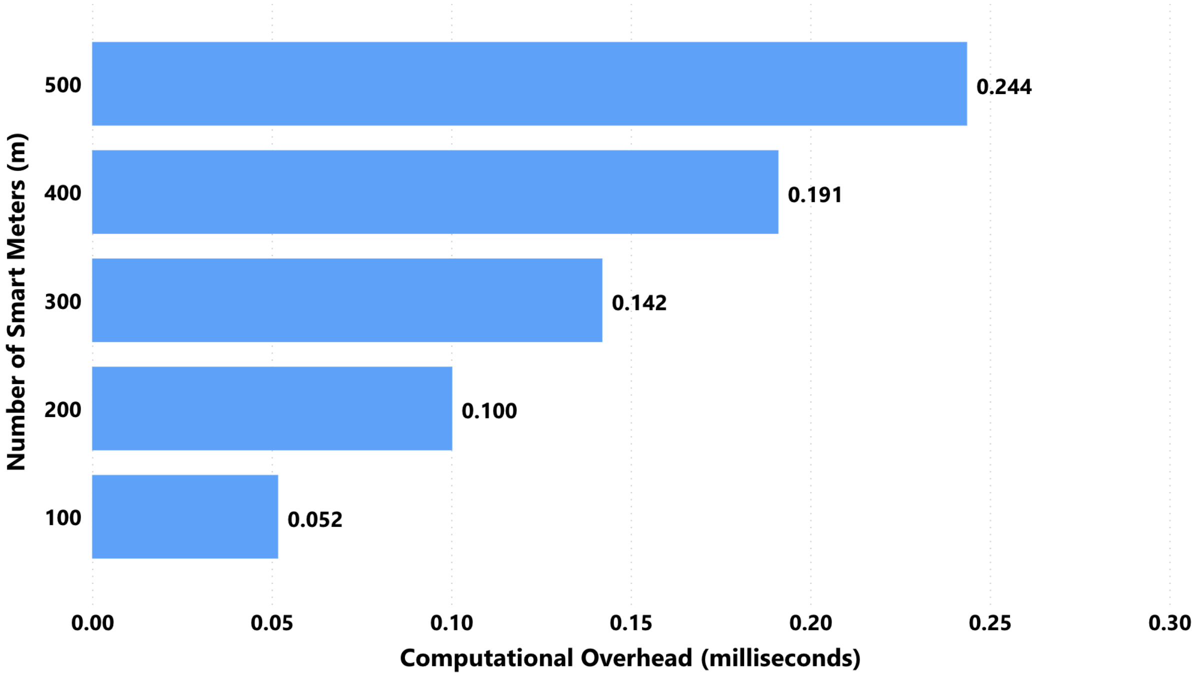


Figure 6.3: Computational Overhead of updating Spatial Registers (Spatial Aggregation) associated with Dedicated Aggregators in a Semi-Honest Setting

In this subsection, we analyze the computational overhead associated with the DAs for reporting spatially aggregated reading to the ESP in a semi-honest setting. The shares created by each SM employing Shamir's Secret Sharing Scheme are being reported to the corresponding DAs participating in the spatial aggregation. Each DA has a single spatial register that is responsible for aggregating all the shares received from the given set of SMs in the SM_{List}. As depicted in Figure 6.3, the computational overhead linked to updating the spatial register grows proportionally with the increasing number of SMs. This update process takes place at regular intervals of time. Once the spatial registers for the designated DAs are updated, the DAs engage in secure multiparty computation (SMPC) by exchanging their spatial registers and subsequently reconstructing the polynomial associated with the spatial aggregation. As the degree of secret sharing polynomials selected by the given set of SMs is same ($k - 1 = n - 1$), the aggregated shares in the spatial registers represents the shares of the polynomial reconstructed in the Spatial Aggregation Phase. This is possible due to the homomorphic properties of the Shamir's Secret Sharing Scheme. As seen in Figure 6.4, the computational overhead increases with increase in number of DAs, since more number of spatial registers entries are needed to be exchanged to reconstruct an $(n-1)$ degree polynomial related to the spatial aggregation.

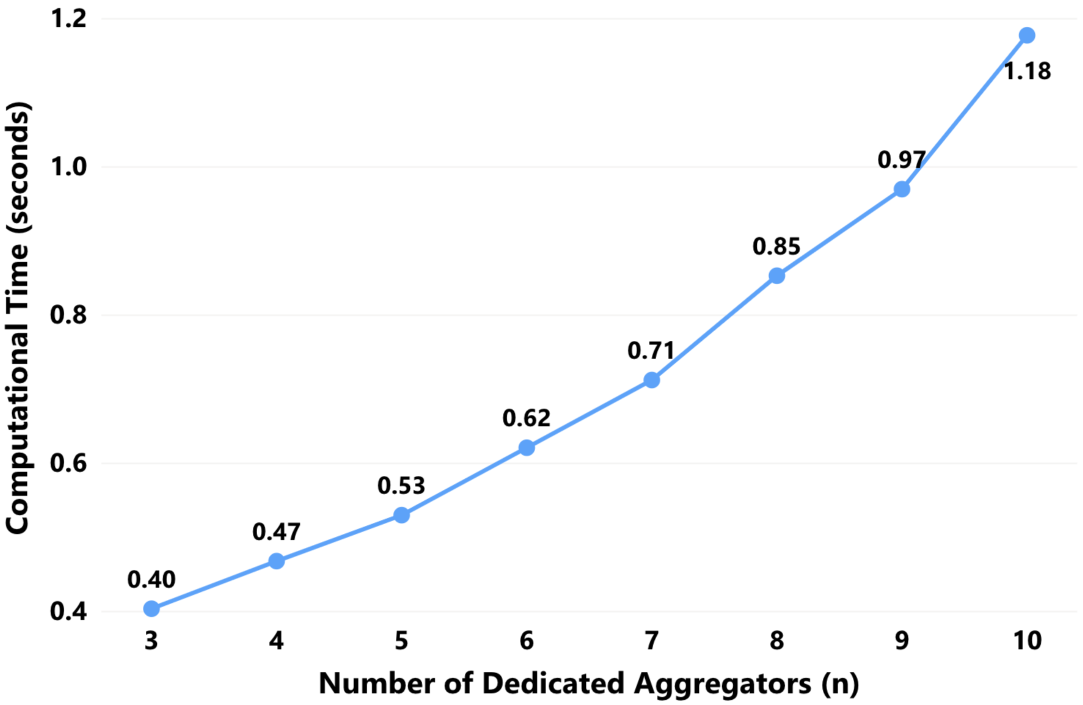


Figure 6.4: Computational Overhead of reconstructing polynomial related to Spatial Aggregation in a Semi-Honest Setting

6.4.3 Overall Computational Overhead associated while reporting Spatially Aggregated Reading to the Electrical Service Provider in a Semi-Honest Setting

In this subsection, we analyze the overall computational overhead associated with all the entities (SMs, DAs, ESP) for reporting spatially aggregated reading to the ESP in a privacy-preserving fashion. We consider a given scenario with SMs $(m = 500)$, DAs $(n = \{3, 4, \ldots, 10\})$ and, the ESP (Fig. 6.5). The SMs are responsible for creating given set of shares based on the number of DAs participating in the spatial aggregation. As we can see in Figure 6.5, the computational overhead on the SMs due to Shamir's Secret Sharing Scheme increases with increase in the number of SMs. The shares that are distributed by the SMs are aggregated and stored in the spatial register corresponding to each DA. The computational overhead associated with the updation of the spatial register is constant (Fig. 6.5) since the given number of SMs are fixed ($m = 500$). Post updation of the spatial register, the DAs employ SMPC to reconstruct the polynomial related to spatial aggregation. The computational overhead on the DAs for reconstructing the polynomial increases with increase in the number of DAs as more number of spatial registers are shared with each other along with constructing a higher degree polynomial $(k-1) = (n-1)$ with increase in the number of DAs. The reconstructed polynomial is same across the DAs in the DA_{List} since they employ same basis polynomials and same set of spatial registers. The constant of the reconstructed polynomial represents the spatially aggregated reading across given set of SMs, this is possible due to homomorphic properties of Shamir's Secret Sharing Scheme. Post reporting the spatially aggregated reading to the ESP, the DAs reset their spatial registers to zero (so that it can be ready for use in the next time instance). There is no computational overhead on ESP as it receives spatially aggregated reading from the DAs for every given instance of time. As spatially aggregated reading is reported to the ESP, it cannot derive the instantaneous reading and link it to a given SM from the SM_{List}, thereby preserving customer privacy.

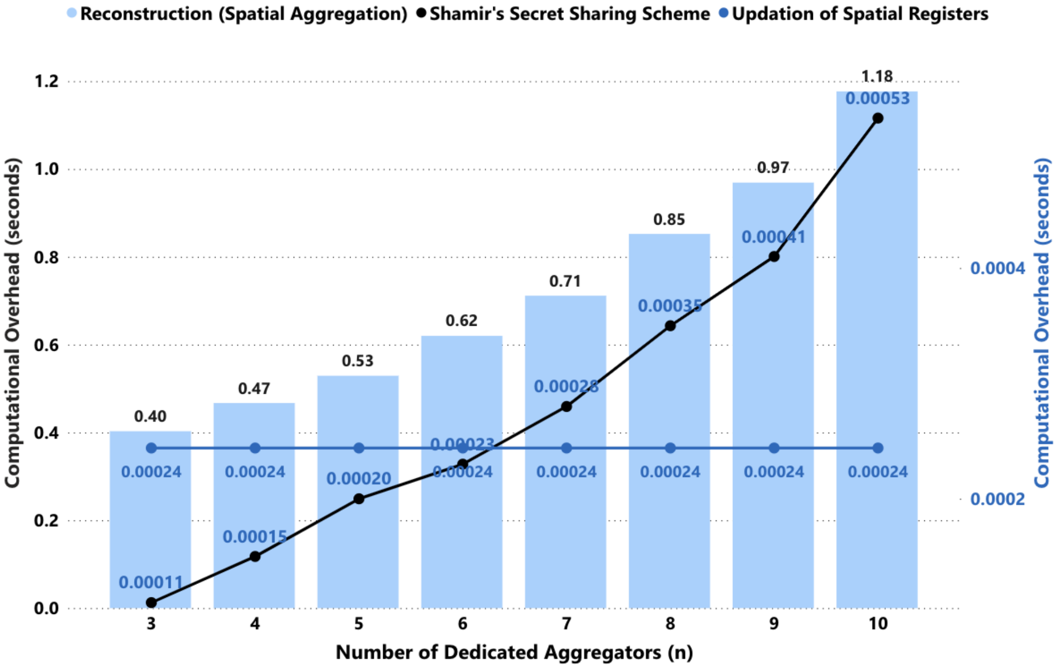


Figure 6.5: Overall Computational Overhead associated while reporting Spatially Aggregated Reading to the Electrical Service Provider in a Semi-Honest Setting

6.4.4 Computational Overhead on Smart Meters due to Temporal Aggregation in a Semi-Honest Setting

In the previous subsection, we were able to report spatially aggregated reading from SMs to the ESP via the DAs in a semi-honest-setting. However, in subsection we adapt our proposed framework order to support temporal aggregation that is required to provide billing functionalities. Importantly, we accomplish this without imposing any additional computational overhead on the SMs. We utilize the same set of shares generated through the application of Shamir's Secret Sharing Scheme, as illustrated in Figure 6.1.

6.4.5 Computational Overhead on Dedicated Aggregators due to Temporal Aggregation in a Semi-Honest Setting

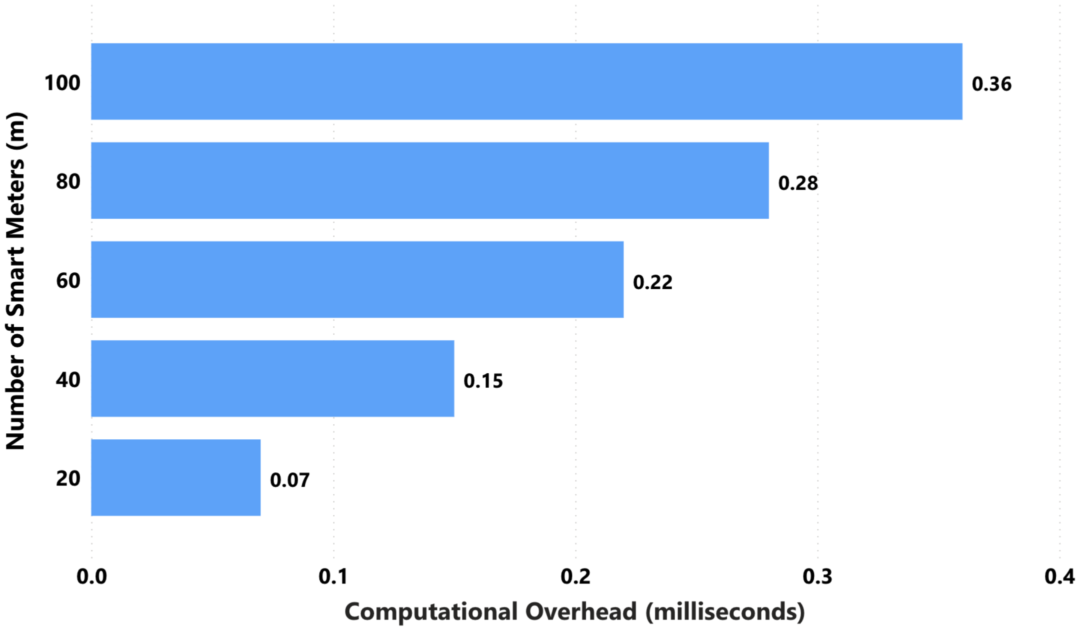


Figure 6.6: Computational Overhead of updating Temporal Registers (Temporal Aggregation) associated with Dedicated Aggregators in a Semi-Honest Setting

In this subsection, we analyze the computational overhead associated with the DAs for reporting temporally aggregated reading to the ESP in a semi-honest setting. The shares created by each SM employing Shamir's Secret Sharing Scheme are being reported to the corresponding DAs participating in the temporal aggregation for Time of Use Billing. Each DA maintains a dedicated temporal register per SM that is responsible for aggregating the shares received from the given set of SMs in the SM_{List} over a period of time. As depicted in Figure 6.6, the computational overhead linked to update the temporal register grows proportionally with the increasing number of SMs. This update process takes place at regular intervals of time (at every given instant). Once the temporal registers for the designated DAs are updated over a given period of time (T), the DAs engage in SMPC at $t = T$ by exchanging their temporal registers and subsequently reconstructing the polynomials associated with the given set of SMs (one polynomial per SM) in the SM_{List}. As the degree of secret sharing polynomials selected by the given SM over a period of time is same ($k - 1 = n - 1$), the aggregated shares in the corresponding temporal registers across the DAs represents the shares of the polynomial reconstructed in the Temporal Aggregation Phase. This is possible due to the homomorphic properties of the Shamir's Secret Sharing Scheme. As seen in Figure 6.7, the computational overhead increases with increase in number of SMs, since more number of temporal registers are needed to be exchanged to reconstruct an $(n - 1)$ degree polynomial related to each of the m SMs in the SM_{List}.

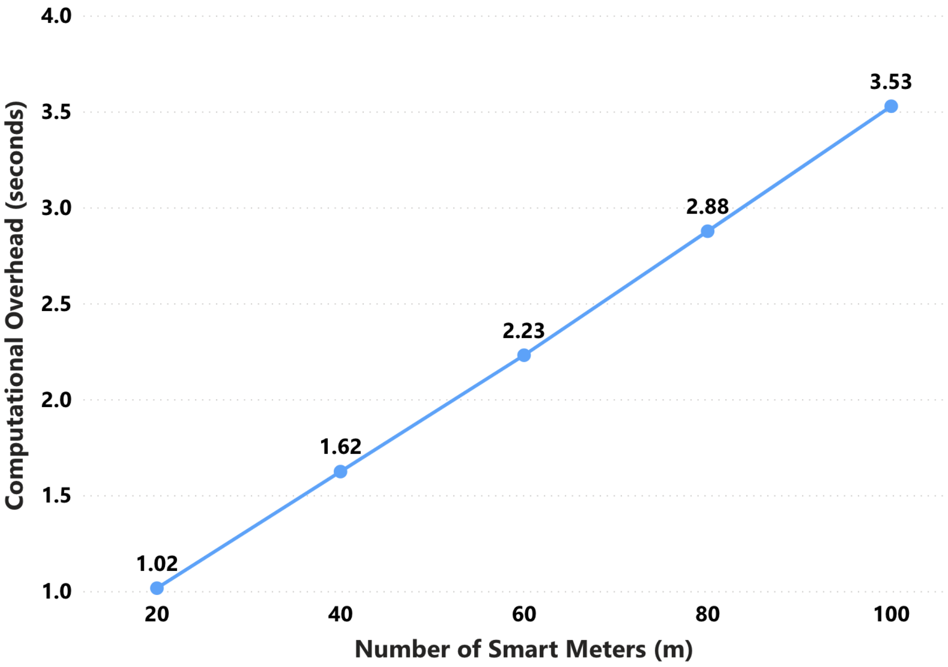


Figure 6.7: Computational Overhead of reconstructing polynomial related to Temporal Aggregation in a Semi-Honest Setting

6.4.6 Overall Computational Overhead associated while reporting Temporally Aggregated Reading to the Electrical Service Provider in a Semi-Honest Setting

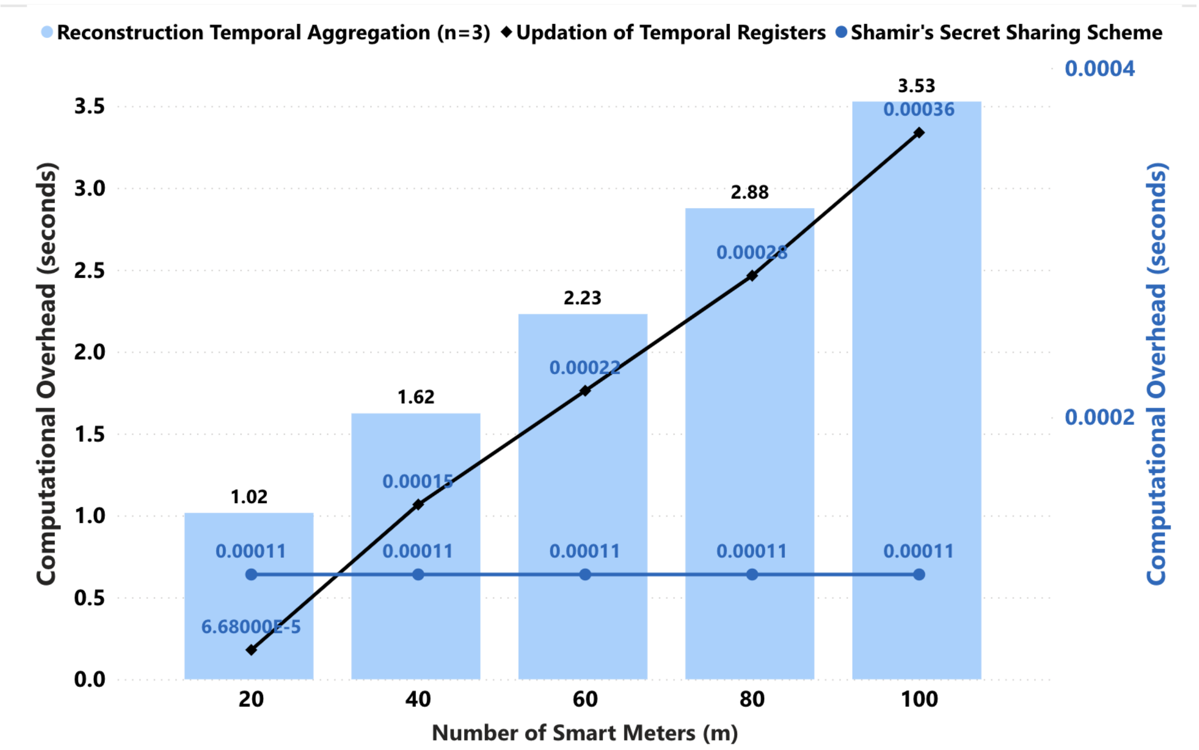


Figure 6.8: Overall Computational Overhead associated while reporting Temporally Aggregated Reading to the Electrical Service Provider in a Semi-Honest Setting

In this subsection, we analyze the overall computational overhead associated with all the entities (SMs, DAs, ESP) for reporting temporally aggregated reading to the ESP in a privacy-preserving fashion. We consider a given scenario with SMs ($m = \{20, 40, \ldots, 100\}$), DAs ($n = 3$) and, the ESP (Fig. 6.8). The SMs are responsible for creating given set of shares based on the number of DAs participating in the temporal aggregation. As we can see in Figure 6.8, the computational overhead on the SMs due to Shamir's Secret Sharing Scheme is constant since each SM has to create only three shares corresponding to each DA for a given instance of time. The shares that are distributed by the SMs are aggregated and stored in the corresponding temporal registers. Each DA has m temporal registers, one per SM. The computational overhead associated with the updation of the temporal registers is directly proportional to the number of SMs (Fig. 6.8). Post updation of the temporal registers, the DAs employ SMPC to reconstruct the polynomials related to temporal aggregation. The computational overhead on the DAs for reconstructing the polynomials related to temporal aggregation increases with increase in the number of SMs as more number of temporal registers are shared with each other along with increased number of reconstructed polynomials related to temporal aggregation. The reconstructed polynomials with respect to each SM is same across the DAs in the DA_{List} since they employ same basis polynomials and same set of temporal registers corresponding to each SM. The constants of the reconstructed polynomials represents the corresponding bill computed by applying Time of Use Billing Tariff across given set of SMs, this is possible due to homomorphic properties of Shamir's Secret Sharing Scheme. Post reporting the computed bill to the ESP, the DAs reset their temporal registers to zero (so that it can be ready for use in the next time period). There is no computational overhead on ESP as it receives computed bill from the DAs for every given instance of time (usually end of each month). As computed bill is reported to the ESP, it cannot derive the instantaneous reading and link it to a given SM from the SM_{List}, thereby preserving customer privacy.

6.4.7 Computational Overhead on Smart Meters due to Spatial Aggregation in a Malicious Setting

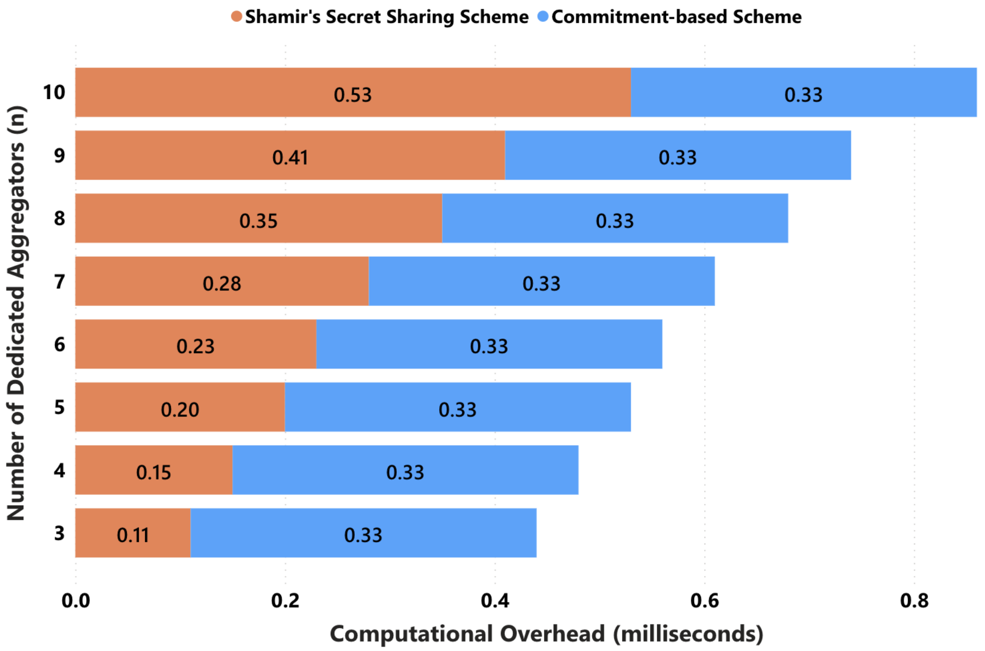


Figure 6.9: Computational Overhead on a Smart Meter in a Malicious Setting

Figure 6.9 represents the computational overhead associated with a given SM due to spatial aggregation in a malicious setting. The overhead on the SM is due to the Shamir's Secret Sharing Scheme and Commitment Scheme, that utilizes the initialization parameters from the ESP (such as degree of secret sharing polynomial, list of DAs, commitment creation parameters, seed list, and the prime number) along with its instantaneous reading from the UMass dataset to create the corresponding shares and overall commitment for the DAs. For our work, we set the degree of the polynomial to $(n-1)$ as more number of shares imply higher privacy and resilience against cyberattacks. This setting ensures that all the n shares are required to successfully recover the instantaneous reading associated with a given SM. Observing Figure 6.9, we notice that the computational overhead linked to Shamir's Secret Sharing Scheme escalates with the growing number of DAs. In contrast, the Commitment-based Scheme incurs a constant overhead as the commitment is created based on the overall instantaneous reading and not the corresponding shares.

6.4.8 Computational Overhead on Dedicated Aggregators due to Spatial Aggregation in a Malicious Setting

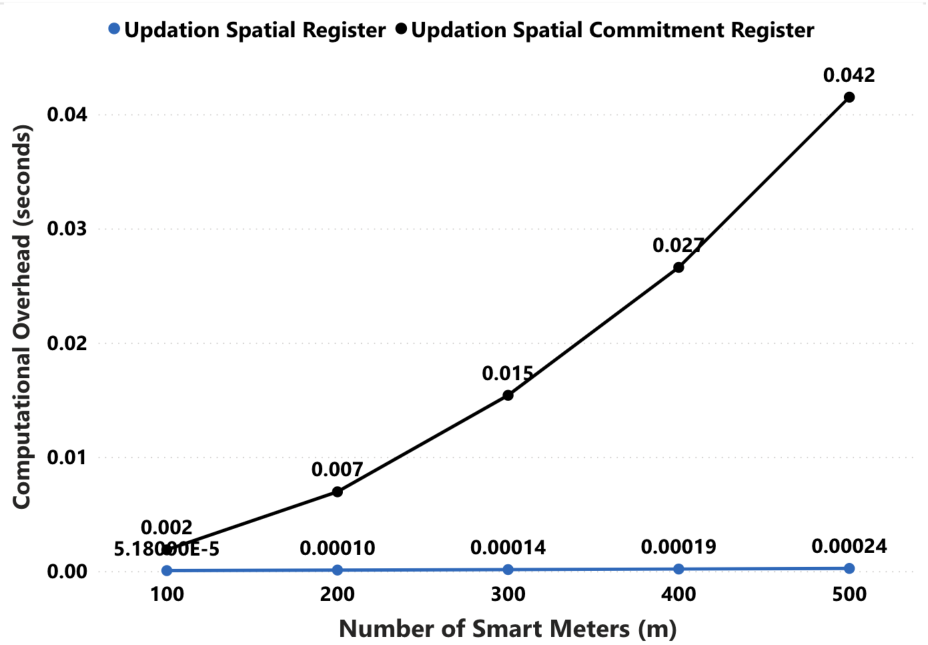


Figure 6.10: Computational Overhead of updating Spatial Registers and Spatial Commitment Registers (Spatial Aggregation) associated with Dedicated Aggregators in a Malicious Setting

In this subsection, we analyze the computational overhead associated with the DAs for reporting spatially aggregated reading to the ESP in a malicious setting. The shares and the commitments created by each SM employing Shamir's Secret Sharing Scheme and Commitment-based Scheme are being reported to the corresponding DAs participating in the spatial aggregation. Each DA has a single spatial register and spatial commitment register that is responsible for aggregating all the shares and commitments received from the given set of SMs in the SM_{List}. As shown in Figure 6.10, the computational overhead associated with updating the spatial register and spatial commitment register increases proportionally with the increasing number of SMs. However, the spatial commitment register incurs a higher computational overhead than the spatial registers because the commitments are multiplied with each other, whereas for spatial registers, the shares are simply added together. This update process takes place at regular intervals of time (at every given instant). Once the spatial registers and spatial commitment registers for the designated DAs are updated, the DAs engage in SMPC by exchanging their spatial registers (Note: spatial commitment registers are not exchanged with other DAs) and subsequently reconstructing the polynomial associated with the spatial aggregation. As the degree of secret sharing polynomials selected by the given set of SMs is same ($k - 1 = n - 1$), the aggregated shares in the spatial registers represents the shares of the polynomial reconstructed in the Spatial Aggregation Phase. This is possible due to the homomorphic properties of the Shamir's Secret Sharing Scheme. As previously seen in Figure 6.4, the computational overhead increases with increase in number of DAs, since more number of spatial registers are needed to be exchanged to reconstruct a $(n - 1)$ degree polynomial related to the spatial aggregation.

6.4.9 Overall Computational Overhead associated while reporting Spatially Aggregated Reading to the Electrical Service Provider in a Malicious Setting

In this subsection, we analyze the overall computational overhead associated with all the entities (SMs, DAs, ESP) for reporting spatially aggregated reading to the ESP in a malicious setting. We consider a given scenario with SMs ($m = 500$), DAs ($n = \{3, 4, \ldots, 10\}$) and, the ESP (Fig. 6.11). The SMs are responsible for creating a given set of shares and commitments based on the number of DAs participating in the spatial aggregation. As we can see in Figure 6.11, the overall computational overhead on the SMs due to Shamir's Secret Sharing Scheme and Commitment-based Scheme increases with an increase in the number of SMs. The shares and commitments distributed by the SMs are aggregated and stored in the spatial register and spatial commitment registers, respectively. The computational overhead associated with updating the spatial register and spatial commitment register remains constant since the given number of SMs is fixed ($m = 500$). After updating the spatial register, the DAs employ SMPC to reconstruct the polynomial related to spatial aggregation (Note: spatial commitment registers are not exchanged with other DAs). The overall computational overhead on the DAs for reconstructing the polynomial increases with an increase in the number of DAs, as more spatial registers are shared with each other and a higher degree polynomial ($k-1=n-1$) is constructed with more DAs (Fig. 6.11).

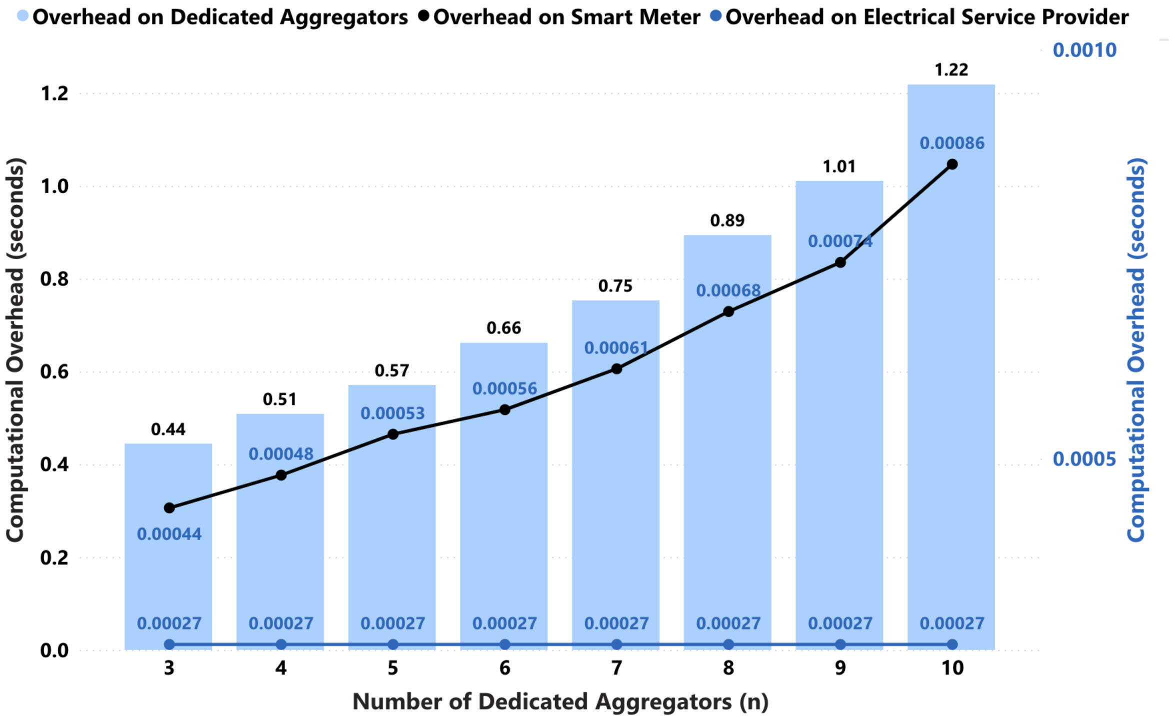


Figure 6.11: Overall Computational Overhead associated while reporting Spatially Aggregated Reading to the Electrical Service Provider in a Malicious Setting

The reconstructed polynomial is the same across the DAs in the DA_{List} since they use the same basis polynomials and the same set of spatial registers. The constant of the reconstructed polynomial represents the spatially aggregated reading across the given set of SMs, which is possible due to the homomorphic properties of Shamir's Secret Sharing Scheme. After reporting the reconstructed polynomial regarding spatial aggregation and the corresponding spatial commitment registers, the DAs reset their spatial registers to zero and spatial commitment registers to one (to prepare for use in the next time instance). The computational overhead on ESP remains constant (Fig. 6.11) as it performs integrity verification on the reconstructed polynomials (which are the same across DAs) and utilizes the spatial commitment registers received from the DAs. As the spatially aggregated reading is derived by the ESP, it cannot derive the instantaneous reading and link it to a specific

SM from the SM_{List}, thereby preserving customer privacy.

6.4.10 Computational Overhead on Smart Meters due to Temporal Aggregation in a Malicious Setting

In the previous subsection 6.4.9, we were able to report spatially aggregated reading from SMs to the ESP via the DAs in a malicious setting. In this subsection, we adapt our proposed framework to support temporal aggregation that is required to provide billing functionalities. Importantly, we accomplish this without imposing any additional computational overhead on the SMs. We utilize the same set of shares and commitment generated through the application of Shamir's Secret Sharing Scheme and Commitment-based Scheme, as illustrated in Figure 6.9.

6.4.11 Computational Overhead on Dedicated Aggregators due to Temporal Aggregation in a Malicious Setting

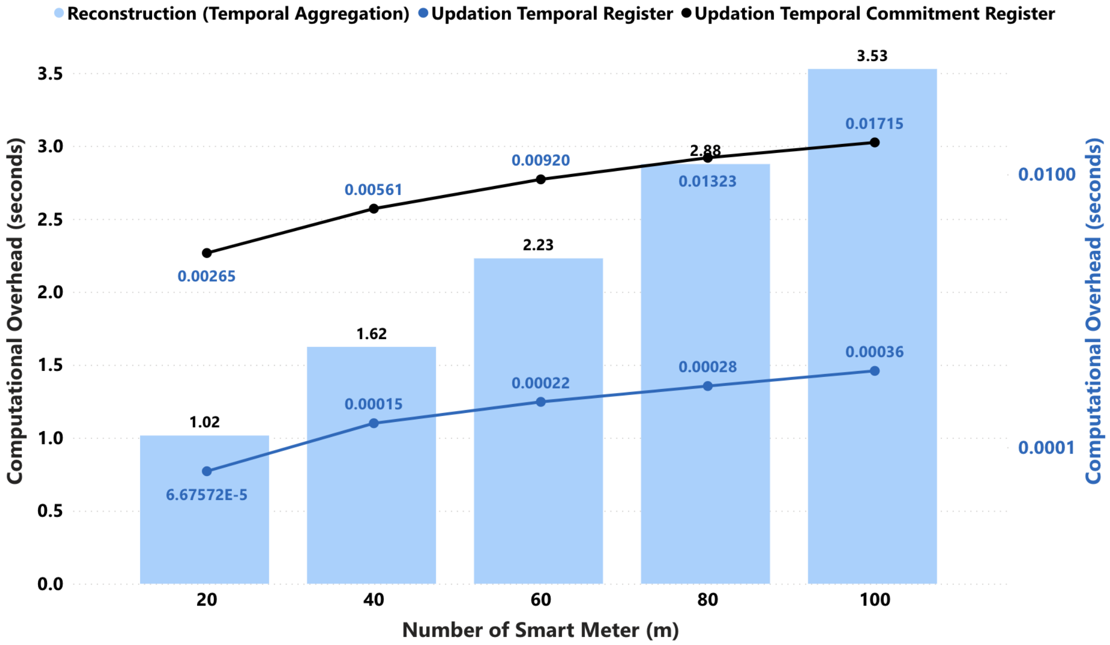


Figure 6.12: Computational Overhead on Dedicated Aggregators due to Temporal Aggregation in a Malicious Setting

In this subsection, we analyze the overall computational overhead associated with the DAs ($n = 3$) when reporting temporally aggregated readings to the ESP for varying numbers of SMs in a malicious setting. Each SM employs Shamir's Secret Sharing Scheme and Commitment-based Scheme to create shares and commitments, which are then reported to the corresponding DAs participating in temporal aggregation. Each DA has m temporal registers and m temporal commitment registers, both of which are responsible for aggregating the shares and commitments received from the corresponding SMs in the SM_{List}. As shown in Figure 6.12, the computational overhead associated with updating the temporal register and temporal commitment register increases proportionally with the increasing number of SMs. However, the temporal commitment register incurs a higher computational overhead than the temporal registers because the commitments are multiplied with each other, whereas for temporal registers, the shares are simply added together. This update process takes place at regular intervals of time (at every given instant). Once the temporal registers and temporal commitment registers for the designated DAs are updated, the DAs engage in SMPC by exchanging their temporal registers at $t = T$ (Note: temporal commitment registers are not exchanged with other DAs) and subsequently reconstructing the polynomials associated with the SMs for temporal aggregation. As the degree of secret sharing polynomials selected by the given set of SMs over time is the same ($k - 1 = n - 1$), the aggregated shares in the temporal registers represent the shares of the polynomial reconstructed in the Temporal Aggregation Phase. This is possible due to the homomorphic properties of the Shamir's Secret Sharing Scheme. As seen in Figure 6.11, the computational overhead associated with the reconstruction of polynomials increases with an increase in the number of SMs since more number of temporal registers are needed to be exchanged to reconstruct $(n - 1)$ degree polynomials associated with m SMs for temporal aggregation.

6.4.12 Overall Computational Overhead associated while reporting Temporally Aggregated Reading to the Electrical Service Provider in a Malicious Setting

In this subsection, we analyze the overall computational overhead associated with all the entities (SMs, DAs, ESP) for reporting temporally aggregated reading to the ESP in a malicious setting. We consider a given scenario with SMs ($m = \{20, 40, \ldots, 100\}$), DAs ($n = 3$) and, the ESP (Fig. 6.13). The SMs are responsible for creating a given set of shares and commitments based on the number of DAs participating in the temporal aggregation. As depicted in Figure 6.13, the overall computational overhead on the SMs due to Shamir's Secret Sharing Scheme and Commitment-based Scheme remains constant, as the SMs are expected to generate three shares (one per DA) and a commitment (same for all DAs). The shares and commitments distributed by the SMs are aggregated and stored in the corresponding temporal register and temporal commitment registers, respectively. Note: Each DA possesses one temporal register and one temporal commitment register for each SM. The computational overhead associated with updating the temporal register and temporal commitment register grows proportionally with increase in the number of SMs. After updating the temporal register and temporal commitment register, the DAs employ SMPC to reconstruct the polynomials related to temporal aggregation (Note: temporal commitment registers are not exchanged with other DAs). The overall computational overhead on the DAs for reconstructing the polynomial increases with an increase in the number of SMs, as more temporal registers are shared with each other and a more number of polynomials with degree $(k-1=n-1)$ are constructed as number of SMs increases (Fig. 6.13). The reconstructed polynomial related to each SM is the same across the DAs in the DA_{List} since they use the same basis polynomials and the same set of temporal registers. The constant of the reconstructed polynomial represents the computed billing by employing Time Of Use Billing for each SM, which is possible due to the homomorphic properties of Shamir's Secret Sharing Scheme.

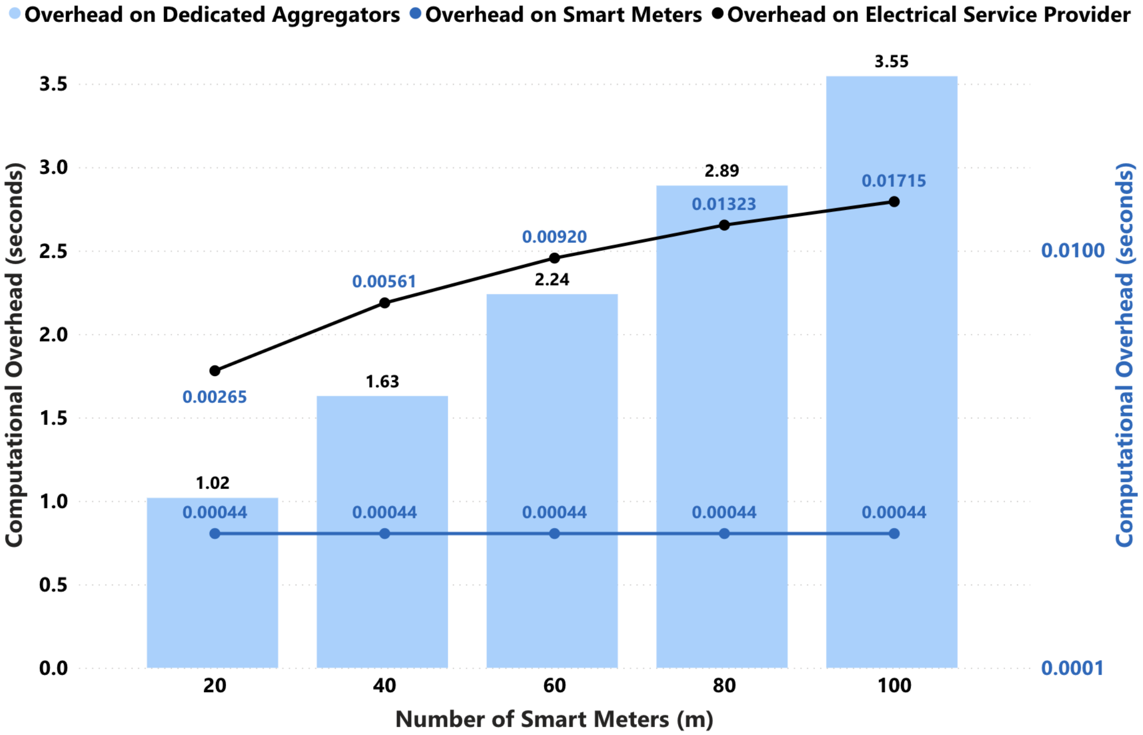


Figure 6.13: Overall Computational Overhead associated while reporting Temporally Aggregated Reading to the Electrical Service Provider in a Malicious Setting

After reporting the reconstructed polynomials regarding temporal aggregation and the corresponding temporal commitment registers, the DAs reset their temporal registers to zero and temporal commitment registers to one (to prepare for use in the next time instance). The computational overhead on ESP increases with increase in the number of SMs (Fig. 6.13) as it has to perform integrity verification on each of the reconstructed polynomials. As the computed bill by employing Time of Use Billing is obtained by the ESP, it cannot derive the instantaneous reading and link it to a specific SM from the SM_{List}, thereby preserving customer privacy.

6.4.13 Comparative Analysis of Computational Overhead on Smart Meter due to Spatial Aggregation

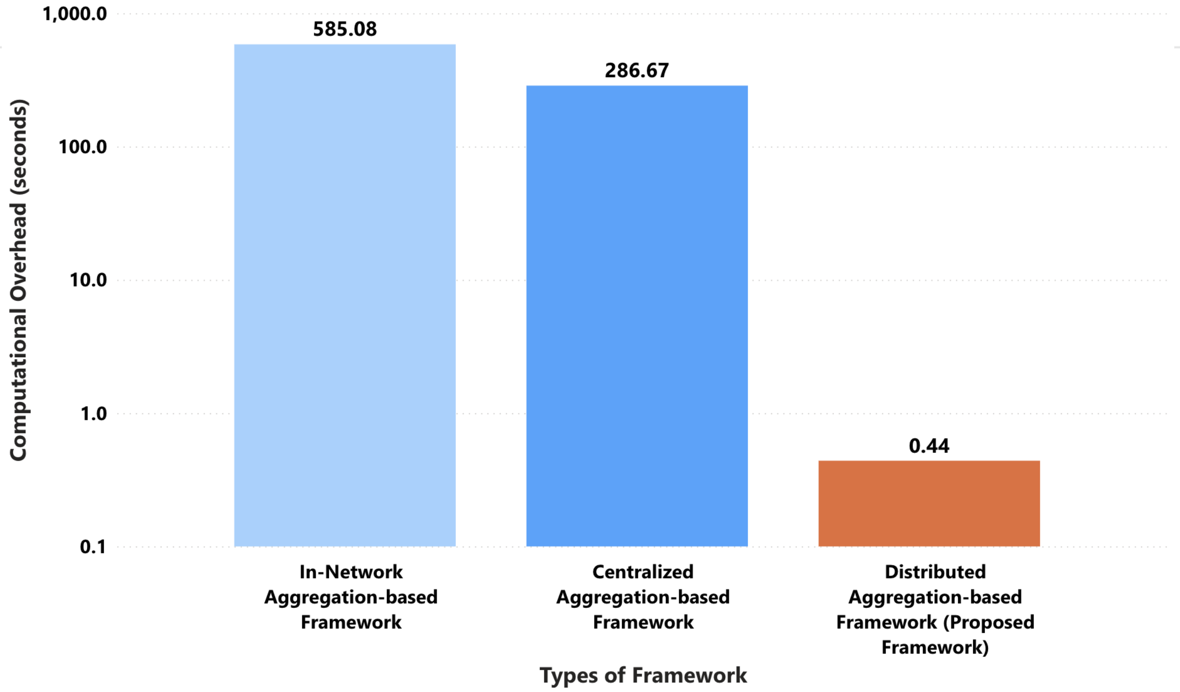


Figure 6.14: Comparative Analysis of Computational Overhead on Smart Meter due to Spatial Aggregation

Figure 6.14 presents an analysis of the computational overhead associated with spatial aggregation for our proposed framework in malicious setting and comparative frameworks in relation to a single SM. As observed, our proposed work exhibits significantly lower computational overhead compared to the In-Network Aggregation-based Framework and Centralized Aggregation-based

Frameworks. The Centralized and In-Network approaches experience high computational overhead due to the involvement of Paillier operations. Specifically, the Centralized Aggregation-based Framework [11] utilizes a single Paillier Encryption and trust key locking phase, whereas the In-Network Aggregation-based Framework [14] necessitates two Paillier Encryption operations: one for encrypting the reading and another for encrypting the decommitment value, in addition to computing the commitment. In contrast, our framework involves the computation of shares and the commitment for the instantaneous reading (high-frequency reading). The computational overhead for three DAs has been considered in our proposed framework (Fig. 6.14).

6.4.14 Comparative Analysis of Computational Overhead on Smart Meters due to Spatio-Temporal Aggregation

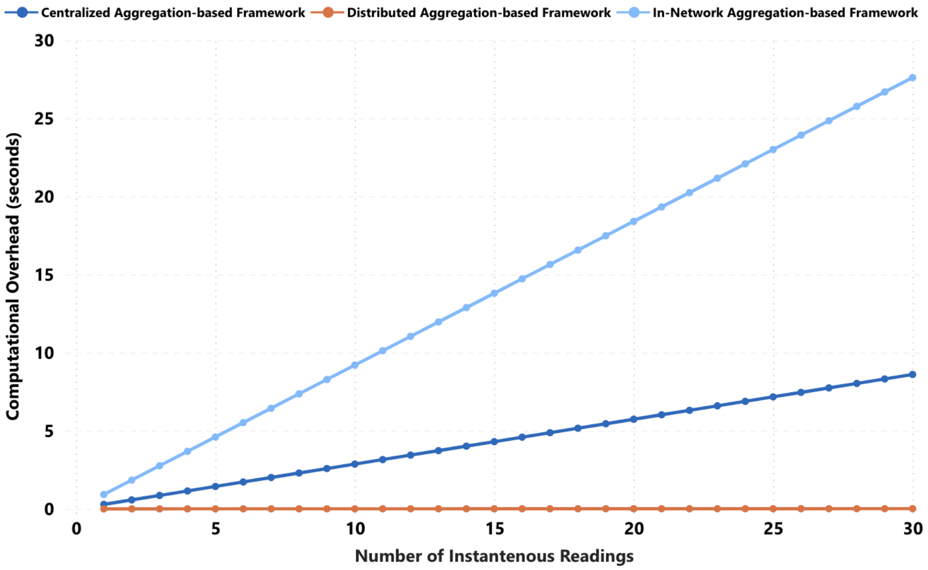


Figure 6.15: Comparative Analysis of Computational Overhead on Smart Meters due to Spatio-Temporal Aggregation

In Figure 6.15, our focus is on the computational overhead experienced by resource-constrained SMs over a span of 30 time instances ($T = 30$). Each instance represents the reporting of one high-frequency meter reading to the ESP while preserving privacy. As observed in Figure 6.15, the computational overhead for the In-Network Aggregation-based Framework is significantly higher since it requires two encryption operations, two homomorphic operations and, commitment generation for each time instance. In contrast, the Centralized Aggregation-based Framework involves simpler operations compared to the In-Network Aggregation-based Framework, namely one Homomorphic Encryption and locking with the trust key for each time instance. Our proposed scheme exhibits the least overhead as it generates shares and one commitment for each time instance.

6.4.15 Comparative Analysis of End-to-End Delay due to Spatial Aggregation

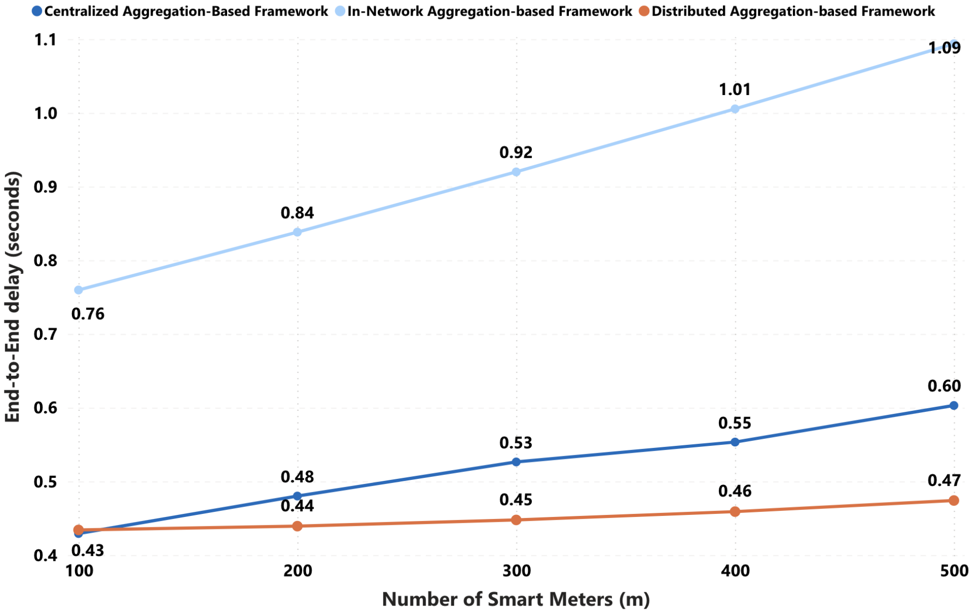


Figure 6.16: Comparative Analysis of End-to-End Delay due to Spatial Aggregation

As emphasized in Section 6.3, the end-to-end delay is a critical metric for enabling faster decision-making by the ESP. In Figure 6.16, we compare the end-to-end delay for spatial aggregation across a given range of SMs. It is evident that as the number of SMs increases, the end-to-end delay due to spatial aggregation also increases for all the frameworks. We consider both the In-Network Aggregation-based Framework at level 1, where each SM interacts with a parent SM responsible for computing aggregation and reporting to the ESP. The end-to-end delay exhibits a linear nature in the proposed framework, as the DA-based aggregation process is a linear function of the number of SMs. In contrast, the In-Network Aggregation-based Framework and Centralized Aggregation-based Frameworks demonstrate an exponential nature. Furthermore, our framework exhibits better scalability compared to other comparative frameworks. This indicates the potential of our framework in deployments involving a large number of SMs while maintaining the desired privacy level controlled by n. For a smaller number of SMs, it becomes crucial to identify the optimal number of DAs to strike a balance between privacy and the introduced delay due to aggregation.

6.5 Resilience Evaluation

In this section, we assess the security and privacy aspects of our framework against different collusion attack scenarios. We also examine how our framework is able to detect modifications to metering data while preserving privacy. It is important to note that the adversary has the capability to compromise up to $(n-1)$ DAs, as specified in our threat model (Section 5.2.2). Additionally, it is assumed that the adversary possesses prior knowledge of the initialization parameters received by the DAs from the ESP.

6.5.1 Resilience to Security Attacks

In this subsection, the adversary's primary objective is to disrupt the grid functionalities by providing false metering data to the ESP. As described in the threat model, the adversary has the capability to compromise up to $(n-1)$ DAs. The adversary possesses the ability to modify various components, including the incoming shares, spatial and temporal registers, spatial commitment registers, and temporal commitment registers. The timing of these modifications allows for further classification into two cases: 1) Pre Secure Multiparty Computation and 2) Post Secure Multiparty Computation.

If the adversary modifies the incoming shares and/or the registers associated with spatial and temporal aggregation associated with the compromised DA(s) before the SMPC, then the reconstructed polynomial will be constructed incorrectly across the entire set of DAs. Since the honest DA will be unaware of the modifications made by the malicious adversary, it would result in all DAs computing and reporting the wrong polynomial to the ESP, assuming there are no changes in the commitment registers. The ESP checks the computed commitment for spatially aggregated reading in the Integrity Verification Phase (Spatial Aggregation), and a similar process is followed for integrity of temporal metering data in the Integrity Verification Phase (Temporal Aggregation). However, the Integrity Verification processes will not succeed as the commitments will not match due to the binding property. This enables ESP to detect the modification in a privacy preserving manner and take appropriate measures. If the registers associated with commitments are modified by the malicious adversary, the honest DA will report the correct set of commitments to the ESP. Hence, the ESP will be alerted to the modification when it receives different commitments across the DAs. In this way, the ESP can detect the modification and take appropriate measures. If the registers associated with commitments are modified by the corrupted DAs, the honest DA will report the correct commitment to the ESP. However, the ESP will be alerted to the modification when it receives different commitments across the DAs.

Regarding post-SMPC modification, if the corrupted DAs initially participated honestly until the SMPC and later plan to deviate from the protocol, they can modify the reconstructed polynomial (with respect to spatial aggregation and/or temporal aggregation) and/or commitment register(s). However, since the honest DA will report the correct reconstructed polynomial and commitment to the ESP, the ESP will be able to detect the discrepancy and take appropriate measures accordingly. Therefore, our framework successfully detects any malicious modifications made by the adversary to shares and/or registers for a given instance of time.

Figure 6.17 represents a generalized flowchart of demonstrate resilience of our proposed framework to security attacks on spatial and/or temporal metering data.

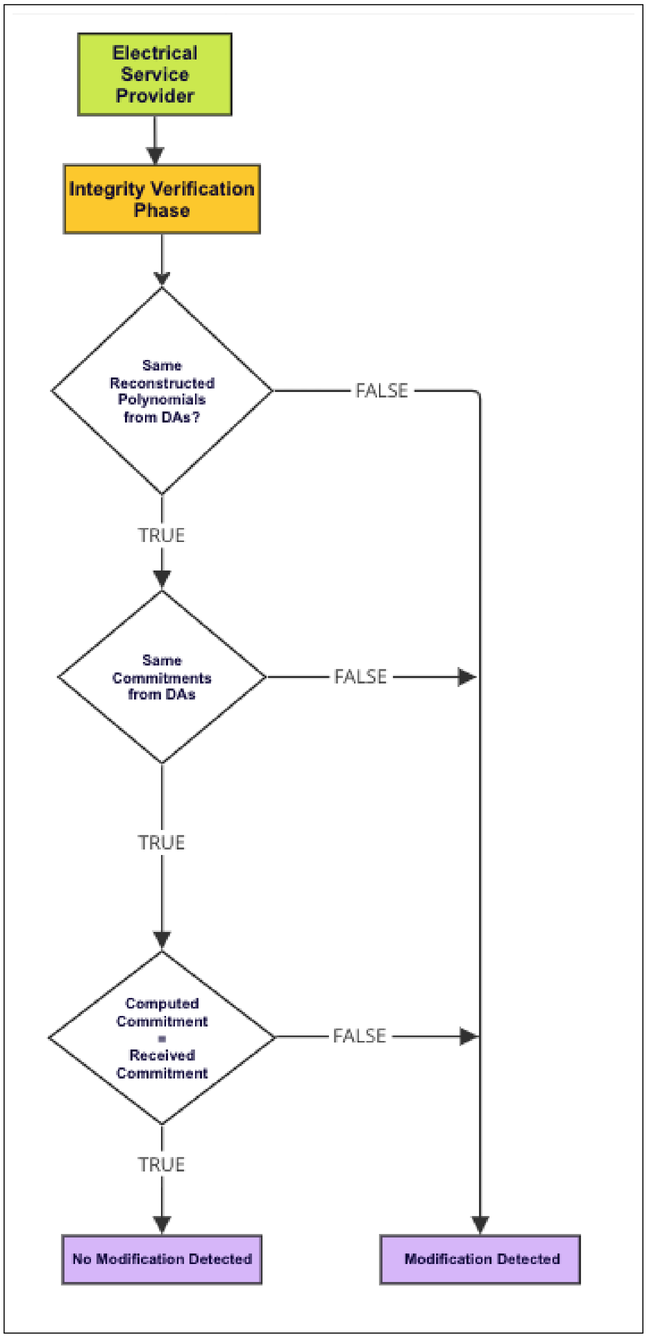

Electrical Service Provider
Integrity Verification Phase
Same Reconstructed Polynomials from DAs?
FALSE
TRUE
Same Commitments from DAs
FALSE
TRUE
Computed Commitment = Received Commitment
FALSE
TRUE
No Modification Detected
Modification Detected

Figure 6.17: A Generalized Flowchart to demonstrate Resilience of our Proposed Framework to Security Attacks

6.5.2 Resilience to Privacy Attacks

In this subsection, we evaluate the resilience of our framework against privacy attacks. As previously mentioned in Chapter 1, the adversary can breach customer privacy by derive useful patterns by accessing high-frequency metering data and linking it to the SM (customer). However, in our framework, the SMs are considered to be semi-honest and can only interact with the DAs, which means they cannot learn the readings of other SMs.As for the DAs, they only learn about the granular shares of each SMs. However, since the honest DA possesses the remaining share and never discloses it to the DAs during interactions involving aggregated shares in the spatial and temporal registers, the remaining share cannot be derived, thereby preserving customer privacy. Post SMPC, the DAs learn about the reconstructed polynomial representing either the spatially aggregated reading or the temporally aggregated reading over the given period of time, but they cannot derive the instantaneous reading from it. The ESP, too, cannot breach privacy as it only receives the aggregated readings in terms of spatial and temporal aggregation. Therefore, our framework is able to report spatio-temporal metering data to the ESP for providing grid and billing functionalities while preserving customer privacy.

6.6 Qualitative Analysis

We perform a qualitative comparative analysis of our framework with other aggregation-based privacy-preserving designs in the literature [13, 14, 22, 23, 25, 26, 31, 37, 40, 48, 53, 55, 56, 60, 63, 64, 65, 74, 76, 84, 86, 87, 93, 95, 100, 102, 103, 104, 108, 111] (Table 6.1). The comparison metrics are defined as follows:

- **Framework Design:** This refers to the type of infrastructure (In-network / Centralized / Distributed) that is deployed to support the aggregation functionality.
- **Threat Model:** This defines the type of adversarial model (active/passive) that is considered for the aggregation framework.
- **Smart Meter Overhead:** This metric represents the computational overhead in terms of complex cryptographic operations and/or reliance on other SMs for aggregation.
- **Integrity Verification (Spatial Aggregation):** This aspect focuses on the framework's ability to detect modifications of spatially aggregated metering data.

- **Integrity Verification (Temporal Aggregation):** This aspect focuses on the framework's ability to detect modifications to temporally aggregated metering data.

Table 6.1: Comparison of Aggregation-based Privacy-Preserving Frameworks

Framework	**Framework Design**	**Threat Model**	**Smart Meter Overhead**	**Integrity Verification (Spatial Aggregation)**	**Integrity Verification (Temporal Aggregation)**
Borges *et* al. [14]	In-network	Passive	High	Yes	Yes
Erkin *et* al. [25]	In-network	Passive	High	No	No
Hoepman Jaap [40]	In-network	Passive	High	No	No
Li *et* al. [53]	In-network	Passive	High	Yes	No
Tonyali *et* al. [100]	In-network	Passive	High	Yes	No
Song *et* al. [95]	In-network	Active	High	Yes	No
Bohli *et* al. [13]	Centralized	Passive	Low	No	No
Efthymiou *et* al. [23]	Centralized	Passive	Low	Yes	Yes
Rottondi *et* al. [87]	Centralized	Passive	Low	No	No
Guan *et* al. [37]	Centralized	Passive	High	Yes	Yes
Li *et* al. [55]	Centralized	Passive	High	Yes	No
Lu *et* al. [63]	Centralized	Passive	High	No	No
Lu *et* al. [64]	Centralized	Passive	High	Yes	No
Zuo *et* al. [111]	Centralized	Passive	High	Yes	No
Wang *et* al. [108]	Centralized	Passive	High	No	No
Gope *et* al. [31]	Centralized	Active	Low	Yes	Yes
Shen *et* al. [93]	Centralized	Active	Low	Yes	No
Lu *et* al. [65]	Centralized	Active	High	Yes	No
Li *et* al. [56]	Centralized	Active	High	Yes	Yes
Liu *et* al. [60]	Centralized	Active	High	Yes	No
Khan *et* al. [48]	Centralized	Active	High	Yes	Yes
Fan *et* al. [26]	Centralized	Active	High	Yes	No
Dong *et* al. [22]	Centralized	Active	High	Yes	Yes
Ni *et* al. [74]	Centralized	Active	High	Yes	No
Ohara *et* al. [76]	Centralized	Active	High	Yes	Yes
Rial *et* al. [84]	Centralized	Active	High	No	Yes
Rottondi *et* al. [86]	Distributed	Passive	Low	No	No
Wagh *et* al. [102]	Distributed	Passive	Low	No	No
Wagh *et* al. [103]	Distributed	Passive	Low	No	No
Wagh *et* al. [104]	Distributed	Passive	Low	No	No
Proposed Framework	**Distributed**	**Active**	**Low**	**Yes**	**Yes**

6.7 Summary

In this chapter, we addressed three identified research questions by implementing a proof of concept to evaluate the performance of our proposed framework. We then conducted a comparative analysis, considering metrics such as computational overhead and, end-to-end delay. We also evaluated the resilience of our framework to withstand security and privacy attacks, specifically within a malicious threat model with a dishonest majority of DAs. Finally, we presented a qualitative comparative analysis of our proposed framework with other related privacy-preserving frameworks from the literature, considering the identified metrics.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

In this dissertation, we addressed the privacy issues associated with the collection of high-frequency metering data in the smart grid. Through our critical literature analysis, we identified research gaps in existing aggregation-based frameworks, including: a) high computational overhead on resource-constrained smart meters (SMs), b) prone to single points of compromise due to reliance on a centralized entity, c) lack of support for dynamic billing integration while preserving customer privacy and, d) lack of integrity verification of spatio-temporal metering data. To overcome these limitations, we proposed a Distributed Aggregation-based Framework that considered the security, privacy, and infrastructure requirements outlined by NIST. The framework comprised SMs, Dedicated Aggregators (DAs), and Electrical Service Provider (ESP). To convert their instantaneous readings into shares, the SMs employed Shamir's Secret Sharing Scheme. These shares were then distributed to the corresponding DAs participating in spatial aggregation through secure multiparty computation (SMPC). The spatially aggregated reading was reported to the ESP in a privacy-preserving manner, as it played a crucial role in providing grid functionalities.

We further extended the framework to support temporal aggregation by utilizing the same set of shares employed in spatial aggregation. This expansion facilitated the inclusion of different types of billing tariffs, such as Flat Rate, Cumulative, and Time of Use Billing, within the framework. By distributing and offloading the majority of the work to the DAs, we addressed the issues related to high computational overhead and single points of compromise.

Moreover, we extended the threat model to a malicious setting involving a dishonest majority of the DAs. Since metering data was outsourced to the DAs for aggregation, it became susceptible to potential modifications. To mitigate this risk, we integrated a Commitment-based Scheme and imposed an additional constraint in the Shamir's Secret Sharing Scheme to enable integrity verification of spatio-temporal metering data. The framework effectively leveraged the homomorphic properties of the Shamir's Secret Sharing Scheme and the Commitment-based Scheme.

To evaluate the performance of our proposed framework, we developed a proof of concept utilizing both an embedded and cloud-based environment, using real-world SM traces. We conducted an analysis of the framework's performance compared to two relevant comparative frameworks from the literature. Additionally, we evaluated the resilience of our framework against potential security and privacy attacks involving a dishonest majority of the DAs. In the end, we presented a qualitative comparison of our proposed framework with other Aggregation-based Frameworks, considering the identified metrics.

To summarize the contributions of this dissertation, they are as follows:

- Development of a distributed privacy-preserving framework in a semi-honest setting
- Development of a distributed privacy-preserving framework in a malicious setting with a dishonest majority of DAs
- Creation of an open source proof of concept to compare our proposed privacy-preserving framework with other existing frameworks

7.2 Future Work

The conducted evaluations related to our proposed framework have motivated several areas worthy of future explorations. We summarize some of these areas as follows:

- **Authentication of SMs:** The current proposed framework focused on security and privacy requirements provided by NIST [32, 81]. Another aspect that has not been addressed is the authentication of SMs. Therefore, the current capabilities of our framework can be enhanced by integrating an authentication mechanism. By incorporating this functionality into the framework, impersonation attacks by dummy SMs introduced by adversaries can be prevented. The adversary may compromise the DAs and introduce dummy SMs to compromise the security and privacy of our framework, thereby disrupting the spatio-temporal functionalities. Although the existing Shamir's Secret Sharing Scheme [91] does not provide authentication, recent studies by the authors [38] indicate that the scheme can be extended to incorporate authentication.

- **Optimization of the Implementation:** By pre-computing the calculations related to the Share Creation Phase and generating a new set of shares for upcoming time instances, it is possible to reduce both the computational overhead and the end-to-end delay. This optimization would enable the ESP to make faster decisions, effectively mitigating the damage and disruption caused by adversaries.

- **Recoverability of Metering Data:** Our current framework estimates the metering data based on past knowledge. However, it is important to address the issue of fault tolerance and provide a mechanism for recovering metering data in the event of entity (DA(s)) failures. This can be achieved by integrating a robust fault-tolerant mechanism into our proposed framework, ensuring that metering data can be recovered and maintained even in the presence of entity failures.

- **Utilization of a Real Hardware:** In our current work, we utilized a Raspberry Pi to mimic a SM. However, exploring the use of a real SM with limited computing resources as hardware for evaluations in terms of computational overhead and end-to-end delay can be further explored.

- **Extending the Threat Model:** In our proposed research, we examine a malicious threat model where the adversary can compromise up to $(n-1)$ DAs, where n represents the total number of DAs in the framework. However, the framework can also be further explored to address the threat of malicious SMs reporting false metering data to the ESP.

It should be noted that our proposed framework can be generalized to other domains such as disaster response, environmental monitoring, and urban planning which require reporting of spatio-temporal data in a privacy-preserving manner.

Bibliography

[1] Asmaa Abdallah and Xuemin Sherman Shen. A lightweight lattice-based homomorphic privacy-preserving data aggregation scheme for smart grid. *IEEE Transactions on Smart Grid*, 9(1):396–405, 2018.

[2] Aysajan Abidin, Abdelrahaman Aly, Sara Cleemput, and Mustafa A Mustafa. An mpc-based privacy-preserving protocol for a local electricity trading market. In *Proceedings of the International Conference on Cryptology and Network Security*, pages 615–625. Springer, 2016.

[3] Gergely Ács and Claude Castelluccia. I have a dream! (differentially private smart metering). In *Proceedings of the International Workshop Information Hiding*, pages 118–132. Springer, 2011.

[4] David Adrian, Karthikeyan Bhargavan, Zakir Durumeric, Pierrick Gaudry, Matthew Green, J. Alex Halderman, Nadia Heninger, Drew Springall, Emmanuel Thomé, Luke Valenta, Benjamin VanderSloot, Eric Wustrow, Santiago Zanella-Béguelin, and Paul Zimmermann. Imperfect forward secrecy: How diffie-hellman fails in practice. In *Proceedings of the SIGSAC Conference on Computer and Communications Security*, page 5–17. ACM, 2015.

[5] Varunya Attasena, Jérôme Darmont, and Nouria Harbi. Secret sharing for cloud data security: a survey. *The International Journal on Very Large Data Bases*, 26(5):657–681, 2017.

[6] Michael Backes and Sebastian Meiser. Differentially private smart metering with battery recharging. In *Proceedings of the Data Privacy Management and Autonomous Spontaneous Security*, pages 194–212. Springer, 2014.

[7] Sainath Bandhakavi, Greg Thompson, and Joseph Beer. Application integration in utility smart metering using AWS. `https://aws.amazon.com/blogs/industries/`

application-integration-in-utility-smart-metering-using-aws/, Mar 2022. Last accessed on Nov, 2022.

[8] Antonella Barletta, Christian Callegari, Stefano Giordano, Michele Pagano, and Gregorio Procissi. Privacy preserving smart grid communications by verifiable secret key sharing. In *Proceedings of the International Conference on Computing and Network Communications*, pages 199–204. IEEE, 2015.

[9] Nipun Batra, Jack Kelly, Oliver Parson, Haimonti Dutta, William Knottenbelt, Alex Rogers, Amarjeet Singh, and Mani Srivastava. NILMTK: an open source toolkit for non-intrusive load monitoring. In *Proceedings of the International Conference on Future Energy Systems*, pages 265–276. ACM, 2014.

[10] Scott Becker. Estimated vs. Actual Readings – How They Affect Your NYSEG Bill. https://blog.solstice.us/solstice-blog/nyseg-bill-estimated-versus-actual-readings, Nov 2018. Last accessed on Dec, 2021.

[11] Rihem Ben Romdhane, Hamza Hammami, Mohamed Hamdi, and Tai-Hoon Kim. Privacy-preserving spatial and temporal data aggregation for smart metering. In *Proceedings of the Asia Conference on Electrical, Power and Computer Engineering*, pages 1–4, 2022.

[12] Samaresh Bera, Sudip Misra, and Joel JPC Rodrigues. Cloud computing applications for smart grid: A survey. *IEEE Transactions on Parallel and Distributed Systems*, 26(5):1477–1494, 2014.

[13] Jens-Matthias Bohli, Christoph Sorge, and Osman Ugus. A privacy model for smart metering. In *Proceedings of the International Conference on Communications Workshops*, pages 1–5. IEEE, 2010.

[14] Fábio Borges, Denise Demirel, Leon Böck, Johannes Buchmann, and Max Mühlhäuser. A privacy-enhancing protocol that provides in-network data aggregation and verifiable smart meter billing. In *Symposium on Computers and Communications*, pages 1–6. IEEE, 2014.

[15] Le Chen, Rongxing Lu, and Zhenfu Cao. PDAFT: A privacy-preserving data aggregation scheme with fault tolerance for smart grid communications. *Peer-to-Peer networking and applications*, 8(6):1122–1132, 2015.

[16] Ronald Cramer, Ivan Bjerre Damgård, and Jesper Buus Nielsen. *Secure multiparty computation*. Cambridge University Press, 2015.

[17] George Danezis, Cédric Fournet, Markulf Kohlweiss, and Santiago Zanella-Béguelin. Smart meter aggregation via secret-sharing. In *Proceedings of the Workshop on Smart energy grid security*, pages 75–80. ACM, 2013.

[18] George Danezis, Markulf Kohlweiss, and Alfredo Rial. Differentially private billing with rebates. In *Proceedings of the International Workshop on Information Hiding*, pages 148–162. Springer, 2011.

[19] Fábio Borges de Oliveira. *On privacy-preserving protocols for smart metering systems*. Springer, 2017.

[20] Tassos Dimitriou and Mohamad Khattar Awad. Secure and scalable aggregation in the smart grid resilient against malicious entities. *Ad Hoc Networks*, 50:58–67, 2016.

[21] Susen Döbelt, Markus Jung, Marc Busch, and Manfred Tscheligi. Consumers' privacy concerns and implications for a privacy preserving Smart Grid architecture—Results of an Austrian study. *Energy Research & Social Science*, 9:137–145, 2015.

[22] Yihui Dong, Jian Shen, Sai Ji, Rongxin Qi, and Shuai Liu. A novel appliance-based secure data aggregation scheme for bill generation and demand management in smart grids. *Connection Science*, 33(4):1116–1137, 2021.

[23] Costas Efthymiou and Georgios Kalogridis. Smart grid privacy via anonymization of smart metering data. In *International Conference on Smart Grid Communications*, pages 238–243. IEEE, 2010.

[24] Zekeriya Erkin, Juan Ramón Troncoso-Pastoriza, Reginald L Lagendijk, and Fernando Pérez-González. Privacy-preserving data aggregation in smart metering systems: An overview. *IEEE Signal Processing Magazine*, 30(2):75–86, 2013.

[25] Zekeriya Erkin and Gene Tsudik. Private computation of spatial and temporal power consumption with smart meters. In *Proceedings of the International Conference on Applied Cryptography and Network Security*, pages 561–577. Springer, 2012.

[26] Chun-I Fan, Shi-Yuan Huang, and Yih-Loong Lai. Privacy-enhanced data aggregation scheme against internal attackers in smart grid. *IEEE Transactions on Industrial informatics*, 10(1):666–675, 2013.

[27] Xi Fang, Satyajayant Misra, Guoliang Xue, and Dejun Yang. Smart grid—the new and improved power grid: A survey. *IEEE Communications Surveys & Tutorials*, 14(4):944–980, 2011.

[28] A. Faruqui and C. Bourbonnais. The tariffs of tomorrow: Innovations in rate designs. *IEEE Power and Energy Magazine*, 18(3):18–25, 2020.

[29] Giulio Giaconi, Deniz Gündüz, and H Vincent Poor. Optimal demand-side management for joint privacy-cost optimization with energy storage. In *Proceedings of the International Conference on Smart Grid Communications*, pages 265–270. IEEE, 2017.

[30] Giulio Giaconi, Deniz Gunduz, and H Vincent Poor. Privacy-aware smart metering: Progress and challenges. *IEEE Signal Processing Magazine*, 35(6):59–78, 2018.

[31] Prosanta Gope and Biplab Sikdar. Lightweight and privacy-friendly spatial data aggregation for secure power supply and demand management in smart grids. *IEEE Transactions on Information Forensics and Security*, 14(6):1554–1566, 2018.

[32] Avi Gopstein, Cuong Nguyen, Cheyney O'Fallon, Nelson Hastings, and David A. Wollman. NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 4.0. `https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=931882`, February 2021. Last accessed on Jan, 2023.

[33] Ulrich Greveler, Peter Glösekötterz, Benjamin Justusy, and Dennis Loehr. Multimedia content identification through smart meter power usage profiles. In *Proceedings of the International Conference on Information and Knowledge Engineering*, pages 1–8. The Steering Committee of The World Congress in Computer Science, 2012.

[34] Zhitao Guan, Jing Li, Longfei Wu, Yue Zhang, Jun Wu, and Xiaojiang Du. Achieving efficient and secure data acquisition for cloud-supported internet of things in smart grid. *IEEE Internet of Things Journal*, 4(6):1934–1944, 2017.

[35] Zhitao Guan, Guanlin Si, Jun Wu, Liehuang Zhu, Zijian Zhang, and Yinglong Ma. Utility-privacy tradeoff based on random data obfuscation in internet of energy. *IEEE Access*, 5:3250–3262, 2017.

[36] Zhitao Guan, Guanlin Si, Xiaosong Zhang, Longfei Wu, Nadra Guizani, Xiaojiang Du, and Yinglong Ma. Privacy-preserving and efficient aggregation based on blockchain for power grid communications in smart communities. *IEEE Communications Magazine*, 56(7):82–88, 2018.

[37] Zhitao Guan, Guanlin Si, Xiaosong Zhang, Longfei Wu, Nadra Guizani, Xiaojiang Du, and Yinglong Ma. Privacy-preserving and efficient aggregation based on blockchain for power grid communications in smart communities. *IEEE Communications Magazine*, 56(7):82–88, 2018.

[38] Kishor Datta Gupta, Md Lutfar Rahman, Dipankar Dasgupta, and Subash Poudyal. Shamir's secret sharing for authentication without reconstructing password. In *Proceedings of the Computing and Communication Workshop and Conference*, pages 0958–0963. IEEE, 2020.

[39] Xingze He, Xinwen Zhang, and C-C Jay Kuo. A distortion-based approach to privacy-preserving metering in smart grids. *IEEE Access*, 1:67–78, 2013.

[40] Jaap-Henk Hoepman. Privacy friendly aggregation of smart meter readings, even when meters crash. In *Proceedings of the Workshop on Cyber-Physical Security and Resilience in Smart Grids*, pages 3–7. ACM, 2017.

[41] Marek Jawurek, Martin Johns, and Florian Kerschbaum. Plug-in privacy for smart metering billing. In *Proceedings of the International Symposium on Privacy Enhancing Technologies Symposium*, pages 192–210. Springer, 2011.

[42] Hyo Jin Jo, In Seok Kim, and Dong Hoon Lee. Efficient and privacy-preserving metering protocols for smart grid systems. *IEEE Transactions on Smart Grid*, 7(3):1732–1742, 2015.

[43] Marc Joye and Benoît Libert. A scalable scheme for privacy-preserving aggregation of time-series data. In *Proceedings of the International Conference on Financial Cryptography and Data Security*, pages 111–125. Springer, 2013.

[44] Malik Ali Judge, Asif Khan, Awais Manzoor, and Hasan Ali Khattak. Overview of smart grid implementation: Frameworks, impact, performance and challenges. *Journal of Energy Storage*, 49:104056, 2022.

[45] Deniz Gündüz1 Georgios Kalogridis and Mustafa A Mustafa. Privacy in smart metering systems. In *Proceedings of the International Workshop on Information Forensics and Security*, pages 1–100. IEEE, 2015.

[46] Georgios Kalogridis, Costas Efthymiou, Stojan Z Denic, Tim A Lewis, and Rafael Cepeda. Privacy for smart meters: Towards undetectable appliance load signatures. In *Proceedings of the International Conference on Smart Grid Communications*, pages 232–237. IEEE, 2010.

[47] Georgios Kalogridis, Zhong Fan, and Sagar Basutkar. Affordable privacy for home smart meters. In *Proceedings of the International Symposium on Parallel and Distributed Processing with Applications Workshops*, pages 77–84. IEEE, 2011.

[48] Hayat Mohammad Khan, Abid Khan, Farhana Jabeen, and Arif Ur Rahman. Privacy preserving data aggregation with fault tolerance in fog-enabled smart grids. *Sustainable Cities and Society*, 64:102522, 2021.

[49] Young-Sam Kim and Joon Heo. Device authentication protocol for smart grid systems using homomorphic hash. *Journal of Communications and Networks*, 14(6):606–613, 2012.

[50] J Zico Kolter and Matthew J Johnson. REDD: A public data set for energy disaggregation research. In *Proceedings of the Workshop on data mining applications in sustainability*, volume 25, pages 59–62, 2011.

[51] Klaus Kursawe, George Danezis, and Markulf Kohlweiss. Privacy-friendly aggregation for the smart-grid. In *Proceedings of the International Symposium on Privacy Enhancing Technologies Symposium*, pages 175–191. Springer, 2011.

[52] Michael LeMay, Rajesh Nelli, George Gross, and Carl A Gunter. An integrated architecture for demand response communications and control. In *Proceedings of the International Conference on System Sciences*, pages 174–174. IEEE, 2008.

[53] Fengjun Li and Bo Luo. Preserving data integrity for smart grid data aggregation. In *Proceedings of the International Conference on Smart Grid Communications*, pages 366–371. IEEE, 2012.

[54] Fengjun Li, Bo Luo, and Peng Liu. Secure information aggregation for smart grids using homomorphic encryption. In *Proceedings of the International Conference on Smart Grid Communications*, pages 327–332. IEEE, 2010.

[55] Hongwei Li, Xiaodong Lin, Haomiao Yang, Xiaohui Liang, Rongxing Lu, and Xuemin Shen. Eppdr: An efficient privacy-preserving demand response scheme with adaptive key evolution in smart grid. *IEEE Transactions on Parallel and Distributed Systems*, 25(8):2053–2064, 2014.

[56] Kunchang Li, Yifan Yang, Shuhao Wang, Runhua Shi, and Jianbin Li. A lightweight privacy-preserving and sharing scheme with dual-blockchain for intelligent pricing system of smart grid. *Computers & Security*, 103:102189, 2021.

[57] Yehuda Lindell. Secure multiparty computation. *Communications of the ACM*, 64(1):86–96, 2020.

[58] Benjamin Lipton. *Smart grid privacy through distributed trust.* Rochester Institute of Technology, New York, United States, 2017.

[59] Jing Liu, Yang Xiao, Shuhui Li, Wei Liang, and CL Philip Chen. Cyber security and privacy issues in smart grids. *IEEE Communications Surveys & Tutorials*, 14(4):981–997, 2012.

[60] Yining Liu, Wei Guo, Chun-I Fan, Liang Chang, and Chi Cheng. A practical privacy-preserving data aggregation (3pda) scheme for smart grid. *IEEE Transactions on Industrial Informatics*, 15(3):1767–1774, 2018.

[61] Dylan Locsin and John Pressley. Duke energy and aws are innovating for a smarter, cleaner energy future. `https://aws.amazon.com/blogs/industries/duke-energy-and-aws-are-innovating-for-a-smarter-cleaner-energy-future/`, Nov 2022. Last accessed on Jan, 2023.

[62] Rongxing Lu. Privacy-preserving data aggregation with data integrity and fault tolerance. In *Privacy-Enhancing Aggregation Techniques for Smart Grid Communications*, pages 153–177. Springer, 2016.

[63] Rongxing Lu. Privacy-preserving multifunctional data aggregation. In *Privacy-Enhancing Aggregation Techniques for Smart Grid Communications*, pages 85–110. Springer, 2016.

[64] Rongxing Lu, Kevin Heung, Arash Habibi Lashkari, and Ali A Ghorbani. A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced iot. *IEEE access*, 5:3302–3312, 2017.

[65] Rongxing Lu, Xiaohui Liang, Xu Li, Xiaodong Lin, and Xuemin Shen. Eppa: An efficient and privacy-preserving aggregation scheme for secure smart grid communications. *IEEE Transactions on Parallel and Distributed Systems*, 23(9):1621–1631, 2012.

[66] Lingjuan Lyu, Karthik Nandakumar, Ben Rubinstein, Jiong Jin, Justin Bedo, and Marimuthu Palaniswami. Ppfa: Privacy preserving fog-enabled aggregation in smart grid. *IEEE Transactions on Industrial Informatics*, 14(8):3733–3744, 2018.

[67] Felix Gomez Marmol, Christoph Sorge, Osman Ugus, and Gregorio Martínez Pérez. Do not snoop my habits: preserving privacy in the smart grid. *IEEE Communications Magazine*, 50(5):166–172, 2012.

[68] Daisuke Mashima, Aidana Serikova, Yao Cheng, and Binbin Chen. Towards quantitative evaluation of privacy protection schemes for electricity usage data sharing. *ICT Express*, 4(1):35–41, 2018.

[69] Michael Matz. THE GRID IS MOVING TO THE CLOUD. `https://eprijournal.com/the-grid-is-moving-to-the-cloud/`, May 2021.

[70] Akanksha Maurya, Alper Sinan Akyurek, Baris Aksanli, and Tajana Simunic Rosing. Time-series clustering for data analysis in smart grid. In *Proceedings of the International Conference on Smart Grid Communications*, pages 606–611. IEEE, 2016.

[71] Stephen McLaughlin, Patrick McDaniel, and William Aiello. Protecting consumer privacy from electric load monitoring. In *Proceedings of the Conference on Computer and Communications Security*, pages 87–98. ACM, 2011.

[72] Amin Mohammadali and Mohammad Sayad Haghighi. A privacy-preserving homomorphic scheme with multiple dimensions and fault tolerance for metering data aggregation in smart grid. *IEEE Transactions on Smart Grid*, 12(6):5212–5220, 2021.

[73] Andrés Molina-Markham, Prashant Shenoy, Kevin Fu, Emmanuel Cecchet, and David Irwin. Private memoirs of a smart meter. In *Proceedings of the Workshop on Embedded Sensing Systems for Energy Efficiency in Building*, pages 61–66. ACM, 2010.

[74] Jianbing Ni, Khalid Alharbi, Xiaodong Lin, and Xuemin Shen. Security-enhanced data aggregation against malicious gateways in smart grid. In *In Proceedings of the Global Communications Conference*, pages 1–6. IEEE, 2015.

[75] University of Massachusetts. Umass repository. `http://traces.cs.umass.edu`, Jan 2013. Last accessed on Jul, 2022.

[76] Kazuma Ohara, Yusuke Sakai, Fumiaki Yoshida, Mitsugu Iwamoto, and Kazuo Ohta. Privacy-preserving smart metering with verifiability for both billing and energy management. In *Proceedings of the Workshop on ASIA Public-Key Cryptography*, pages 23–32. ACM, 2014.

[77] Pascal Paillier. Public-key cryptosystems based on composite degree residuosity classes. In *Proceedings of the International Conference on the Theory and Applications of Cryptographic techniques*, pages 223–238, Prague, Czech Republic, 1999. Springer.

[78] Marco Pau, Edoardo Patti, Luca Barbierato, Abouzar Estebsari, Enrico Pons, Ferdinanda Ponci, and Antonello Monti. A cloud-based smart metering infrastructure for distribution grid services and automation. *Sustainable Energy, Grids and Networks*, 15:14–25, 2018.

[79] Torben Pryds Pedersen. Non-interactive and Information-theoretic Secure Verifiable Secret Sharing. In *Proceedings of the International Cryptology Conference*, pages 129–140. Springer, 1991.

[80] Ronald Petrlic. A privacy-preserving concept for smart grids. *Sicherheit in vernetzten Systemen*, 18:A1–A14, 2010.

[81] Victoria Pillitteri and Tanya Brewer. Cybersecurity User's Guide to the Guidelines for Smart Grid Cybersecurity NISTIR 7628 Vol. 1. `https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=915613`, February 2014. Last accessed on Jan, 2021.

[82] Elias Leake Quinn. Privacy and the new energy infrastructure. *Available at Social Science Research Network 1370731*, 2009.

[83] S Raj Rajagopalan, Lalitha Sankar, Soheil Mohajer, and H Vincent Poor. Smart meter privacy: A utility-privacy framework. In *Proceedings of the International Conference on Smart Grid Communications*, pages 190–195. IEEE, 2011.

[84] Alfredo Rial and George Danezis. Privacy-preserving smart metering. In *Proceedings of the workshop on Privacy in the electronic society*, pages 49–60. ACM, 2011.

[85] Cristina Rottondi, Giacomo Verticale, and Antonio Capone. A security framework for smart metering with multiple data consumers. In *Proceedings of the INFOCOM Workshops*, pages 103–108. IEEE, 2012.

[86] Cristina Rottondi, Giacomo Verticale, and Antonio Capone. Privacy-preserving smart metering with multiple data consumers. *Computer Networks*, 57(7):1699–1713, 2013.

[87] Cristina Rottondi, Giacomo Verticale, and Christoph Krauss. Distributed privacy-preserving aggregation of metering data in smart grids. *IEEE Journal on Selected Areas in Communications*, 31(7):1342–1354, 2013.

[88] Sushmita Ruj and Amiya Nayak. A decentralized security framework for data aggregation and access control in smart grids. *IEEE Transactions on Smart Grid*, 4(1):196–205, 2013.

[89] Lalitha Sankar, S Raj Rajagopalan, and Soheil Mohajer. Smart meter privacy: A theoretical framework. *IEEE Transactions on Smart Grid*, 4(2):837–846, 2013.

[90] Dongwon Seo, Heejo Lee, and Adrian Perrig. Secure and efficient capability-based power management in the smart grid. In *Proccedings of the International Symposium on Parallel and Distributed Processing with Applications Workshops*, pages 119–126. IEEE, 2011.

[91] Adi Shamir. How to share a secret. *Communications of the Association for Computing Machinery (ACM)*, 22(11):612–613, 1979.

[92] Mohammadhadi Shateri, Francisco Messina, Pablo Piantanida, and Fabrice Labeau. Real-time privacy-preserving data release for smart meters. *IEEE Transactions on Smart Grid*, 11(6):5174–5183, 2020.

[93] Gang Shen, Yixin Su, Danhong Zhang, Cheng Zhang, and Mingwu Zhang. A robust, distributed, and privacy-preserving aggregation scheme for smart grid communications. *Journal of the Chinese Institute of Engineers*, 42(1):54–65, 2019.

[94] Elaine Shi, TH Hubert Chan, Eleanor Rieffel, Richard Chow, and Dawn Song. Privacy-preserving aggregation of time-series data. In *Proceedings of the Network and Distributed System Security Symposium*, pages 1–17. Internet Society, 2011.

[95] Jingcheng Song, Yining Liu, Jun Shao, and Chunming Tang. A dynamic membership data aggregation (dmda) protocol for smart grid. *IEEE Systems Journal*, 14(1):900–908, 2019.

[96] Siddharth Sridhar, Adam Hahn, and Manimaran Govindarasu. Cyber-physical system security for the electric power grid. *Proceedings of the IEEE*, 100(1):210–224, 2012.

[97] Onur Tan, Jesús Gómez-Vilardebó, and Deniz Gündüz. Privacy-cost trade-offs in demand-side management with storage. *IEEE Transactions on Information Forensics and Security*, 12(6):1458–1469, 2017.

[98] Onur Tan, Deniz Gunduz, and H Vincent Poor. Increasing smart meter privacy through energy harvesting and storage devices. *IEEE Journal on Selected Areas in Communications*, 31(7):1331–1341, 2013.

[99] Song Tan, Debraj De, Wen-Zhan Song, Junjie Yang, and Sajal K Das. Survey of security advances in smart grid: A data driven approach. *IEEE Communications Surveys & Tutorials*, 19(1):397–422, 2017.

[100] Samet Tonyali, Kemal Akkaya, Nico Saputro, A Selcuk Uluagac, and Mehrdad Nojoumian. Privacy-preserving protocols for secure and reliable data aggregation in IoT-enabled smart metering systems. *Future Generation Computer Systems*, 78:547–557, 2018.

[101] Samet Tonyali, Ozan Cakmak, Kemal Akkaya, Mohamed MEA Mahmoud, and Ismail Guvenc. Secure data obfuscation scheme to enable privacy-preserving state estimation in smart grid ami networks. *IEEE Internet of Things Journal*, 3(5):709–719, 2015.

[102] Gaurav S Wagh, Sahil Gupta, and Sumita Mishra. A distributed privacy preserving framework for the smart grid. In *Proceedings of the Power & Energy Society Innovative Smart Grid Technologies Conference*, pages 1–5. IEEE, 2020.

[103] Gaurav S Wagh and Sumita Mishra. A cyber-resilient privacy framework for the smart grid with dynamic billing capabilities. In *Proceedings of the International Conference on*

Communications, Control, and Computing Technologies for Smart Grids, pages 1–6. IEEE, 2020.

[104] Gaurav S Wagh and Sumita Mishra. Divide & Conquer: A privacy safeguarding framework for the smart grid. In *Proceedings of the International Conference on Communications*, pages 1–6. IEEE, 2021.

[105] Gaurav S Wagh and Sumita Mishra. A distributed privacy-preserving integrity verification framework for the smart grid. In *Proceedings of the International Symposium on Technologies for Homeland Security*, pages 1–7. IEEE, 2022.

[106] Gaurav S Wagh and Sumita Mishra. A distributed approach to privacy-preservation and integrity assurance of smart metering data. In *Proceedings of the International Conference on Future Energy Systems*, page 60–65. ACM, 2023.

[107] Gaurav Shivaji Wagh. A privacy safeguarding framework for the smartgrid. `https://github.com/gauravwagh16193/APrivacySafeguardingFrameworkForSmartGrid`.

[108] Xiaodi Wang, Yining Liu, and Kim-Kwang Raymond Choo. Fault-tolerant multisubset aggregation scheme for smart grid. *IEEE Transactions on Industrial Informatics*, 17(6):4065–4072, 2020.

[109] Ye Yan, Yi Qian, and Hamid Sharif. A secure data aggregation and dispatch scheme for home area networks in smart grid. In *Proccedings of the Global Telecommunications Conference*, pages 1–6. IEEE, 2011.

[110] Lei Zhang and Jing Zhang. Publicly verifiable spatial and temporal aggregation scheme against malicious aggregator in smart grid. *Applied Sciences*, 9(3):490–510, 2019.

[111] Xiangjian Zuo, Lixiang Li, Haipeng Peng, Shoushan Luo, and Yixian Yang. Privacy-preserving multidimensional data aggregation scheme without trusted authority in smart grid. *IEEE Systems Journal*, 15(1):395–406, 2020.

Appendices

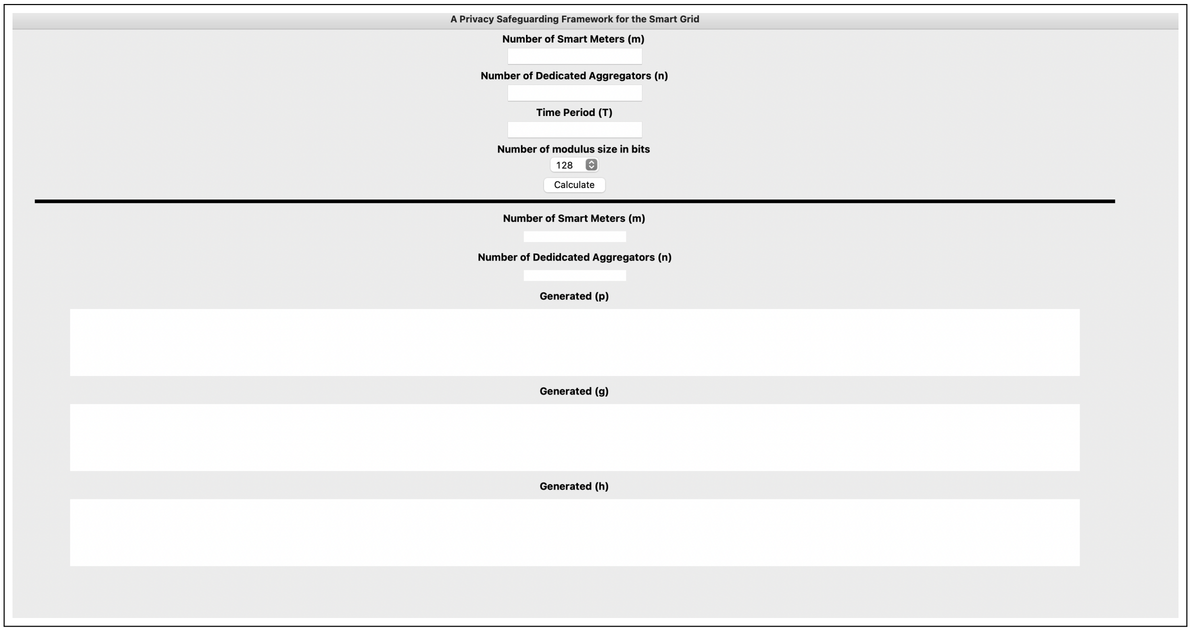
Appendix A

First Appendix

A.1 Graphical User Interface

Figure A.1, represents the Graphical User Interface (GUI) for our developed proof of concept. It takes the following inputs from the user and generates the initialization parameters for our proposed framework. The backend of the GUI consumes the sanitized dataset mentioned in A.2.

- Number of Smart Meters (m)
- Number of Dedicated Aggregators (n)
- Total Time Period (T)
- Number of bits



A Privacy Safeguarding Framework for the Smart Grid

Number of Smart Meters (m)

Number of Dedicated Aggregators (n)

Time Period (T)

Number of modulus size in bits

128

Number of Smart Meters (m)

Number of Dedidcated Aggregators (n)

Generated (p)

Generated (g)

Generated (h)

Figure A.1: Graphical User Interface of our Proposed Work

The following initialization parameters are generated at the backend and stored on the cloud deployed using AWS.

- Prime number (p)
- Commitment creation parameter (g)
- Commitment creation parameter (h)
- List of Dedicated Aggregators (DA_{List})
- List of Smart Meters (SM_{List})
- Tariff ($Tariff$)
- Time period (T)
- Number of SM (m)
- Number of DA (n)
- Maximum coefficient
- Instantaneous readings
- Bits
- Seed List ($Seed_{List}$)

A.2 UMass Dataset

The UMass Dataset is sanitized in a Python environment, where operations are performed to identify null (Fig. A.2) and missing values (Fig. A.3). If any missing values are found, they are filled using the built-in forward-fill method in Python. Since our research focuses on reporting metering data from the Smart Meter (SM) to the Electrical Service Provider (ESP), we aggregate the individual appliance readings to calculate the instantaneous reading for the specific SM at each moment in time.

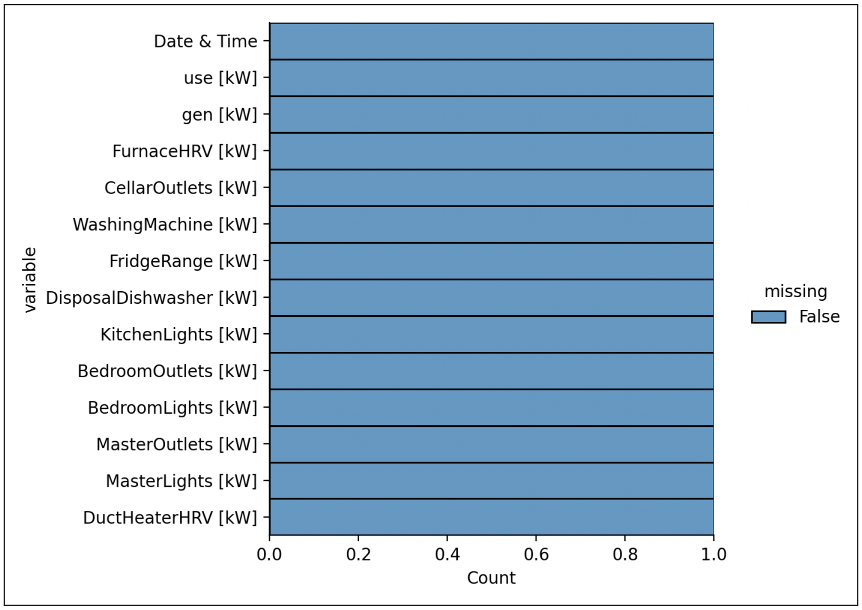


Figure A.2: Visualization of missing data of UMass Dataset using Bar Plot

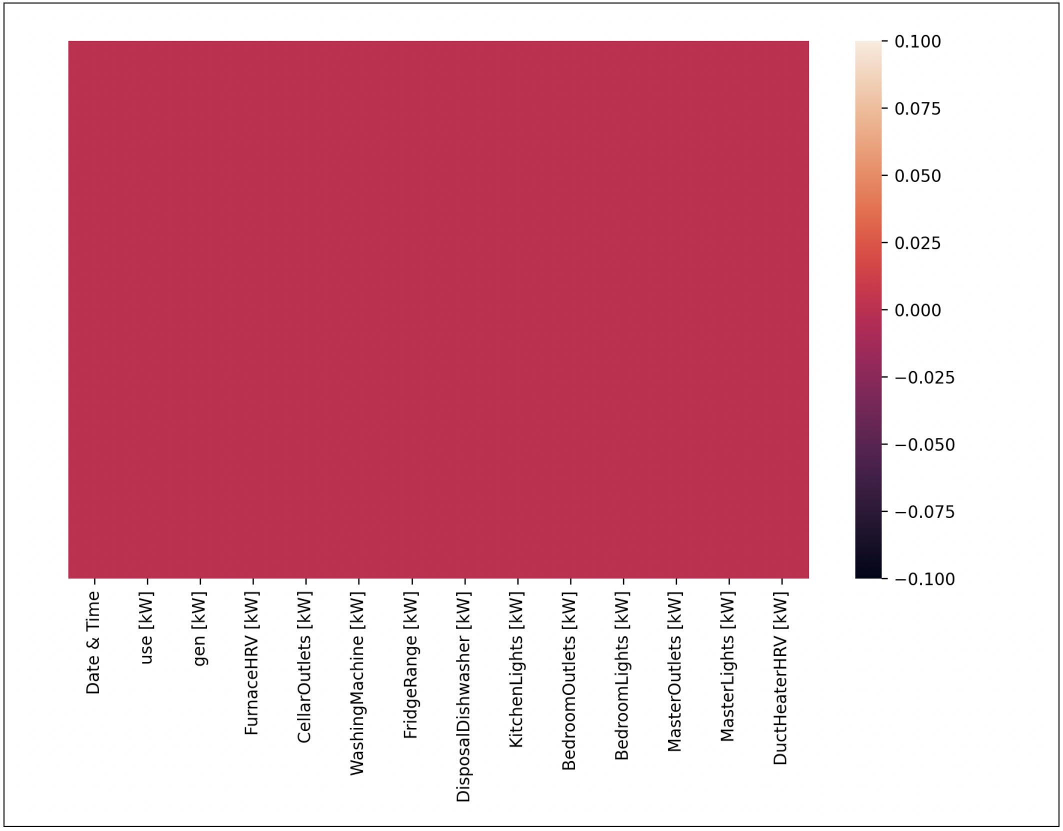


Figure A.3: Visualization of null data of UMass Dataset using Heatmap

,