Achieving Continuous Privacy-Preserving Histogram Query in Smart Grid Communications

by

Kingsley Kwame Baah Larbi

B.Sc. degree in Computer Science, Kwame Nkrumah University of Science and Technology, Ghana, 2014

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Computer Science

In the Graduate Academic Unit of Computer Science

Supervisor(s):	Rongxing Lu, Ph.D, Computer Science
Examining Board:	Saqib Hakak, Ph.D, Faculty of Computer Science, Chair
	Suprio Ray, Ph.D, Faculty of Computer Science
	Zhen Lei, Ph.D, Department of Civil Engineering

This thesis is accepted by the
Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

December, 2020

Abstract

Privacy has been taken very seriously in recent times with the introduction of the General Data Protection Regulation (GDPR) by the European Union (EU) and the addition of new rules to the existing Personal Information Protection and Electronic Documents Act (PIPEDA act) by the Canadian government. Governments are strengthening their stance on privacy by ensuring that organizations respect individuals' privacy rights. As such privacy within the smart grid in terms of usage data of customers must be treated with utmost importance. It is in this vein that this research is embarked on, a thesis which involves achieving a continuous histogram query in smart grid communications in a privacy preserving manner. Specifically, this research first gives a brief description of the smart grid from the aspects of characteristics for smart grid design, architecture of the smart grid, advantages and challenges of smart grid, current research focus in smart grid, security and privacy issues in smart grid communications and related works on privacy-preserving smart grid. Then, we employ Paillier Homomorphic Encryption to propose a continuous privacy-preserving histogram query scheme

for secure smart grid communications, which can generate a histogram for a user-specified time period while preserving the privacy of residential users. The proposed scheme presents residential users' electricity usage data to the control center without violating their privacy. It does this by presenting all the users' electricity data into two forms of histogram data. The first form is the sum of each class of data, which sums up all the electricity usage data within a particular range and presents it to the control center. The second is the count of each class of data, which counts all the electricity usage data that has been added within a particular range and presents to the control center. Our scheme contains three phases, i.e., Report Generation, Report Aggregation and Array Recovery phase. We analyze the security of each phase and evaluate its performance. The results show that our scheme is privacy-preserving and efficient. Especially, the average time consumption for each of the phases is less than 20 ms in our evaluation.

Dedication

First I would like to dedicate this thesis to Almighty God for his mercies in preserving my health throughout the program. I would also like to dedicate this thesis to my parents, Rev. Benjamin Larbi and Mrs. Florence Larbi, for their unconditional love and support, without whom I wouldn't have been able to complete this chapter of my life.

Acknowledgements

I would like to express my sincerest gratitude to my supervisor, Dr. Rongxing Lu for his patience, guidance, encouragement and support throughout my time as his student. I was very fortunate to have a supervisor who cared so much about my work. I would also like to thank all the members of staff of the University of New Brunswick who have been there to help me with every aspect of academics and student life. I would like to acknowledge my parents, Rev. Benjamin Larbi and Mrs. Florence Larbi, for their continued love and support. I would also like to thank Christian Akpanya and Ebenezer Usher, whom I see as big brothers, for their care and support throughout the difficult times. I would also like to acknowledge Yunguo Guan, Yandong Zheng, Songnian Zhang and Kwasi Boakye Boateng for their guidance. I would also give thanks and praise to my Lord and Saviour, Jesus Christ for his love and mercies upon my life.

Table of Contents

List of Tables

List of Figures

Abbreviations

AMI Advanced Metering Infrastructure

AODV Ad-hoc On-demand Distance Vector

BAN Building Area Network

CC Control Center

CPS Cyber Physical System

DER Distributed Energy Resource

DMS Distribution Management System

DR Demand Response

EMS Energy Management System

EM Electromobility

EPON Ethernet Passive Optical Networks

EU European Union

FAN File Area Network

GDPR General Data Protection Regulation Area

GW Gateway

HAN Home Area Network

IAN Internet Area Network

ICT Information Communications Technology

IEC International Electrotechnical Commission

IED Intelligent Electronic Device

LAN Local Area Network

MDMS Meter Data Management System

MPLS Multi-Protocol Label Switching

NAN Neighbourhood Area Network

NIST National Institute of Standards and Technology

PIPEDA Personal Information Protection and Electronic Documents Act

PLC Programmable Logic Controller

PMU Phasor Measurement Unit

RTU Remote Terminal Unit

SCADA Supervisory Control and Data Acquisition

WAMS Wide Area Measurement System

Chapter 1

Introduction

The smart grid or smart power grid as a cyber physical system (CPS) is a tremendous improvement over the traditional power grid and can be considered as the next revolutionary innovation in electric power generation, transmission and distribution. The traditional power grid's focal point is in the generation, transmission, distribution and control of the flow of electricity and it has the features of one way communication, manual control and recovery, very few sensors, minimal extent of control and less consumer participation. The smart grid improves upon almost everything associated with the traditional power grid and is intelligent. It has the features of two way communication, more sensors on the network, remote checks, an increase in customer participation and ability to recover from cyber-attacks.

The smart grid can assist in the creation of a better and more active elec-

tricity market as the participation of consumers grow. The grid's ability to deliver a steady source of power supply makes it capable of handling difficult demand and response cycles. It has the capability of automatically adapting to avert and manage blackouts and outages. In addition, the smart grid can make support available for different new applications and it provides easy integration for renewable energy sources encompassing different kinds of distributed power generators and electric vehicles.

1.1 Background of smart grid

The traditional power grid or the conventional electrical grid has not seen any significant changes in over 10 decades causing the grid to be antiquated. A grid such as this is environmentally wasteful, inadequate and a consumer of fossil fuels that emit particulates and greenhouse gases and affect the ecosystem. Meanwhile, the traditional grid may not be able to meet the ever increasing demand for power in the future. Hence, utilities are forecasting that the increasing requirements for electrical grid sustainability, efficiency and resiliency will make the shift to a smart grid paradigm inevitable. A lot of countries around the world have implemented smart grid technologies in their electricity networks as shown in Figure 1.1.

Smart grid does not have a single unique definition and it is defined differently by a number of organizations. We list these definitions as follows [23].

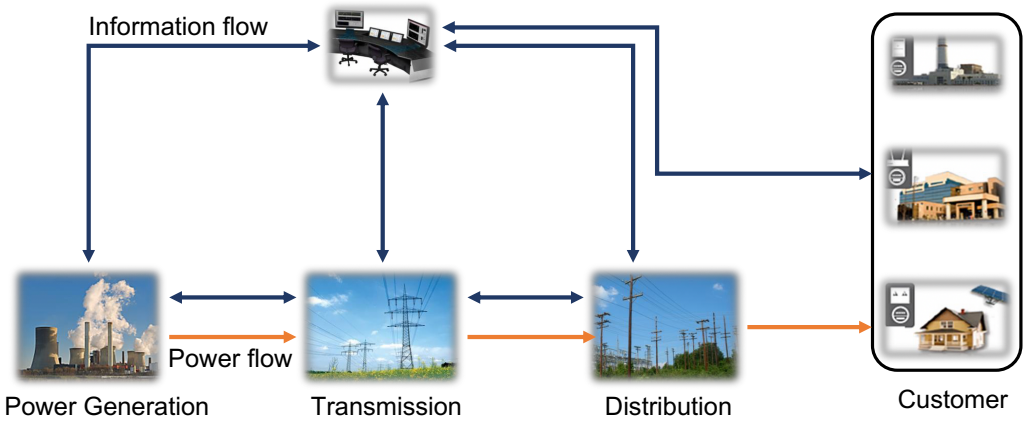

Figure 1.1: The envision of smart grid

• The European technology platform [35] [41] [29] defines that *"A smart grid is an electricity network that can intelligently integrate the actions of all users connected to it-generators, consumers and those that do both in order to efficiently deliver sustainable, economic and secure electricity supplies."*

• Smarter grids [2] defines that *"A smart grid uses sensing, embedded processing and digital communications to enable the electricity grid to be observable (able to be measured and visualized), controllable (able to be manipulated and optimized), automated (able to adapt and self-heal), fully integrated (fully interoperable with existing systems and with the capacity to incorporate a diverse set of energy sources)."*

• The Canadian Electricity Association [7] defines that *"The smart grid is a suite of information-based applications made possible by increased automation of the electricity grid, as well as the underlying automation itself; this suite of technologies integrates the behavior and actions of all connected supplies*

and loads through dispersed communication capabilities to deliver sustainable, economic and secure power supplies."

• The U.S. department of energy defines that *"A smart grid uses digital technology to improve reliability, security and efficiency (both economic and energy) of the electrical system from large generation, through the delivery systems to electricity consumers and a growing number of distributed-generation and storage resources."*

• The IEC defines that *"The smart grid is a developing network of transmission lines, equipment, controls and new technologies working together to respond immediately to our 21st Century demand for electricity."*

• The IEEE defines that *"The smart grid is a revolutionary undertaking-entailing new communications-and control capabilities, energy sources, generation models and adherence to cross jurisdictional regulatory structures."*

The smart grid is a cyber physical system, which comprises of a computing system, a communication channel network and physical elements as shown in Figure 1.2. Controllers takes charge of the physical dynamics of smart grid equipments; sensors are placed and connected through the communication network and they aids in measuring the overall condition of the smart grid and sends relevant data back to the controllers. Specifically, a cyber physical system comprises of the following components [37].

Physical dynamics: Every physical aspect of a cyber physical system are part of this component. This can be the voltage of distribution lines of the

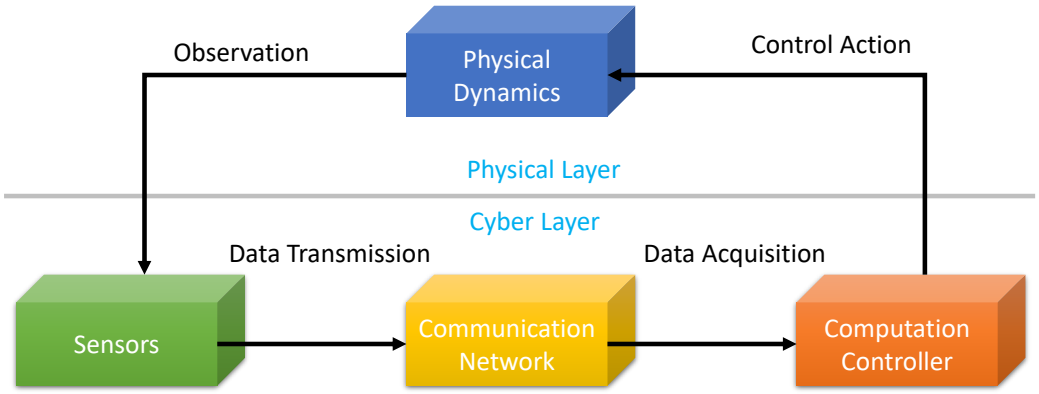

Figure 1.2: A Cyber Physical System

electrical power grid, the weight of a vehicle on the highway or the velocity and position of robots. The physical dynamics change with time and with the law of evolution which can be computed through the system itself and by the control actions of system.

Sensors: To be able to have an awareness of the physical dynamics, sensors are positioned that can be used directly or indirectly; a lot of sensors can exist in an individual CPS [51] [69] [88] [57]. Smart grid has many sensors that work together to sample dynamics of electricity and compute phases of different frequencies among others.

Controller: The sensors deliver reports to their controllers and the controllers act accordingly. Location of computations and actuation actions may be different hence a communication channel is needed. In smart grid, valve position and actuation are controlled and measured by controllers(frequency controllers), depending on frequency reports they receive from various sen-

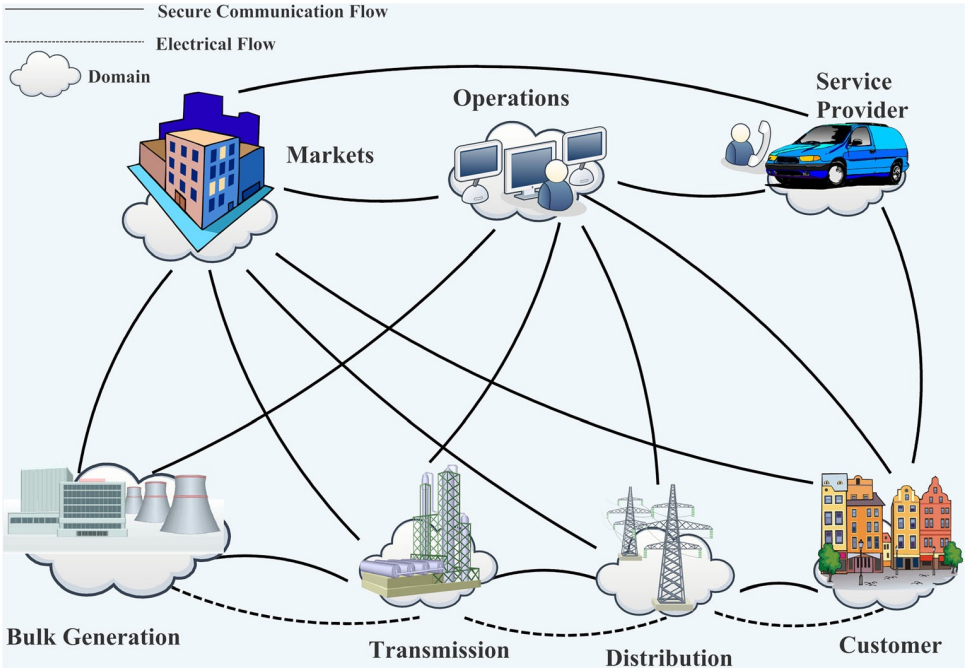

Figure 1.3: NIST Reference Model defined for the smart grid

sors.

Communication network: In smart grid, the location of sensors and controllers is different in general and a communication network is needed for the interaction between these two important components. The network could be wired or wireless depending on the needs of the specific power grid.

1.1.1 Architecture of the smart grid

A number of models have been developed for the smart grid architecture but most people and organizations follow the reference model proposed by

the National Institute of Standards and Technology (NIST). The model proposed by NIST comprises of seven different logical domains [31]. The initial four logical domains are bulk generation, distribution, customers and transmission which conserve, supply and produce electricity in a two way flow of information and electricity (Figure 1.3).

The remaining three are operations, markets and service providers that control the flow of electricity, essential information, utility services and consumers. There exist a number of customers in the defined model: Industrial Area Network (IAN), Building Area Network (BAN) and Home Area Network (HAN). In these locations, Advanced Metering Infrastructure (AMI) is employed to observe and record communication and the inward or outward flow of electricity.

Based on the architecture of the smart grid, the smart grid contains the following components.

1. **Advanced Metering Infrastructure (AMI):** It is the combination of a number of technologies, concerning smart connections amidst system operators and customers [67] [62] [20] [38] [33]. The major applications that are present on the Home Area Network (HAN) are operational gateways, meter data management systems (MDMSs) and smart meters [8] [22]. The AMI infrastructure concentrates on customers and acquiring real-time pricing of electricity, which enables them to better manage their electricity consumption.

2. **Supervisory control and data acquisition:** Its sole purpose is for the real-time control and monitoring of power delivery network [39] [12] [84] [21]. Two subsystems which are: energy management system (EMS) and distribution management system (DMS) are a crucial part of the supervisory control and data acquisition (SCADA) system [66] [17] [81] [14].

3. **Communication standards and protocols:** The benchmark of communication for the electricity industry are developed by top organizations in communications like IEEE, Distributed Network Protocol (DNP3) and International Electrotechnical Commission (IEC) [19] [27] [59].

1.1.2 Characteristics for smart grid Design

In 2007, the U.S. National Energy Technology Laboratory came up with seven characteristics for designing smart grids. Later in 2009, the U.S. Department of Energy (DOE) combined two of them (self heals and resist attacks). We list these characteristics as follows [56].

1. **Enabling participation from the customer:** In the smart grid system, customers are more enlightened and communication is two way unlike the traditional power grid. Customers have the option of adjusting load on their network to meet their needs for better electricity

management. The response is directly related to the demand and also real-time pricing is an addition to the smart grid system.

2. **Accommodating storage and generation:** The smart grid system adopts various DERs like solar, geothermal energy, tidal and wind with the assistance of a flexible network. This network assists in controlling the loads and backing-up energy requirements that aid in boosting the productivity of smart grid.

3. **New products, markets and consumer services:** With smart applications and intelligent electronic devices (IEDs) being made available to customers, customers can remotely adjust their power consumption by choosing various services that are more appropriate for them. Various grid parameters like service quality, capacity, time and rate will influence the market and customers will have authority over them.

4. **Need-based power quality for customers:** One customer's needs may vary from another, for instance, in a home where the requirement for power is much less as compared to a large industry. Also the demand in quality of power (such as fluctuation on voltage, interruption, voltage volume among others) varies. Meeting these requirements is a tough task for smart grids.

5. **Optimized utilization of different operations:** Essential functions of the smart grid include: productive operations and administration of

various facilities, DERs and grid applications. An enhanced use of all these resources lowers the investment cost and power consumption. Hence a tough and effective administration is required for the smart grid system.

6. **Resiliency against cyber-attacks and disasters:** It is important that the smart grid is reliable. Failure of the smart grid could result in devastating effects for the nation in question. Hence smart grids must be able to recover from harmful events like cyber-attacks, physical attacks and natural disasters. Self-healing ability is a feature that is required of the smart grid with the presence of technologies such as: automatic control devices, advance sensing, accurate and timely detection among others.

1.2 Advantages and Challenges of smart grid

The smart grid is an improvement over the traditional power grid. There are many advantages and challenges that plague this new technology [75]. We take a look at them in this section.

1.2.1 Advantages

The smart grid has the advantages of self-healing, interactive, security, asset optimization, distributed generation, market empowerment, environment

friendly as follows.

1.2.1.1 Self-Healing

With the aid of real time sensors like Phasor Measurement Units (PMUs) the smart grid can sense the physical state of the power. With its self-healing attribute the smart grid can predict, identify and respond to defects and blackouts [5] [3] [9]. It does this by using PMU and automatic control center. It also takes advantage of intelligent sensors that have the ability to begin, end and change the flow of electricity to guard against additional issues [90] [54].

1.2.1.2 Interactive

One of the essential properties of smart grid is its two way power flow. At a particular instance one can be both a consumer and a provider [24]. Dynamic pricing motivates users to use minimal power during peak hours, which ultimately minimizes the peak demand. This conduct eventually progresses towards the unity Load Factor which is preferable.

1.2.1.3 Security

The smart grid network and control system is created in such a way that it is capable of resisting cyber-attacks [46] [64]. Real time monitoring with PMU

makes it possible for the operators to predict future issues that may affect the Grid. This enables operators to take necessary steps to guard against such issues or problems. Its distributed generation and microgrid feature guarantee the security of the supply.

1.2.1.4 Asset Optimization

With the use of Geographic Information System (GIS), we can create competent network using minimum transmission network and other tools [6]. It will allow condition and performance based maintenance. Smart grids will enhance efficiency by lowering technical and non-technical line losses.

1.2.1.5 Distributed Generation

Distributed Generation (DG) implies dispersed generating units taking the place of a centralized network. The advantages of distributed generation can be split into two sections: economic and operational. From an economic standpoint, distributed generation delivers power support when there is peak demand, thereby reducing interruption that may cause blackouts. It also lowers the cost of investment because of the flexibility of its capacity and installation placement. DG lowers operational costs when stationed near to the customer load because it prevents improving or creating a new transmission and distribution network hence saving cost. Using local renewable

energy sources (RES) will assist in lowering the reliance on imported fossil fuels and reduce increasing energy prices [87].

1.2.1.6 Market Empowerment

The smart grid includes customers by making them a part of the electricity market as active participants. It will give utilities the power to keep up with the ever changing consumer expectation and provide greater clarity and option in energy purchasing [74]. It will create the need for cost-effective and energy-efficient products. Smart grids will aid in schooling the average consumer, promoting change in new energy management services and lower the cost and environmental impact of providing electricity.

1.2.1.7 Environment Friendly

The energy preservation and advances in end-use efficiency provided by the smart grid can help lower significantly the proportion of CO2 emissions [72]. The Pacific Northwest National Laboratory (PNNL) came out with a report in which carbon savings from the introduction of smart grid technologies are estimated to the year 2030. The PNNL estimates direct carbon savings from equipment such as smart meters at 12 percent and indirect savings from items like stronger grid support for renewable electricity generation at 6 percent. Environmental improvements can be observed by controlling the peak load through demand response rather than spinning reserves. Smart

grid will lower transmission and distribution loss whiles controlling theft, making sure there is better accessibility in the hinterlands and increasing reliability and quality of supply in urban areas. Minimizing these losses would mean an investment of about $20,000 (reactive power compensation, phase shifting transformer) to $75,000 (power electronics steering and control) per MW. Smart grid system allows for continuous feedback on electricity use, which allows consumers to control their usage in reaction to pricing and consumption hence lowering annual CO2 emissions. Optimized use of existing generation, transmission and distribution through this system reduces the new infrastructure constructions.

1.2.2 Challenges

The smart grid has the following challenges: financial resources, government support, compatible equipment, consumer education, cost assessment, cyber security and data privacy, capacity to absorb advanced technology and strengthening the grid as follows.

1.2.2.1 Financial Resources

Large capital investments are required to implement the smart grid in existing electrical grids. In effect notable financial investments are needed to generate the necessary distributed network and other establishments [68].

1.2.2.2 Government Support

Huge financial investments are not the only hurdles one must overcome in order to implement smart grid. For an efficient implementation of smart grid one must have a willing government and an effective energy policy [53] [61].

1.2.2.3 Compatible Equipment

The smart grid is a technologically advanced system. A lot of equipment are required which must be compatible with the system. Old equipment must be removed from the existing electrical grid since they are incompatible with smart grid technologies. This may create an issue for utilities, regulators and customers.

1.2.2.4 Consumer Education

Consumer education and participation is a critical aspect of the successful implementation of the smart grid. A large portion of the benefits of smart meter depend on consumer participation. Therefore the consumers have to be informed and smart to get the maximum benefit.

1.2.2.5 Cost Assessment

Costs could be more costly than estimated. This is due to the fact that the standards and protocols required to create and run an advanced metering in-

frastructure are always changing. Therefore making investments now before the standards are settled could risk having an obsolete infrastructure [65].

1.2.2.6 Cyber Security and Data Privacy

The use of digital communication networks and the exchange of information on consumption patterns raises concerns in cyber-security and the likelihood of misuse of private data [48] [71] [89].

1.2.2.7 Capacity to Absorb Advanced Technology

Smart grid makes use of sophisticated devices. Technology is always moving forward and never at a standstill. The smart grid must have the ability to absorb current and advanced technology.

1.2.2.8 Strengthening the Grid

The smart grid though a modernization of the electrical grid faces some challenges which should be our area of concern. The organizations and the manpower that setup the smart grid face some unexpected occurrences and hazards which the smart grid must overcome. Some of these hazards are: Attacks of cyber thieves, Weak base, Inefficient Control System, Corrosion, Smart Meter Authentication and blackouts among others.

1.3 Current Research Focus in smart grid

Driven by the advantages and challenges in the smart grid, the research on smart grid has received considerable attention. In this section we present the various categories in smart grid research. They can be categorized as distribution automation (DA), grid management (GM), energy storage (ES), market (MA), generation and Distributed Energy Resources (Gen & DER), electromobility (EM), smart homes/buildings (SH), smart cities(SC), demand response (DR), information and communication technologies (ICT), cybersecurity (CS) and advanced metering infrastructure (AMI) [40].

1.3.1 Distribution Automation (DA)

This category is centered around the automatization of the grid, implying control and monitoring systems that boost the system's dependability and power standard. Digital sensors, switches and intelligent electronic devices (IED) are used for this activity. Actions like voltage monitoring, reactive power management, volt-Var control and fault detection are carried out in conjunction with distributed automation (DA) and grid management (GM), which helps to enhance the network strength and lowers cost [13].

1.3.2 Grid Management (GM)

Grid management improves the operation of transformers and other components to enable the incorporation with transmission systems. The major topics covered are: substation automation, automation of distributed networks, self-healing networks, real-time monitoring of the grid, advanced control systems and microgrids. The research goals in this area are: reliability, integration of distributed generation, efficiency, voltage control and reactive power [78].

1.3.3 Energy Storage (ES)

Energy storage comprises of the action of transforming electrical energy into a form that can be conserved and supplying this energy in the form of electricity back into the grid when needed. Various technologies can be used for this objective, whereas the electrical energy can be supplied to the transmission or distribution grid or at the customer's premises. Storage is very essential for the evolution of smart grids, since it allows for the higher integration of renewable energy generation, which is associated with irregularity. Storage can lead to a more effective energy management and can lower the energy waste. Examples of energy storage system types are: batteries, pumped-storage hydroelectricity, underground thermal ES, compressed air ES, thermochemical, chemical-hydrogen storage, flywheels, supercapacitors and superconducting magnetic ES [76].

1.3.4 Market (MA)

Market activities are associated with the influence of smart grids on the electricity market and they can also involve the requirements of creating a product for the market. Technological market barriers in smart grids are also looked at as well as modeling of new financial frameworks and market relations between actors in smart grids [74].

1.3.5 Generation and Distributed Energy Resources (Gen & DER)

Distributed energy resources (DER) can either be sources of generation or storage that are linked to the distribution system. Problems like non radial power flow and increased fault current duty might occur and they need to be solved. The generation can be split into two categories: classic generation (dispatchable) and generation that is not controlled by the system operator. Wind energy and solar energy created by photovoltaics (PVs) are produced by DER. Energy from DER is crucial for smart grids so as to lower CO2 emissions and generate energy in an eco-friendly way. Other examples of DER are: biomass, hydro fuel cell, tidal waves or gas power plants [83].

1.3.6 Electromobility (EM)

The concept of EM is not restricted to electric vehicles and their usage but includes their interaction with the electric grid of which communication infrastructure is key. The topic of EM is large and there exist several aspects like: energy efficiency, power quality, energy management and vehicle autonomy, ES, interoperability, citizen behavior, security and safety, grid load impact and environmental impact. Vehicle charging is a very essential element and special attention is being given on areas like charging technologies, charging plugs, charging power options and charging infrastructure [25] [43].

1.3.7 Smart Homes/Building (SH)

Smart homes/building means that present-day telecommunication and automatized systems are used within homes to observe and manage smart devices, whereas communication with smart meters is made possible. Because of the presence of these advanced systems, the consumer is better informed of their usage and can become an active player in the successful implementation of smart grids. Motivation to shift or limit loads are provided, which can be based, for example, on different tariffs. Sensor or submeter networks are also a part of this category especially within smart buildings, which enable an efficient energy management. Activities under this category include: temperature control, lighting, sensors technology, power quality, smart appliances, DR and energy management strategies [36].

1.3.8 Smart Cities (SC)

Smart cities have been receiving some attention from the scientific community in recent times in the form of several initiatives and leading proposals, like EU SC partnership marketplace and the SC research cluster. The SC concept comprises of advanced services to citizens like transportation, lighting, internet access or other technologies that enhance the standard of everyday life.

1.3.9 Demand Response (DR)

In a wider sense, DR includes making all citizens involved in participating more in better energy management. DR means load shifting or curtailing in order to avoid highs in the overall usage curves. This would create a more solid and well-grounded grid. DR comprises of tariff incentives to customers and it implies deliberately modifying consumers' usual pattern of electricity usage. It is a topic that is currently under evaluation with 346 competitive projects in Europe and topics like baselining, product design, measurement and validation activities and market models are being looked at. Activities involved in DR are: SH and smart building, demand modeling, DR management systems and customer energy management systems [74].

1.3.10 Information and Communication Technologies (ICT)

Information and communication technologies solutions comprises of all the hardware, software, telecommunication equipment and systems that allows information transmission among the different actors of a smart grid. This topic consists of a broad field of technologies and applications. There are different types of networks (eg, WAN, FAN, NAN, LAN and HAN) on which smart grid information and communication technologies can be applied. In addition there are various wireless technologies that can be applied, for example: GSM, GPRS, 3G, LTE, Zigbee, WiFi and 6LowPAN among others. There also exists a number of wired solutions that can be applied, for example: the power line communications (NB-PLC and BB-PLC) and optic fibers among others. Activities in this category are: remote equipment configuration, system status monitoring, event management and response automation among others [77].

1.3.11 Cybersecurity

Cybersecurity is necessary for smart grids to exist. The large data transmission means that advanced techniques should be applied in order to safeguard critical information and confidential data. It also comprises of all the measures for safeguarding communication devices against unapproved access or

actions that could give rise to alteration or theft of data. Activities associated with CS are risk assessment, risk response, confidentiality and privacy, authorization and authentication among others [56].

1.3.12 Advanced metering infrastructure (AMI)

Smart meters are important devices in smart grids. This is because they provide two-way communication between energy providers and customers and they can also help in load consumption control and monitoring. By 2020 around 200 million smart meters are likely to be installed in Europe. There are various technologies that can be used for setting up smart meters including wireless and wired solutions. Some of the applications in this category are: Installation and configuration, pricing and billing and communication issues among others [67].

1.4 Security and Privacy Issues in smart grid Communications

In this section, we take a look at the various cyber security and privacy issues affecting the smart grid in detail [56].

1.4.1 Cyber security Issues

The cyber security of the smart grid comprises of all IT and communication issues that bother the operation of power delivery systems and the management of the utilities as follows.

1.4.1.1 Device Issues

Devices like PLC's (Programmable Logical Controllers), RTU's (Remote Terminal Units) and IED's (Intelligent Electronic Devices) are widely used in power delivery systems and they give administrators the ability to conduct maintenance or to carry out functionalities from a remote location. This ability allows unauthorized users to control the device and disrupt normal operations of the grid such as shutting down running devices to disconnect power services or manipulating sensing data to misguide the decisions of the operators. For meter devices, a conventional physical meter can be altered by reversing the internal usage counter or be tweaked to control the calculation of electric flow. A number of algorithms have been introduced in [52], [10] to prevent the meter data from being compromised.

1.4.1.2 Networking Issues

Ethernet Passive Optical Networks (EPON) has been touted as a promising solution for the smart grid in the area of wired networks. However EPON

can be easily attacked by methods such as spoofing, DoS and eavesdropping. By making use of identity-based cryptography (IBC) and challenge-response technology, a secure communication protocol was proposed for the EPON. In sensor networks, researchers have come up with a unanimous decision that wireless mesh networks should be used in the AMI. One of the main reasons is that mesh networks can overcome bad links by using redundant communication paths. However wireless mesh technologies suffer from a number of attacks such as: cross-layer traffic injection, node impersonation, route injection, message modification. To solve this problem Zigbee Alliance released a standard based on Zigbee Pro and 802.15.4 standards. Bennet and Wicker [11], however, argued that the conventional Zigbee protocol would experience severe delays due to the multi-tier feature of the cluster-tree based routing strategy. To speed up the transfer rate, the authors came up with an idea of adding a new layer between layers 2 and 3 of Zigbee networks. This layer will use a modified multi-protocol label switching (MPLS) layer 2.5 protocol to reduce end to end delays. They also suggested that the routing protocol in Zigbee networks must be pure AODV (Ad Hoc on Demand Distance Vector), which could significantly shorten the time needed to establish a path. The authors also discovered that the AODV protocol suffers from black hole attacks that discard path establishing messages. To solve this problem, they came up with the solution of establishing a dedicated path between two communication principals.

1.4.1.3 Dispatching and Management Issues

The smart grid has the following dispatching and management issues.

• Take down the server: If information like the IP of the SCADA server and the network path are acquired by the attacker, the server can easily be brought down or turned off by the traditional denial of service errors or by just deleting the system files. Denial of service can be performed if the TCP/IP can be overwhelmed. Deleting the files can be accomplished by hacking the user passwords or gaining access to the physical system. These attacks pose a threat to future services also.

• Gaining control over the system: This is done by deploying a Trojan or by backdoor entry into the system registries. This is the highest form of security threat where a false alarm and manipulated controls can be created and sent to RTUs leading to large scale collapses.

• Stealing corporate data: These issues happen when the enterprise security level is weak and the software architecture used is not highly capable. The corporate data can be stolen from the database and used to the advantage of competitors that get hold of the data.

• Fiddling with billing information: Intruders might be able to gain unauthorized entry into the system and access billing and other financial information. This information can later be misused and can create a lot of issues for the consumers. There has to be a strong firewall in place to protect the servers

from losing such important information.

• Key logger software: Attackers are capable of using the logged key strokes of the system keyboard and gaining access to the system passwords and usernames.

• Gain competitive advantage: Attackers from a service provider can gain unauthorized entry into the system of another service provider. Access their data and get to know their plans and strategies and thus change their own plans in a way that would benefit them in a competitive environment.

• Using the SCADA servers to attack the other servers in the system thereby gaining access to valuable information from the utility companies.

• Intentionally alter mathematical data points to mislead the utility operators who then identify a false alarm and then begin to shut down or rescale the system causing unnecessary delays.

• Changing user logged data in a distant and remote DBMS, which can bother the innocent users as well as the utility companies.

1.4.1.4 Anomaly Detection Issues

Unsecure timing information can be misused for performing replay attacks, which has a huge impact on security protocols. Meanwhile, Faulty RTUs can affect the network negatively in so many ways. Accuracy of data transfer and service quality guarantee are not possible.

1.4.1.5 Other Issues

All modern data communication protocols follow a messaging protocol that is properly documented and available to the public. An example of a documented protocol that is available to the public is the DNP protocol. The DNP protocol is used by a lot of electric utilities across North America. Using these documented protocols makes it possible for an intruder to perform reverse engineering of the data acquisition protocol and exploit the protocol using a "Man-in-the-middle" attack. The resulting effect could include sending incorrect data to the field device or control center operator leading to 1) financial loss if the attack results in excess generation output; 2) physical danger if a line is carrying power while linemen are in the field servicing the line; 3) equipment damage if control commands sent to the field lead to overload conditions.

IEC 62351 is a support standard for IEC 61850 that concentrates on security and technical requirements of vendors. Fries et al. [32] gave a summary of both documents and stated that IEC 62351 must be updated because of some new use cases in the smart grid. Those use cases are mainly derived from customer participation and demand response in the grid. From the results of IEC 62351, the authors asserted that the MMS and XML should be further upgraded to guarantee the integrity of the application layer. When a central command is forwarded by an intermediate substation, the current MMS version is not capable of guaranteeing its integrity in the application layer. To

solve this problem, the authors came up with a possible solution which is by adding a "cryptotoken" to the command packet. First, it establishes a TLS connection on every hop with corresponding session keys on the transport level. Secondly, it establishes an end-to-end communication channel on the application level and negotiates the session key during the handshake phase. Thirdly, it uses this session key to secure all successive traffic. Through these steps integrity in the application layer is realized.

1.4.2 Privacy Issues

In this section we take a look at the various privacy issues that affect the smart grid. In accordance with the study of NIST [34], four areas of privacy concern in the smart grid should be considered.

• First and foremost, fraud should be noted, in the case where energy consumption is ascribed to a different location (e.g., in PHEV's case). The metering system (either physical recording or electronically and remotely metering systems) must not tolerate personnel abuse or alter the collected data. The NIST's report [34] has taken into consideration two relevant privacy use cases and has analyzed them into details. One case is concerning a landlord with tenants that own PHEVs and prefer to be charged separately. In order to preserve the privacy of the tenants, utility is included to validate communications between the smart meter and PHEVs through a secure line and energy services communication interface (ESCI) supplied by the util-

ity and/or vehicle manufacturer. Another case is concerning PHEV general registration and enrollment process. To be able to complete initial setup for PHEVs, NIST maintains that utilities should provide these services to customers: 1) enrollment, 2) registration, 3) initial connection, 4) ability to continuously re-establish connection between a utility and PHEV, 5) capability of supplying a PHEV tariff or charging status information to customer interfaces, and 6) correct bill.

• Second, data within the smart meter and HAN could disclose certain actions of home smart appliances. It can also be used for monitoring specific times and locations of energy consumption in certain parts of the home, which may further show the appliances used and/or types of activities. For example, appliance vendors may require this kind of data to know both how and why persons used their product in a certain manner. Such information could affect appliance warranties. Also other entities may require this data to perform target marketing. Georgios et al. [44] created a system that makes use of a power router and a rechargeable battery to conceal load signatures in a home area. In this system there is the assumption that the home will contain a lot of energy storage and generation devices in future. This model has been further enhanced and it's called ElecPrivacy.

• Third, acquiring near-real-time data regarding energy consumption may help in knowing whether a residence or facility is occupied, location of people in the building, what they are up to and so on. Authors in [52] presented a

data aggregation approach for all level meters based on spanning tree topology. Making use of homomorphic encryption secures the data from home meters to the data center. In [85], researchers noted that customers may use a separate measurement device at home to effectively observe their power usage. The redundant meter data, if transmitted in an unsecured wireless environment can disclose customer's information to a malicious intruder. By compressing the data to a rate below its entropy, the authors introduced a coding method that solved this problem.

• Finally, personal lifestyle information gotten from energy consumption data may be beneficial to some vendors or parties. For instance vendors may make use of this information for the purposes of targeted marketing, which may not be tolerated by those targets. This useful information may be disclosed by new technologies like smart meters, time of use and demand rates, and direct load control of equipment. They can be further sold and used for energy management analysis and peer comparisons. Costas et al. [28] introduced an escrow-based anonymization scheme that makes it impossible for personal information to be tracked by unauthorized third parties. They sorted metering data into two parts: "high-frequency" and "low-frequency". Then corresponding setup and communication procedures were created for each type of data. These procedures are both regular PKI authentication approaches. The anonymity degree of the system depends on the size of the anonymity set hence to widely deploy a scheme like this requires further investigation. In addition two aspects of the smart grid data need to be

looked at when evaluating existing laws and regulatory policies to guarantee that new types of data are catered for: 1) granular and available data on use of individual appliances by time and location; 2) public awareness of contractual agreements about data ownership and what may be disclosed concerning people's daily activities.

1.5 Related Works on Privacy-Preserving Communication Protocols in smart grid

In the smart grid, a number of meters and sensors are widely deployed across the power system and they collect and report a huge amount of real-time data to the system control center. With the received data, the control center can automatically and timely monitor grid status, balance electricity load, maintain system operation, optimize energy consumption, etc. Specifically, all the intelligent electric appliances in the residential user's home are connected to a smart meter, which periodically records the power consumption of appliances and reports the metering data to a local area substation. Since the sensors collect and report the data at a high frequency, say 50 Hz, how to efficiently achieve data reporting becomes challenging. To address this challenge, the secure aggregation technique was proposed. In the secure aggregation technique, the customer's privacy should be preserved while transmission. Existing data aggregation schemes also stress the same consideration during

aggregating, individual user's data should not be exposed. Most of them use homomorphic encryption to encrypt users' data so that the semi-trusted aggregator can aggregate all users' data without decryption. However, all of these schemes can only be used to compute the summation of users' data as the aggregation, while the control center may need to compute more statistics such as the mostly common used variance. Therefore, how to compute more complex statistics of users' data without disclosing individual user's data is a challenging problem in smart grid communications. Actually, this problem is also mentioned as one of the open research challenges in Internet of Things.

To better understand the research challenge of data aggregation in smart grid communications, we first review some existing research efforts made on data aggregation. Data aggregation is a crucial technique in smart grid communications, which is used for collecting the most critical data in an energy efficient manner with minimum data latency. In past decade, many data aggregation schemes have been proposed, including flat network based data aggregation [49,50] and hierarchical network based data aggregation [16, 55]. However, all these schemes do not consider the security challenges, which results in the possible disclosure of sensitive data, e.g., customer privacy, and the pollution of aggregated data due to malicious devices involved. To address the security issues, Przydatek et al. [73] propose a framework for secure data aggregation, and present an approach called aggregate-commit-prove to enable the data receiver to verify the authenticity of the information

provided by the aggregator with efficient random sampling mechanisms and interactive proofs. Although Przydatek et al.'s framework enables secure data aggregation; the effectiveness of the approach still needs to be verified by extensive simulations and real experiments. Meanwhile, data privacy is also not well protected in the framework.

To address the privacy issue in data aggregation, Shi et al. [79] propose the privacy-preserving aggregation of time-series data, which allows a group of nodes to periodically upload encrypted values to a data aggregator, and the latter can calculate the sum of all users' values in every time period, but cannot learn anything else. In [42], Joye and Libert observe that Shi et al.'s protocol just supports small plaintext spaces, and thus present a practical privacy-preserving aggregation protocol to accommodate large plaintext spaces. To support both spatial and temporal aggregation, EKin and Tsudik [30] explore some simple yet relatively efficient privacy techniques for the aggregation of smart meter measurements. Although the above three protocols [30, 42, 79] achieve privacy-preserving property, they are not fault-tolerant, that is, once a node is malfunctioning, the whole aggregated results cannot be correctly decrypted. Therefore, fault-tolerant and privacy-preserving aggregation is still challenging. To address this challenge, Chan et al. [15] propose a privacy-preserving stream aggregation with fault tolerance. Though it is resilient to user failure and compromise, the protocol is not quite efficient in terms of transmission cost. In addition to the above privacy-preserving data aggregation works, privacy-preserving data aggregation [80]

is also applied to people-centric urban situation to achieve user privacy, yet the authentication is not well considered. Finally, we should take a note that, homomorphic encryption is an important technique to implement most privacy-preserving aggregation protocols [30], however, since homomorphic encryption is a time-consuming public key technique, it is not efficient for designing privacy-preserving data aggregation, especially for low-cost nodes in smart grid communications.

Although a range of schemes targeting different kinds of data aggregation have been proposed, the continuous histogram aggregation/query is out of reach for the current research. Aiming at the issue, in this thesis, we propose a continuous privacy-preserving histogram query scheme for secure smart grid communications.

1.6 Summary of Our Contributions

Our major objective is to propose a continuous privacy-preserving histogram query scheme for secure smart grid communications. In this thesis, the following contributions have been made:

1. We give a brief description of the smart grid with areas like: characteristics for smart grid design, architecture of the smart grid, advantages and challenges of smart grid, current research focus in smart grid, security and privacy issues in smart grid communications and related works

on privacy-preserving smart grid.

2. We establish our system model, security model and our design goals on continuous privacy-preserving histogram query in smart grid communications.

3. We present our continuous privacy-preserving histogram query scheme for secure smart grid communications, which generates a histogram for a user-specified time period while preserving the privacy of residential users.

4. Finally we present an evaluation on the performance of our proposed continuous privacy-preserving histogram query scheme.

1.7 Thesis Organization

This thesis is organized as follows: In Chapter 2, we discuss the following topics into details: Symmetric Encryption, Asymmetric Encryption, Homomorphic Encryption, Paillier Homomorphic Encryption, Hash Function and Digital Signature. In Chapter 3, we formalize the system model, security model and identify our design goals. In Chapter 4, we present our proposed continuous privacy-preserving histogram query scheme in details. In Chapter 5, we discuss our security analysis and performance evaluation of the continuous privacy-preserving histogram query scheme. In Chapter 6, we discuss

the conclusion that has been drawn from this work and our potential future work.

Chapter 2

Preliminaries

In this chapter, we talk into details about the following topics: Symmetric Encryption, Asymmetric Encryption (Public-Key Encryption), Homomorphic Encryption, Hash function and Digital Signature.

2.1 Preliminaries to Cryptography

Cryptography involves transforming the intended message into an unreadable form and transmitting the message through an insecure channel. There are two main forms of encryption, namely: Symmetric Encryption and Asymmetric Encryption.

2.1.1 Basic Terms Used in Cryptography

- **Plaintext:** The actual message that an individual wants to send to another person is known as Plaintext. In Cryptography, the raw message that has to be sent to the other party is given a special name known as plaintext. For example, Alice as a person wishes to send "Hi friend, its been long" message to Bob. Here "Hi friend, its been long" is a plaintext message.

- **Ciphertext:** The message that does not make sense or cannot be comprehended by anybody is what is known as Ciphertext. In Cryptography, the intended message is transformed into an unreadable format before the message is sent to the intended recipient. For example, "Bkl8&jqM1234@" is a ciphertext produced.

- **Encryption:** The process of transforming plaintext into ciphertext is known as Encryption. Cryptography makes use of the encryption technique to transmit confidential messages through an insecure channel. Two things are needed for the encryption process, they are: an encryption algorithm and a key. An encryption algorithm denotes the procedure that has been used in encryption. Encryption occurs at the sender side.

- **Decryption:** The opposite process of encryption is known as Decryption. It is the process of transforming ciphertext into plaintext. Cryp-

tography makes use of the decryption technique at the receiver's end to acquire the actual message from the unreadable message (Ciphertext). Two things are needed for the decryption process, they are: a decryption algorithm and a key. A decryption algorithm refers to the procedure that has been used in decryption.

- **Key:** A key can be a numeric or alpha numeric text or probably a special symbol. The key is used in conjunction with the encryption algorithm at the point where encryption occurs on the plaintext and it's also used in conjunction with the decryption algorithm at the point where decryption occurs on the ciphertext. The process of choosing a key in Cryptography is very essential since the security of the encryption algorithm relies directly on this. For example, with Caesar cipher, if Alice uses a key of 3 to encrypt the plaintext "Minister" then the corresponding ciphertext will be "Plqlvwhu".

2.1.2 Purpose of Cryptography

Cryptography makes provision for a number of security goals to guarantee the privacy of data and non-alteration of data among others. Because of the security benefits of cryptography, it is extensively being used today. Find below the various goals of cryptography [86], [26].

- **Confidentiality:** This implies that before the transmission and receiv-

ing of the information by the system, both the sender and the receiver must be authorized to access the data and not anybody else.

- **Authentication:** This ensures that both the sender and receiver of data of any system is verified to make sure that the data is from an authorized person.
- **Integrity:** Only the approved party is permitted to make changes to the information that has been sent. No other person except the sender and the receiver has been granted power to make changes to the message that has been sent.
- **Non Repudiation:** This means that both the sender and receiver of a message cannot deny that they have sent a message.
- **Access Control:** This implies that only the approved parties have the authority to access the given information.

2.2 Symmetric Encryption

Symmetric Encryption is also known as secret-key encryption or shared key encryption. It is a form of encryption that uses the same key for both encryption and decryption. Figure 2.1 shows the symmetric encryption.

The key must be distributed securely to all parties the sender intends to send the encrypted information to. Data Encryption Standard (DES) is a

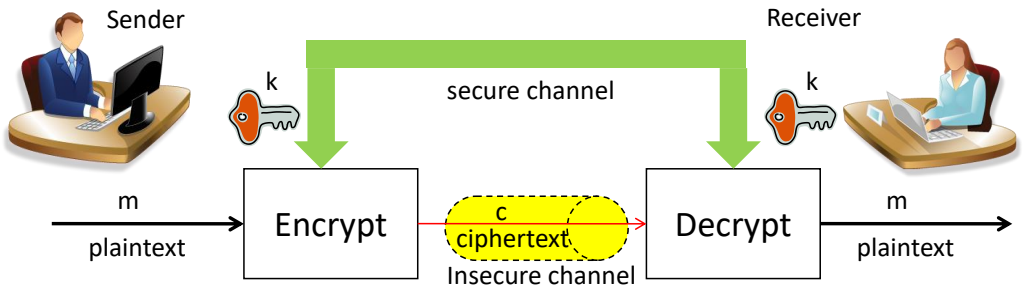

Figure 2.1: Symmetric Encryption. The same key is used for encryption and decryption.

popular symmetric encryption scheme for electronic data, such as electronic funds and transmission in the banking sector. In terms of speed, it is more faster as compared to the asymmetric algorithm and this is due to the fact that it requires less computational power. However, because of the small key Space of the DES algorithm which is 2^{56} possible keys and simple algorithm design, it is susceptible to a number of attacks, which are: brute force, linear cryptanalysis and differential cryptanalysis attacks [4], [82]. Since these existing attacks can break DES, it was replaced with Advanced Encryption Standard (AES). AES provides support for three different key sizes, which are : 128, 192 and 256. They can also be referred to as AES-128, AES-192 and AES-256. Studies have shown that AES is vulnerable to various side-channel attacks if it's not implemented properly [45].

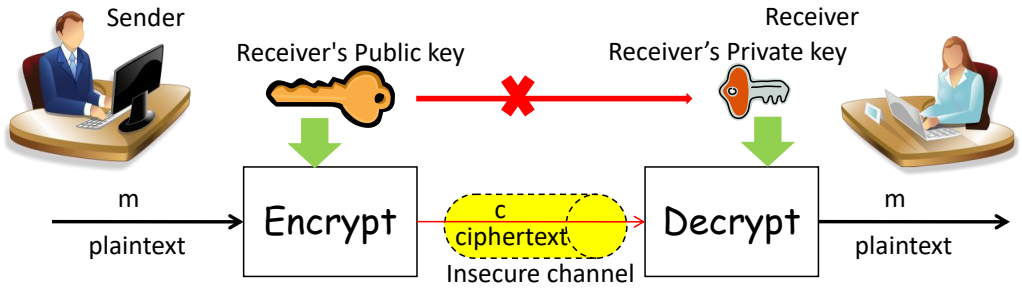

Figure 2.2: Asymmetric Encryption. Different keys are used for encryption and decryption, and the private key cannot be computed from the public key.

2.3 Asymmetric Encryption

Asymmetric Encryption is also known as public-key encryption. Unlike symmetric encryption, it does not use the same key for encryption and decryption. Rather, it uses an encryption key (also known as the public key, which is made known to the public) for encryption and a decryption key (also known as the private key, which is private to the user) for decryption. Figure 2.2 shows the asymmetric encryption.

There are different types of asymmetric encryption schemes, each of which utilizes unique encryption algorithms. Two examples of such schemes are El-Gamal [63] and RSA [91].

2.4 Homomorphic Encryption

Homomorphic encryption allows operations on the encrypted text whose results, when decrypted, yields the same results as the same operations applied on the corresponding plaintexts. This encryption is used when the computing party is not supposed to know the plaintext that it is operating on, thus ensuring that there is no data leakage.

Figure 2.3 is a conceptual example for addition in homomorphic encryption where the plaintexts are m_1, m_2, while $E(m_1)$, $E(m_2)$ are the corresponding ciphertexts.

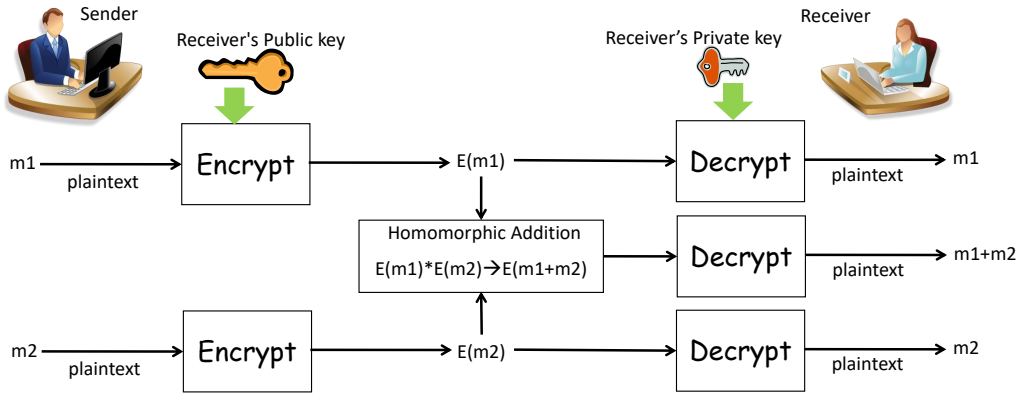

Figure 2.3: A conceptual model for Homomorphic Encryption. HE allows computations on encrypted data.

As shown in the figure, Homomorphic Addition of $E(m_1)$ and $E(m_2)$ provides the same outcome $E(m_1 + m_2)$ as directly encrypting $m_1 + m_2$. Thus, we can obtain the plaintext $m_1 + m_2$ by decrypting $E(m_1) \cdot E(m_2)$ [47]. In other words, $E(m_1 + m_2)$ can be calculated by using only $E(m_1)$ and $E(m_2)$, i.e.,

$E(m_1) \cdot E(m_2) \rightarrow E(m_1 + m_2)$, where $E()$ is one homomorphic encryption algorithm supporting homomorphic addition operation.

2.4.1 Types of Homomorphic Encryption

Homomorphic encryption can be classified as Partially Homomorphic Encryption (PHE), Somewhat Homomorphic Encryption (SWHE) and Fully Homomorphic Encryption (FHE). PHE permits only one operation, for unlimited times, over encrypted data. SWHE supports a limited number of operations for a limited number of times. FHE supports any arbitrary operation for an unlimited number of times [1].

2.4.2 Paillier Public Key Encryption

Paillier PKE is a PHE scheme that is well-known and has been implemented in various privacy-preserving frameworks. In the Paillier theorem, two randomly large prime numbers p and q are chosen such that $p = 2p' + 1$ and $q = 2q' + 1$, where p' and q' are also prime numbers. The two theorems in Paillier cryptosystem state that:

1. For any $x \in \mathbb{Z}_n, (1 + n)^x = 1 + x \cdot n \bmod n^2$.

2. Let λ be the least common multiple of $p - 1$ and $q - 1$, $\lambda = lcm(p - 1, q - 1) = 2p'q'$. For any $x \in \mathbb{Z}^*_{n^2}$, $x^{n\lambda} = 1 \bmod n^2$.

Paillier can achieve homomorphic properties by using three algorithms: Key Generation $KeyGen(\kappa)$, Encryption $Enc(pk, m)$, and Decryption $Dec(sk, c)$ [58].

- $KeyGen(\kappa)$: For a security parameter $\kappa \in \mathbb{Z}^+$, two large prime numbers, $p = 2p' + 1$ and $q = 2q' + 1$ of equal length ($|p| = |q| = \kappa$) will be chosen randomly. Note that p' and q' are also prime numbers. Let $n = pq$ and $\lambda = lcm(p-1, q-1) = 2p'q'$. An integer $g \in \mathbb{Z}^*_{n^2}$ is randomly selected and $\mu = (L(g^\lambda \bmod n^2))^{-1} \bmod n$ is computed, where L is defined as a function $L(x) = \frac{x-1}{n}$. Then, $KeyGen(\kappa)$ generates public key $pk = (n, g)$ and private key $sk = (\lambda, \mu)$.

- $Enc(pk, m)$: Given public key pk and a message $m \in \mathbb{Z}_n$, $c = E(m) = g^m \cdot r^n \bmod n^2$ is the cyphertext of m, where $r \in \mathbb{Z}^*_n$ is a random number and $gcd(r, n) = 1$.

- $Dec(sk, c)$: Given private key sk and ciphertext $c = Enc(m)$, the plaintext message can be computed as $m = L(c^\lambda \bmod n^2) \cdot \mu \bmod n$.

Paillier Correctness: As stated previously, $m = Dec(sk, c) = L(c^\lambda \bmod n^2) \cdot \mu \bmod n$. Here we want to prove that this algorithm outputs the original message [58].

$$
\begin{aligned}
m = Dec(sk, c) = L(c^{\lambda} \bmod n^2) \cdot \mu \bmod n = \frac{L(c^{\lambda} \bmod n^2)}{L(g^{\lambda} \bmod n^2)} \bmod n \\
= \frac{L((g^m \cdot r^n)^{\lambda} \bmod n^2)}{L(g^{\lambda} \bmod n^2)} \bmod n = \frac{L((g^{m\lambda} \cdot r^{n\lambda}) \bmod n^2)}{L(g^{\lambda} \bmod n^2)} \bmod n \\
= \frac{L(g^{m\lambda} \bmod n^2)}{L(g^{\lambda} \bmod n^2)} \bmod n = \frac{L((1+n)^{m\lambda} \bmod n^2)}{L((1+n)^{\lambda} \bmod n^2)} \bmod n \\
= \frac{L((1+mn\lambda) \bmod n^2)}{L((1+n\lambda) \bmod n^2)} \bmod n = \frac{m\lambda}{\lambda} \bmod n = m \bmod n = m
\end{aligned}
\tag{2.1}
$$

Paillier Security: It has been proven that Paillier is secure against chosen plaintext attack and its security analysis can be found in [70].

Paillier Properties: The Paillier PKE satisfies two notable homomorphic properties:

- Homomorphic Addition: The encrypted sum of two plaintexts is equal to the product of their corresponding ciphertexts, $E(m_1) \cdot E(m_2) = E(m_1 + m_2)$.
- Homomorphic Multiplication: The encryption of the product of two plaintexts is equal to the encrypted plaintext to the power of the other plaintext, $E(m_1)^{m_2} = E(m_1 \cdot m_2)$.

It must be noted that Paillier PKE does not support homomorphic multiplication directly over two ciphertexts.

2.5 Hash Function

A hash function is a function that maps data of arbitrary size into a fixed size. The fixed-size value is called a digest or a hash value. For a variable-length message M and a hash function H, the hash value h is calculated as: $h = H(M)$. A hash function is a one-way function, which means that it is computationally impossible to find the associated message given the hash value. A cryptographic hash function is a hash function suitable for cryptography and must be resistant to cryptanalysis. The required properties of cryptography hash functions are stated below [60]:

- Resistance to Pre-image Attacks: The computation of the corresponding message M of a given hash value must not be feasible. Hash functions that do not satisfy this requirement are vulnerable to pre-image attacks.

- Resistance to Second Pre-image Attacks: For a message M_1 and its hash value h, it must not be feasible to compute a different input message M_2 from the hash value. The functions that do not satisfy this property are susceptible to second pre-image attacks.

Table 2.1: Differences between Hash function, Symmetric, and Asymmetric Encryption Algorithms.

Algorithms			
Features	Hash Function	Symmetric	Asymmetric
No. of keys	0	1	2
Key length	256 bits	128 bits	2048 bits
Commonly used	SHA	AES	RSA
Speed	Fast	Fast	Relatively low
Complexity	Medium	Medium	High
Key Sharing	N/A	Hard	Easy and secure

- Resistance to Collision Attacks: It is difficult to find two messages M_1 and M_2 that have the same hash value $H(M_1) = H(M_2)$. Functions failing to satisfy this property are vulnerable to birthday attacks.

There is no key involved in hash function. In the symmetric encryption algorithms, the same key is used in encryption and decryption unlike in the asymmetric algorithms where two different keys are used. According to the National Institute of Standards and Technology (NIST), Table 2.1 shows the differences between the hash function, symmetric, and asymmetric encryption algorithms.

2.6 Digital Signature

Digital signature is used to guarantee the authenticity and integrity of an electronic message. The receiver of a digitally-signed message uses the sig-

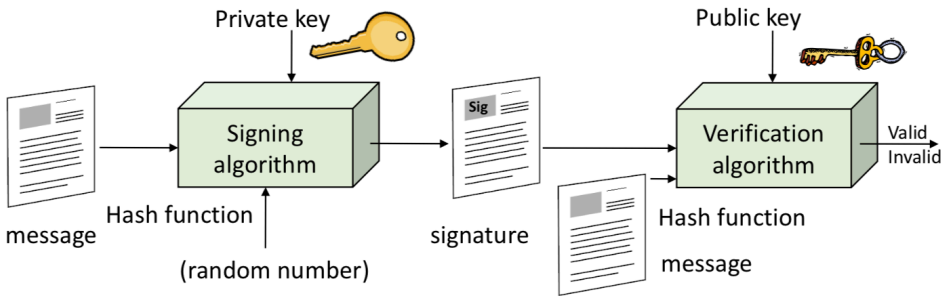

Figure 2.4: The digital signature technique

nature to verify the authenticity of the sender and ensures that the message is unaltered. Figure 2.4 shows how a digital signature is constructed.

Digital signatures use hash functions and public-key cryptography, where the sender generates a hash value from the message, and then encrypts it with its private key. After receiving the message, the receiver uses the public key and the hash function to verify whether the signature matches the message.

Chapter 3

Models and Design Goals

In this chapter, we formalize our system model, security model, and identify our design goals on continuous privacy-preserving histogram query in smart grid communications.

3.1 System Model

Focusing on the continuous privacy-preserving histogram query scenario at the residential users in smart grid communications, our system model mainly includes four types of entities, namely, a trusted authority (TA), a control center (CC), a residential gateway (GW) and a set of residential users $\mathbb{U} = \{U_1, U_2, ..., U_N\}$, as shown in Fig. 3.1, where N indicates the number of residential users in the set $\mathbb{U}$.

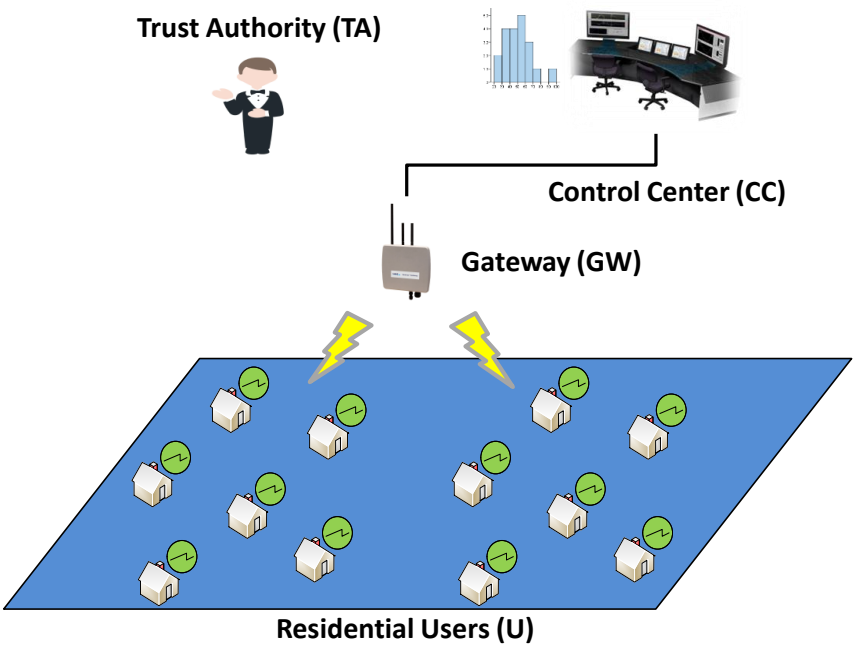

Figure 3.1: System model under consideration

3.1.1 Trust Authority (TA)

TA is a fully trusted entity who is responsible for initializing the whole system. Specifically, it generates secret keys for other entities in the system and distributes the keys through secure channels. After initializing the system, TA will not be involved in the subsequent continuous privacy-preserving histogram query process.

3.1.2 Control Center (CC)

CC is the core entity in the system, who is responsible for data collecting, processing and analyzing the nearly real-time data from $\mathbb{U}$ for monitoring

the health of smart grid.

3.1.3 Residential Gateway (GW)

GW serves as a relay and aggregator role in the system. That is, GW relays the information from CC to $\mathbb{U}$ and at the same time collects and aggregates the data from $\mathbb{U}$, and forwards the aggregated data to the CC.

3.1.4 Residential Users $\mathbb{U}$

Each user $U_i \in \mathbb{U} = \{U_1, U_2, ..., U_N\}$ is equipped with a smart meter, which collects and reports the nearly real-time electricity usage data m_i, e.g., every 30 minutes, to CC via GW. Consider CC divides the time to time periods, $T_1, T_2, T_3, \cdots$. After collecting the data from the users $\mathbb{U}$ at each timer period T_p, CC will generate a Histogram for the time period T_p. Formally, our continuous privacy-preserving Histogram query is defined as follows.

- Assume every user's data m_i is in the range $[0, \alpha)$, e.g., $\alpha = 100$ during every time period T_p. The CC equally divides the range $[0, \alpha]$ into β classes, each class is with size $\delta = \frac{\alpha}{\beta}$. For example, if $\beta = 10$, we have the size $\delta = \frac{\alpha}{\beta} = \frac{100}{10} = 10$, and all classes are shown in Table 3.1.

- Then, for each class j, where $1 \leq j \leq \beta$, CC wants to know how many users' data are within the range $[(j-1) \times \delta, j \times \delta)$ for drawing the

Class	1	2	3	4	5	6	7	8	9	10
Range	[0,9)	[10,19)	[20,29)	[30,39)	[40,49)	[50,59)	[60,69)	[70,79)	[80,89)	[90,99)

Table 3.1: An example of classes and ranges

Histogram in the time period T_p. For the simplification of description, we consider each user's data m_i is an integer in the range $[0, \alpha)$. Then, CC initializes two arrays: Int[] A = new Int[β], Int[] B = new Int[β], where the array A is used to store the sum of data for each class j, while the array B is used to store the count of data for each class j.

- With the help of gateway GW, each user U_i will report its data m_i to CC. If m_i falls into the class j, then

$$A[j] = A[j] + m_i, \qquad B[j] = B[j] + 1$$

Finally, CC can draw the Histogram in the time period T_p. For example, Table 3.2 shows an example of Histogram results, and figures (3.2),(3.3) draw the corresponding Histograms.

Class	1	2	3	4	5	6	7	8	9	10
Sum (A)	15	60	150	245	450	715	650	375	85	0
Count (B)	3	4	6	7	10	13	10	5	1	0

Table 3.2: An example of Histogram results

Privacy Requirements. As the goal of this work is to achieve continuous privacy-preserving Histogram query, each user U_i's data m_i should be privacy-preserving, no one else, including GW and CC, can read the data

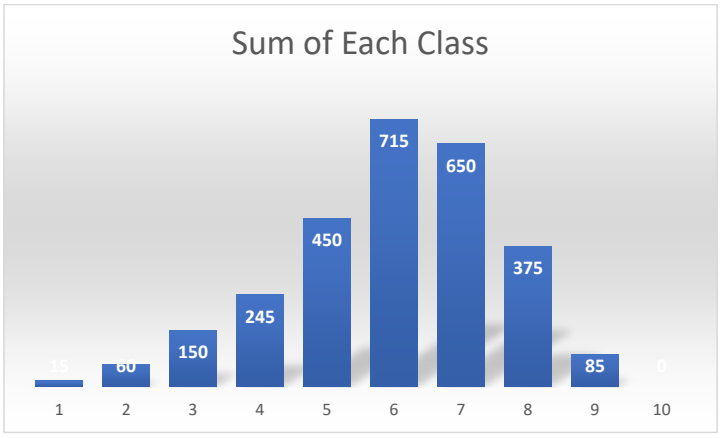

Figure 3.2: Histogram – Sum of Each Class

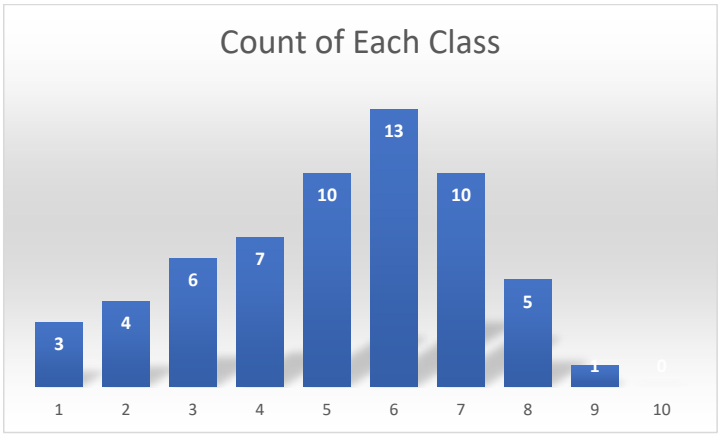

Figure 3.3: Histogram – Count of Each Class

m_i. At the same time, only CC can read the results of Histogram, e.g., the results in Table 3.2.

3.2 Security Model

In our security model, we consider both CC and GW are *honest-but-curious*. That is, they will faithfully follow the continuous privacy-preserving Histogram query protocol, but also attempt to get to know each individual user's data once the condition is reached. In addition, residential users $\mathbb{U} = \{U_1, U_2, ..., U_N\}$ are also honest, i.e., each U_i won't report false data to CC or collude with CC to get other users' individual data.

Note that there are other possible attacks, i.e., false data injection attack [18], Denial of Service (DoS) attack, in smart grid communications. Since our focus is on privacy-preserving histogram query, those attacks are beyond the scope of this work, and will be discussed in our future work.

3.3 Design Goals

Our design goal is to develop a continuous privacy-preserving Histogram query scheme for smart grid communication such that CC can obtain and draw the Histogram for users' electricity consumptions at each time period. Specifically, the following two desirable goals should be satisfied.

- *The proposed scheme should be privacy-preserving.* Only CC can read the Histogram results in the proposed scheme, and no one (including CC and GW) can read each individual user data.

- *The proposed scheme should be efficient.* Not only the encryption at user side, aggregation at gateway, but also the decryption at control center should be efficient in terms of computational cost. In addition, the protocol should achieve communication efficiency.

Chapter 4

Proposed Continuous Privacy-Preserving Histogram Query Scheme

In this chapter, we present our continuous privacy-preserving Histogram query scheme (CPHQ) for secure smart grid communications, which mainly consists of the following four phases: system initialization, user report generation, privacy-preserving report aggregation, and Histogram report reading.

4.1 System Initialization

According to the system model in our previous chapter, it is reasonable to assume the trusted authority (TA) to initialize the whole system.

Given the security parameter κ, e.g., $\kappa = 512$, TA first builds up the Paillier Cryptosystem by running the following steps:

- choose two large prime numbers p, q, where $|p| = |q| = \kappa$;
- compute the RSA modulus $n = pq$ and $\lambda = lcm(p-1, q-1)$;
- define a function $L(u) = \frac{u-1}{n}$;
- choose a base $g \in \mathbb{Z}^*_{n^2}$;
- compute $\mu = (L(g^\lambda \bmod n^2))^{-1} \bmod n$;
- set the public key $pk = (n, g)$, and the corresponding private key $sk = (\lambda, \mu)$.

Given a total of N residential users $\mathbb{U} = \{U_1, U_2, ..., U_N\}$ in our system model, when every user U_i's data m_i is in the range $[0, \alpha)$ during every time period T_p, and CC equally divides the range $[0, \alpha]$ into β classes, we have the size of each class is $\delta = \frac{\alpha}{\beta}$. In reality, it is reasonable to assume the electricity usage of each user is less than 100 in a time period T_p, and CC can set $\beta = 10$, then the size $\delta = \frac{\alpha}{\beta}$ will be $\delta = 10$. After that, CC chooses a super-increasing

sequence

$$\vec{\mathbf{a}} = (a_1 = 1, a_2, \cdots, a_\beta),$$

where $a_2, \cdots, a_\beta$ are large primes, such that

$$\sum_{j=1}^{i-1} a_j \cdot N \cdot \delta < a_i,$$

for $i = 2, \cdots, l$, and

$$\sum_{i=1}^{\beta} a_i \cdot N \cdot \delta < n.$$

Furthermore, TA randomly chooses $(N+1)$ elements $\mathcal{S} = (s_0, s_1, s_2, \cdots, s_N)$, where each $s_i \in \mathbb{Z}_n$ and

$$\sum_{i=0}^{N} s_i \equiv 0 \bmod n.$$

Finally, TA chooses two secure hash functions $H_1, H_2 : \{0,1\}^* \to \mathbb{Z}_n$, sets the public parameter $\mathbf{Pub} = (\alpha, \beta, \delta, pk = (n, g), \vec{\mathbf{a}}, H_1, H_2)$, and publishes **Pub** in the whole system, so that each entity in this system can access **Pub**.

Key Distribution. After publishing the public parameter **Pub**, TA runs the following steps to distribute secret key materials via secure channels.

- TA securely distributes $s_0, sk = (\lambda, \mu)$ to CC.
- TA also securely distributes s_i to each user $U_i \in \mathbb{U}$.

Note that, the gateway GW just serves as the relay and runs the aggregation

operations, TA does not assign any secret key materials to GW. However, GW can still access the public parameter **Pub**.

After the system initialization, TA will be offline, and the continuous privacy-preserving Histogram query (CPHQ) will be run by other entities in the system.

4.2 User Report Generation

At each time period T_p, CC will initialize two integer arrays $A[]$, $B[]$, as

$$\text{Int[] A} = \text{new Int}[\beta], \qquad \text{Int[] B} = \text{new Int}[\beta]$$

and then wait for all users' reports via GW.

While for each user $U_i \in \mathbb{U}$ at each time period T_p, U_i will report his/her data m_i by running the following steps:

- First, the user U_i computes A_i, B_i from m_i as follows.

$$A_i = m_i \bmod \delta; \qquad B_i = \frac{(m_i - A_i)}{\delta}$$

  i.e., $m_i = B_i \cdot \delta + A_i$. For instance, when $\alpha = 100, \beta = 10, \delta = 10$, and $m_i = 78$, we have $A_i = 8$, $B_i = 7$.

- Based on the value of B_i, the user U_i chooses $a_{(B_i+1)}$ from the public parameter $\vec{\mathbf{a}} = (a_1 = 1, a_2, \cdots, a_\beta)$. In the numerical example, we have $a_{(B_i+1)} = a_8$.

- Next, for encrypting A_i, B_i, the user U_i uses the public key $pk = (n, g)$, the secret key s_i to compute the ciphertexts $E[A_i]$, $E[B_i]$ as follows,

  $$E[A_i] = g^{a_{(B_i+1)} \cdot A_i} H_1(T_p)^{s_i} \bmod n^2; \quad E[B_i] = g^{a_{(B_i+1)} \cdot 1} H_2(T_p)^{s_i} \bmod n^2$$

  where T_p represents the current time period, which should be unique and the same for all users during the continuous privacy-preserving Histogram query. Note that, we use $E[\cdot]$ to represent a ciphertext.

- Finally, the user U_i forwards $(E[A_i], E[B_i])$ to the gateway GW.

4.3 Privacy-preserving Report Aggregation

After receiving all ciphertexts $(E[A_i], E[B_i])$ from all users $U_i \in \mathbb{U}$, GW homomorphically runs the privacy-preserving report aggregation as follows:

$$E[A] = \prod_{i=1}^{N} E[A_i] \bmod n^2; \qquad E[B] = \prod_{i=1}^{N} E[B_i] \bmod n^2$$

After that, GW reports $(E[A], E[B])$ to the control center CC.

Note that, according to the constructions of $E[A], E[B]$, we know

$$E[A] = \prod_{i=1}^{N} E[A_i] \bmod n^2$$
$$= g^{(a_1 \cdot \sum_{a_{(B_i+1)}=a_1} A_i)+(a_2 \cdot \sum_{a_{(B_i+1)}=a_2} A_i)+\cdots+(a_\beta \cdot \sum_{a_{(B_i+1)}=a_\beta} A_i)} \cdot H_1(T_p)^{\sum_{i=1}^{N} s_i} \bmod n^2,$$

$$E[B] = \prod_{i=1}^{N} E[B_i] \bmod n^2$$
$$= g^{(a_1 \cdot \sum_{a_{(B_i+1)}=a_1} 1)+(a_2 \cdot \sum_{a_{(B_i+1)}=a_2} 1)+\cdots+(a_\beta \cdot \sum_{a_{(B_i+1)}=a_\beta} 1)} \cdot H_2(T_p)^{\sum_{i=1}^{N} s_i} \bmod n^2.$$

4.4 Histogram Report Reading

After receiving $(E[A], E[B])$, CC uses the secret key s_0 to update $(E[A], E[B])$ as follows,

$$E[A]' = E[A] \cdot H_1(T_p)^{s_0} \bmod n^2$$
$$= g^{(a_1 \cdot \sum_{a_{(B_i+1)}=a_1} A_i)+(a_2 \cdot \sum_{a_{(B_i+1)}=a_2} A_i)+\cdots+(a_\beta \cdot \sum_{a_{(B_i+1)}=a_\beta} A_i)} \cdot H_1(T_p)^{\sum_{i=0}^{N} s_i} \bmod n^2,$$

$$E[B]' = E[B] \cdot H_2(T_p)^{s_0} \bmod n^2$$
$$= g^{(a_1 \cdot \sum_{a_{(B_i+1)}=a_1} 1)+(a_2 \cdot \sum_{a_{(B_i+1)}=a_2} 1)+\cdots+(a_\beta \cdot \sum_{a_{(B_i+1)}=a_\beta} 1)} \cdot H_2(T_p)^{\sum_{i=0}^{N} s_i} \bmod n^2.$$

Let

$$A = (a_1 \cdot \sum_{a_{(B_i+1)}=a_1} A_i) + (a_2 \cdot \sum_{a_{(B_i+1)}=a_2} A_i) + \cdots + (a_\beta \cdot \sum_{a_{(B_i+1)}=a_\beta} A_i)$$

and

$$B = (a_1 \cdot \sum_{a_{(B_i+1)}=a_1} 1) + (a_2 \cdot \sum_{a_{(B_i+1)}=a_2} 1) + \cdots + (a_\beta \cdot \sum_{a_{(B_i+1)}=a_\beta} 1).$$

From the condition set in our system initialization phase,

$$\sum_{i=1}^{\beta} a_i \cdot N \cdot \delta < n.$$

Then, we know both A, B are less than n, i.e., $A < n, B < n$.

Also, because $\sum_{i=0}^{N} s_i \equiv 0 \bmod n$, i.e.,

$$\sum_{i=0}^{N} s_i = k \cdot n \qquad \text{for some integer } k,$$

we know

$$E[A]' = g^A \cdot H_1(T_p)^{kn} \bmod n^2; \qquad E[B]' = g^B \cdot H_2(T_p)^{kn} \bmod n^2$$

are valid Paillier ciphertexts of (A, B), respectively.

Because $(E[A]', E[B]')$ are valid Paillier ciphertexts, CC can use the private key $sk = (\lambda, \mu)$ to recover A, B as

$$A = (a_1 \cdot \sum_{a_{(B_i+1)}=a_1} A_i) + (a_2 \cdot \sum_{a_{(B_i+1)}=a_2} A_i) + \cdots + (a_\beta \cdot \sum_{a_{(B_i+1)}=a_\beta} A_i) < n,$$

$$B = (a_1 \cdot \sum_{a_{(B_i+1)}=a_1} 1) + (a_2 \cdot \sum_{a_{(B_i+1)}=a_2} 1) + \cdots + (a_\beta \cdot \sum_{a_{(B_i+1)}=a_\beta} 1) < n.$$

By invoking the Algorithm 1 with the input A, CC can recover the aggregated data

$$A \rightarrow A[] = (A[1], A[2], \cdots, A[\beta]),$$

where each $A[j] \in A[]$ is

$$A[j] = \sum_{a_{(B_i+1)}=a_j} A_i.$$

Similarly, by invoking the Algorithm 1 with the input B, CC can recover the aggregated data

$$B \rightarrow B[] = (B[1], B[2], \cdots, B[\beta]),$$

where each $B[j] \in B[]$ is

$$B[j] = \sum_{a_{(B_i+1)}=a_j} 1.$$

Algorithm 1 Recover the aggregated report array

procedure RECOVER THE AGGREGATED REPORT ARRAY
Input: $\vec{\mathbf{a}} = (a_1 = 1, a_2, \cdots, a_\beta)$ and a large integer $D < n$
Output: an integer array $D[] = (D[1], D[2], \cdots, D[\beta])$
    Set $X_l = D$
    for $j = \beta$ to 2 **do**
        $X_{j-1} = X_j \bmod a_j$
        $D[j] = \frac{X_j - X_{j-1}}{a_j}$
    end for
    $D[1] = X_1$
    return $D[] = (D[1], D[2], \cdots, D[\beta])$
end procedure

Because $\sum_{j=1}^{i-1} a_j \cdot N \cdot \delta < a_i$ for $i = 2, \cdots, l$, and $\sum_{i=1}^{\beta} a_i \cdot N \cdot \delta < n$, the correctness of Algorithm 1 is obvious.

Finally, CC updates the array A[] by computing

$$A[j] = A[j] + (j-1) \cdot \delta \cdot B[j], \qquad \text{for } i = 1, 2, \cdots, \beta$$

and draws the Histogram for the time period T_p based on the arrays $A[] = (A[1], A[2], \cdots, A[\beta]$ and $B[] = (B[1], B[2], \cdots, B[\beta])$.

Note that, in order to support the continuous Histogram query, the above protocol will be repeated in each time period.

4.5 A Simple Numerical Example

In order for better understanding the continuous Histogram query, we use a simple numerical example to illustrate its correctness.

When $\alpha = 100, \beta = 10, \delta = 10$, we consider $N = 15$ users in our system, their data and the corresponding ciphertexts are shown in Table 4.1.

After the aggregation at the gateway GW, the ciphertexts $E[A], E[B]$ are

$$\begin{aligned} E[A] &= \prod_{i=1}^{N} E[A_i] \bmod n^2 \\ &= g^{(a_1 \cdot \sum_{a_{(B_i+1)}=a_1} A_i) + (a_2 \cdot \sum_{a_{(B_i+1)}=a_2} A_i) + \cdots + (a_\beta \cdot \sum_{a_{(B_i+1)}=a_\beta} A_i)} \cdot H_1(T_p)^{\sum_{i=1}^{N} s_i} \bmod n^2 \\ &= g^{a_1 \cdot 0 + a_2 \cdot 4 + a_3 \cdot 6 + a_4 \cdot 14 + a_5 \cdot 14 + a_6 \cdot 13 + a_7 \cdot 13 + a_8 \cdot 8 + a_9 \cdot 0 + a_{10} \cdot 1} \cdot H_1(T_p)^{\sum_{i=1}^{N} s_i} \bmod n^2 \end{aligned}$$

Table 4.1: A Simple Numerical Example with $N = 15$ Users

User	m_i	A_i	B_i	$E[A_i]$	$E[B_i]$
U_1	$m_1 = 67$	$A_1 = 7$	$B_1 = 6$	$E[A_1] = g^{a_7 \cdot 7} \cdot H_1(T_p)^{s_1} \bmod n^2$	$E[B_1] = g^{a_7 \cdot 1} \cdot H_2(T_p)^{s_1} \bmod n^2$
U_2	$m_2 = 58$	$A_2 = 8$	$B_2 = 5$	$E[A_2] = g^{a_6 \cdot 8} \cdot H_1(T_p)^{s_2} \bmod n^2$	$E[B_2] = g^{a_6 \cdot 1} \cdot H_2(T_p)^{s_2} \bmod n^2$
U_3	$m_3 = 48$	$A_3 = 8$	$B_3 = 4$	$E[A_3] = g^{a_5 \cdot 8} \cdot H_1(T_p)^{s_3} \bmod n^2$	$E[B_3] = g^{a_5 \cdot 1} \cdot H_2(T_p)^{s_3} \bmod n^2$
U_4	$m_4 = 35$	$A_4 = 5$	$B_4 = 3$	$E[A_4] = g^{a_4 \cdot 5} \cdot H_1(T_p)^{s_4} \bmod n^2$	$E[B_4] = g^{a_4 \cdot 1} \cdot H_2(T_p)^{s_4} \bmod n^2$
U_5	$m_5 = 26$	$A_5 = 6$	$B_5 = 2$	$E[A_5] = g^{a_3 \cdot 6} \cdot H_1(T_p)^{s_5} \bmod n^2$	$E[B_5] = g^{a_3 \cdot 1} \cdot H_2(T_p)^{s_5} \bmod n^2$
U_6	$m_6 = 14$	$A_6 = 4$	$B_6 = 1$	$E[A_6] = g^{a_2 \cdot 4} \cdot H_1(T_p)^{s_6} \bmod n^2$	$E[B_6] = g^{a_2 \cdot 1} \cdot H_2(T_p)^{s_6} \bmod n^2$
U_7	$m_7 = 46$	$A_7 = 6$	$B_7 = 4$	$E[A_7] = g^{a_5 \cdot 6} \cdot H_1(T_p)^{s_7} \bmod n^2$	$E[B_7] = g^{a_5 \cdot 1} \cdot H_2(T_p)^{s_7} \bmod n^2$
U_8	$m_8 = 63$	$A_8 = 3$	$B_8 = 6$	$E[A_8] = g^{a_7 \cdot 3} \cdot H_1(T_p)^{s_8} \bmod n^2$	$E[B_8] = g^{a_7 \cdot 1} \cdot H_2(T_p)^{s_8} \bmod n^2$
U_9	$m_9 = 71$	$A_9 = 1$	$B_9 = 7$	$E[A_9] = g^{a_8 \cdot 1} \cdot H_1(T_p)^{s_9} \bmod n^2$	$E[B_9] = g^{a_8 \cdot 1} \cdot H_2(T_p)^{s_9} \bmod n^2$
U_{10}	$m_{10} = 39$	$A_{10} = 9$	$B_{10} = 3$	$E[A_{10}] = g^{a_4 \cdot 9} \cdot H_1(T_p)^{s_{10}} \bmod n^2$	$E[B_{10}] = g^{a_4 \cdot 1} \cdot H_2(T_p)^{s_{10}} \bmod n^2$
U_{11}	$m_{11} = 55$	$A_{11} = 5$	$B_{11} = 5$	$E[A_{11}] = g^{a_6 \cdot 5} \cdot H_1(T_p)^{s_{11}} \bmod n^2$	$E[B_{11}] = g^{a_6 \cdot 1} \cdot H_2(T_p)^{s_{11}} \bmod n^2$
U_{12}	$m_{12} = 77$	$A_{12} = 7$	$B_{12} = 7$	$E[A_{12}] = g^{a_8 \cdot 7} \cdot H_1(T_p)^{s_{12}} \bmod n^2$	$E[B_{12}] = g^{a_8 \cdot 1} \cdot H_2(T_p)^{s_{12}} \bmod n^2$
U_{13}	$m_{13} = 62$	$A_{13} = 2$	$B_{13} = 6$	$E[A_{13}] = g^{a_7 \cdot 2} \cdot H_1(T_p)^{s_{13}} \bmod n^2$	$E[B_{13}] = g^{a_7 \cdot 1} \cdot H_2(T_p)^{s_{13}} \bmod n^2$
U_{14}	$m_{14} = 61$	$A_{14} = 1$	$B_{14} = 6$	$E[A_{14}] = g^{a_7 \cdot 1} \cdot H_1(T_p)^{s_{14}} \bmod n^2$	$E[B_{13}] = g^{a_7 \cdot 1} \cdot H_2(T_p)^{s_{14}} \bmod n^2$
U_{15}	$m_{15} = 91$	$A_{15} = 1$	$B_{14} = 9$	$E[A_{15}] = g^{a_{10} \cdot 1} \cdot H_1(T_p)^{s_{15}} \bmod n^2$	$E[B_{15}] = g^{a_{10} \cdot 1} \cdot H_2(T_p)^{s_{15}} \bmod n^2$

$$
\begin{aligned}
E[B] &= \prod_{i=1}^{N} E[B_i] \bmod n^2 \\
&= g^{(a_1 \cdot \sum_{a_{(B_i+1)}=a_1} 1)+(a_2 \cdot \sum_{a_{(B_i+1)}=a_2} 1)+\cdots+(a_\beta \cdot \sum_{a_{(B_i+1)}=a_\beta} 1)} \cdot H_2(T_p)^{\sum_{i=1}^{N} s_i} \bmod n^2 \\
&= g^{a_1 \cdot 0+a_2 \cdot 1+a_3 \cdot 1+a_4 \cdot 2+a_5 \cdot 2+a_6 \cdot 2+a_7 \cdot 4+a_8 \cdot 2+a_9 \cdot 0+a_{10} \cdot 1} \cdot H_2(T_p)^{\sum_{i=1}^{N} s_i} \bmod n^2
\end{aligned}
$$

After the update with s_0 at the control center, the ciphertexts $E[A], E[B]$ become

$$
\begin{aligned}
E[A]' &= E[A] \cdot H_1(T_p)^{s_0} \bmod n^2 \\
&= g^{a_1 \cdot 0+a_2 \cdot 4+a_3 \cdot 6+a_4 \cdot 14+a_5 \cdot 14+a_6 \cdot 13+a_7 \cdot 13+a_8 \cdot 8+a_9 \cdot 0+a_{10} \cdot 1} \cdot H_1(T_p)^{k \cdot n} \bmod n^2
\end{aligned}
$$

$$
\begin{aligned}
E[B]' &= E[B] \cdot H_1(T_p)^{s_0} \bmod n^2 \\
&= g^{a_1 \cdot 0+a_2 \cdot 1+a_3 \cdot 1+a_4 \cdot 2+a_5 \cdot 2+a_6 \cdot 2+a_7 \cdot 4+a_8 \cdot 2+a_9 \cdot 0+a_{10} \cdot 1} \cdot H_2(T_p)^{k \cdot n} \bmod n^2
\end{aligned}
$$

By decrypting $E[A]', E[B]'$, the control center can obtain

$$A = a_1 \cdot 0 + a_2 \cdot 4 + a_3 \cdot 6 + a_4 \cdot 14 + a_5 \cdot 14 + a_6 \cdot 13 + a_7 \cdot 13 + a_8 \cdot 8 + a_9 \cdot 0 + a_{10} \cdot 1$$

$$B = a_1 \cdot 0 + a_2 \cdot 1 + a_3 \cdot 1 + a_4 \cdot 2 + a_5 \cdot 2 + a_6 \cdot 2 + a_7 \cdot 4 + a_8 \cdot 2 + a_9 \cdot 0 + a_{10} \cdot 1$$

By applying the Algorithm 1, the control center can get the arrays A[1..10] and B[1..10] in Table 4.2.

	1	2	3	4	5	6	7	8	9	10
A[1..10]	0	4	6	14	14	13	13	8	0	1
B[1..10]	0	1	1	2	2	2	4	2	0	1

Table 4.2: The results of A[1..10], B[1..10]

By applying

$$A[j] = A[j] + (j-1) \cdot \delta \cdot B[j], \qquad \text{for } i = 1, 2, \cdots, \beta$$

the control center can obtain the updated results of arrays A[1..10] and B[1..10] in Table 4.3, and can draw the Histograms in Figs.(4.1),(4.2) for the time period T_p.

	1	2	3	4	5	6	7	8	9	10
A[1..10]	0	14	26	74	94	113	253	148	0	91
B[1..10]	0	1	1	2	2	2	4	2	0	1

Table 4.3: The updated results of A[1..10], B[1..10]

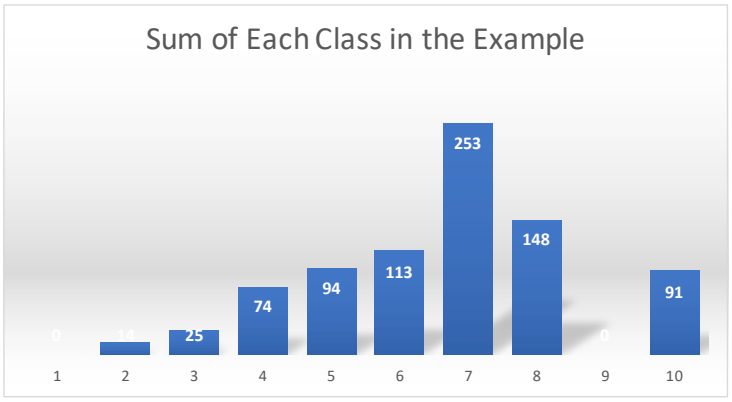

Figure 4.1: Histogram – Sum of Each Class in Example

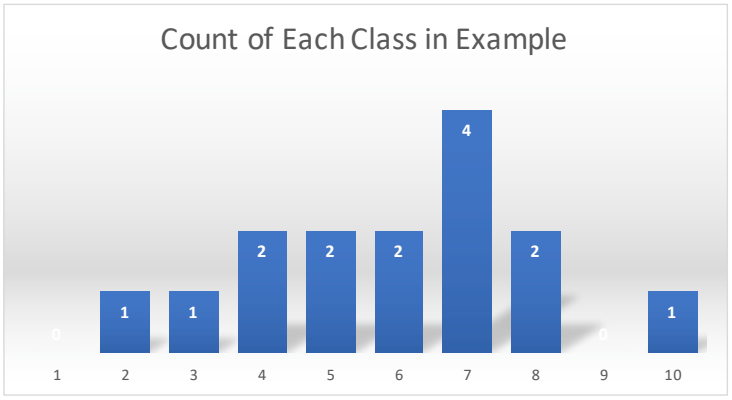

Figure 4.2: Histogram – Count of Each Class in Example

Chapter 5

Security Analysis and Performance Evaluation

In this chapter, we will present the security analysis and performance evaluation of our proposed continuous privacy-preserving histogram query scheme.

5.1 Security Analysis

First, we analyze the security of the proposed continuous privacy-preserving histogram query scheme. Since it is based on the Paillier cryptosystem [70], and Paillier cryptosystem is semantically secure under chosen-plaintext attack, we can ensure the reported data m_i of each user $U_i \in \mathbb{U}$ at each time period T_p is privacy-preserving. The detailed analyses are as follows.

• *In the proposed query scheme, the reported data m_i of each user $U_i \in \mathbb{U}$ at each time period T_p is privacy-preserving, only if there is no collusion among other users $\mathbb{U} \setminus \{U_i\}$, GW, and CC.*

First, given the ciphertexts $E[A_i], E[B_i]$ of the reported data m_i of $U_i \in \mathbb{U}$ at each time period T_p, i.e.,

$$E[A_i] = g^{a_{(B_i+1)} \cdot A_i} H_1(T_p)^{s_i} \bmod n^2; \quad E[B_i] = g^{a_{(B_i+1)} \cdot 1} H_2(T_p)^{s_i} \bmod n^2$$

where $A_i = m_i \bmod \delta$; $B_i = \frac{(m_i - A_i)}{\delta}$, if GW colludes with CC, i.e., GW directly sends $E[A_i], E[B_i]$ to CC, then CC can obtain the $E[A_i], E[B_i]$. In addition, if each user $U_j \in \mathbb{U} \setminus \{U_i\}$ also colludes with CC, i.e., U_j shares the secret key s_j to CC, then CC can obtain

$$\begin{aligned} E[A_i]' &= E[A_i] \cdot H_1(T_p)^{\sum_{j=0, j\neq i}^{N} s_j} = g^{a_{(B_i+1)} \cdot A_i} \cdot H_1(T_p)^{\sum_{j=0}^{N} s_j} \\ &= g^{a_{(B_i+1)} \cdot A_i} \cdot H_1(T_p)^{kn} \bmod n^2; \end{aligned}$$

$$\begin{aligned} E[B_i]' &= E[B_i] \cdot H_2(T_p)^{\sum_{j=0, j\neq i}^{N} s_j} = g^{a_{(B_i+1)} \cdot 1} \cdot H_2(T_p)^{\sum_{j=0}^{N} s_j} \\ &= g^{a_{(B_i+1)} \cdot 1} \cdot H_2(T_p)^{kn} \bmod n^2; \end{aligned}$$

Obviously, both $E[A_i]', E[B_i]'$ are valid Paillier ciphertexts, CC can use the private key $sk = (\lambda, \mu)$ to recover $a_{(B_i+1)} \cdot A_i$, $a_{(B_i+1)} \cdot 1$. As a result, through $a_{(B_i+1)}$, the value B_i can be determined, and the reported data $m_i = A_i + B_i \cdot \delta$ can be obtained by CC.

However, in our system model, we consider all users $\mathbb{U} = \{U_1, U_2, ..., U_N\}$ are

honest, and they will not collude with CC. Therefore, even though CC has the private key $sk = (\lambda, \mu)$, CC cannot obtain each user U_i's data m_i at the time period T_p, let alone other entities without the private key $sk = (\lambda, \mu)$ in the system.

Second, because we consider the continuous histogram query, the ciphertexts of reported data m_{i1} and m_{i2} of a user U_i at different timer periods T_{p1}, T_{p2} could be obtained by other entities. In this case, are the reported data m_{i1}, m_{i2} still secure? In the following, we have some discussions on this case. Given the ciphertexts $E[A_{i1}], E[A_{i2}], E[B_{i1}], E[B_{i2}]$, i.e.,

$$E[A_{iw}] = g^{a_{(B_{iw}+1)} \cdot A_{iw}} H_1(T_{pw})^{s_i} \bmod n^2; \quad E[B_{iw}] = g^{a_{(B_{iw}+1)} \cdot 1} H_2(T_{pw})^{s_i} \bmod n^2$$

where $A_{iw} = m_{iw} \bmod \delta$; $B_{iw} = \frac{(m_{iw} - A_{iw})}{\delta}$, and $w \in \{1, 2\}$. If $H_1(T_{p1}) = H_1(T_{p2})$, $H_2(T_{p1}) = H_2(T_{p2})$, we can obtain the values $g^{a_{(B_{i1}+1)} \cdot A_{i1} - a_{(B_{i2}+1)} \cdot A_{i2}}$ and $g^{a_{(B_{i1}+1)} - a_{(B_{i2}+1)}}$, where

$$\frac{E[A_{i1}]}{E[A_{i2}]} = \frac{g^{a_{(B_{i1}+1)} \cdot A_{i1}} H_1(T_{p1})^{s_i}}{g^{a_{(B_{i2}+1)} \cdot A_{i2}} H_1(T_{p2})^{s_i}} = g^{a_{(B_{i1}+1)} \cdot A_{i1} - a_{(B_{i2}+1)} \cdot A_{i2}}$$

and

$$\frac{E[B_{i1}]}{E[B_{i2}]} = \frac{g^{a_{(B_{i1}+1)} \cdot 1} H_2(T_{p1})^{s_i}}{g^{a_{(B_{i2}+1)} \cdot 1} H_2(T_{p2})^{s_i}} = g^{a_{(B_{i1}+1)} - a_{(B_{i2}+1)}}$$

Because $\vec{\mathbf{a}} = (a_1 = 1, a_2, \cdots, a_\beta)$ are public parameters, we can easily obtain $a_{(B_{i1}+1)}$, $a_{(B_{i2}+1)}$ from $g^{a_{(B_{i1}+1)} - a_{(B_{i2}+1)}}$ by trying all possible cases. Then, because A_{i1}, A_{i2} are within a small space, we can again obtain them from

$g^{a_{(B_{i1}+1)} \cdot A_{i1} - a_{(B_{i2}+1)} \cdot A_{i2}}$ by trying all possible cases. Finally, we can obtain the reported data m_{i1} and m_{i2}.

In the above discussion, the reason we can obtain m_{i1} and m_{i2} is that we make the assumptions that $H_1(T_{p1}) = H_1(T_{p2})$, $H_2(T_{p1}) = H_2(T_{p2})$. However, if we choose secure hash functions H_1, H_2, the probabilities of the collisions $H_1(T_{p1}) = H_1(T_{p2})$, $H_2(T_{p1}) = H_2(T_{p2})$ are negligible. Therefore, only if the hash functions H_1, H_2 are secure, we cannot obtain each user's reported data from the continuous histogram query.

As a result, no one (including CC and GW) can read each individual user data.

• *In the proposed query scheme, only CC can read the Histogram results, if all entities in the system follow the protocol.*

This is obvious. As we discussed above, with the help of GW, CC can obtain

$$E[A]' = g^A \cdot H_1(T_p)^{kn} \bmod n^2; \qquad E[B]' = g^B \cdot H_2(T_p)^{kn} \bmod n^2$$

Because only CC has the private key $sk = (\lambda, \mu)$, and there is no collusion, we can conclude that only CC can use the private key $sk = (\lambda, \mu)$ to read the Histogram results.

Summarizing the above analysis, we can see our proposed continuous histogram query scheme is really privacy-preserving.

5.2 Performance Evaluation

In this section, we evaluate the performance of the proposed Continuous Privacy-preserving Histogram Query (CPHQ) scheme. In specific, we mainly focus on the computational overhead and communication cost of the report generation, report aggregation and array recovery phases. To this end, we first theoretically show the computational complexity and communication cost of the three phases. Then, we implement our proposed CPHQ scheme in Java. Afterthat, we conduct experiments on a MacOS platform with a 3.1 GHz Intel Core i7 processor and a memory of 16GB 2133 MHz LPDDR3 to demonstrate the average time consumption for each phase.

5.2.1 Theoretical Analysis

In this subsection, we respectively show the computational complexity and the communication cost of the proposed scheme. Specifically, we first analyze the computational complexity of the report generation, report aggregation, and array recovery phases. Then, we show the communication cost of the CPHQ scheme.

- *Computational Complexity of the Report Generation Phase:* In the report generation phase, each residential user $U_i \in \mathbb{U}$ reports his/her data

m_i by respectively computing

$$E[A_i] = g^{a_{(B_i+1)} \cdot A_i} H_1(T_p)^{s_i} \bmod n^2 \text{ and } E[B_i] = g^{a_{(B_i+1)} \cdot 1} H_2(T_p)^{s_i} \bmod n^2.$$

Therefore, the computational complexity for each residential user $U_i \in \mathbb{U}$ is O(1).

- *Computational Complexity of the Report Aggregation Phase:* In the report aggregation phase, the gateway GW collects data report $(E[A_i], E[B_i])$ from each user U_i, then it computes

$$E[A] = \prod_{i=1}^{N} E[A_i] \bmod n^2 \quad \textit{and} \quad E[B] = \prod_{i=1}^{N} E[B_i] \bmod n^2$$

Therefore, the computational complexity of the report aggregation phase is O(N).

- *Computational Complexity of the Array Recovery Phase:* In the array recovery phase, the control center CC first respectively decrypts $E[A]$ and $E[B]$. Then, following Algorithm 1, it extracts $A[]$ and $B[]$ based on the property of super-increasing sequences with a computational complexity of $O(\beta)$. Therefore, the overall computational complexity of this phase is $O(\beta)$.

- Communication Cost of the Proposed Scheme: The communication cost of the CPHQ scheme comprises of the following two parts: i) *The communication cost between the residential users and GW:* During the

report generation phase, each user U_i uploads $(E[A_i], E[B_i])$ to the gateway GW, and the length of the message is 4κ bits. Thus, the communication cost between the users and GW is $N \cdot 4\kappa$ bits. ii) *The communication cost between GW and CC:* After report aggregation, GW will upload the aggregated user report to CC, which only contains two ciphertexts, namely, $E[A]$ and $E[B]$. Hence, the overall communication cost of the report aggregation phase is 4κ bits.

5.2.2 Experiments

To further demonstrate the time consumption of the three phases in our proposed scheme, we evaluate them with varied numbers of groups and users. In specific, for each pair of numbers of groups and users, we run 5 rounds of warmup and calculate average time consumption of the other 45 for the three phases. Each group consists of 5 to 50 users with an interval of 5, for example: 5, 10, 15, 20 among others. Similarly, we choose the number of groups ranging from 5 to 50 with an interval of 5, e.g., 5, 10, 15, 20 among others. Figures 5.1 to 5.3 show the average time consumption for the three different phases collected from our experiments, and we will respectively discuss them as follows.

To clearly demonstrate the average time consumption of the Report Generation phase, we evaluate it with the numbers of users to be 10, 30, and 50, and the number of groups ranges from 5 to 50. As show in Figure 5.1, the average

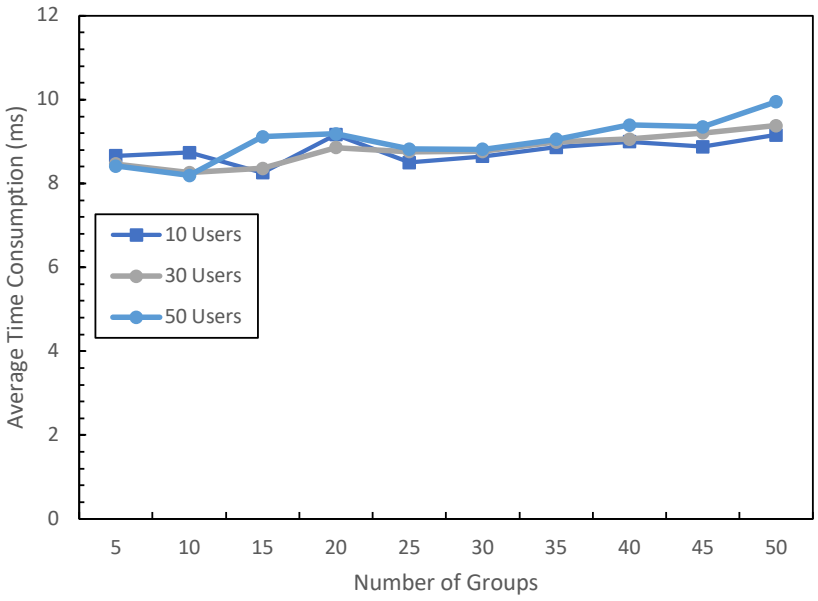

Figure 5.1: The average time consumption for the Report Generation phase for different number of users

time consumption will not be affected by the number of users, and it remains just about the same as the number of groups increase. Specifically, when the number of groups is 50, it will only take about 10 ms for a residential user to generate a report.

As shown in Figure 5.2, the computational cost for the gateway to aggregate reports will only increase with the number of users and is not affected by the number of groups. This is mainly because that, by employing the super-increasing sequence, each report only employs two ciphertexts to represent the values for all groups, so the computational cost for aggregating single user report is $O(1)$.

According to our analysis, the computational cost of this phase should be

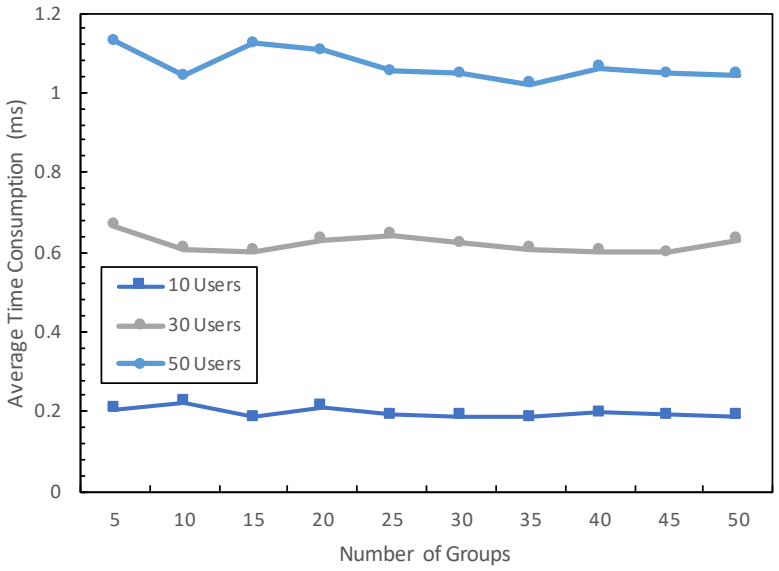

Figure 5.2: The average time consumption for the Report Aggregation phase for different number of users

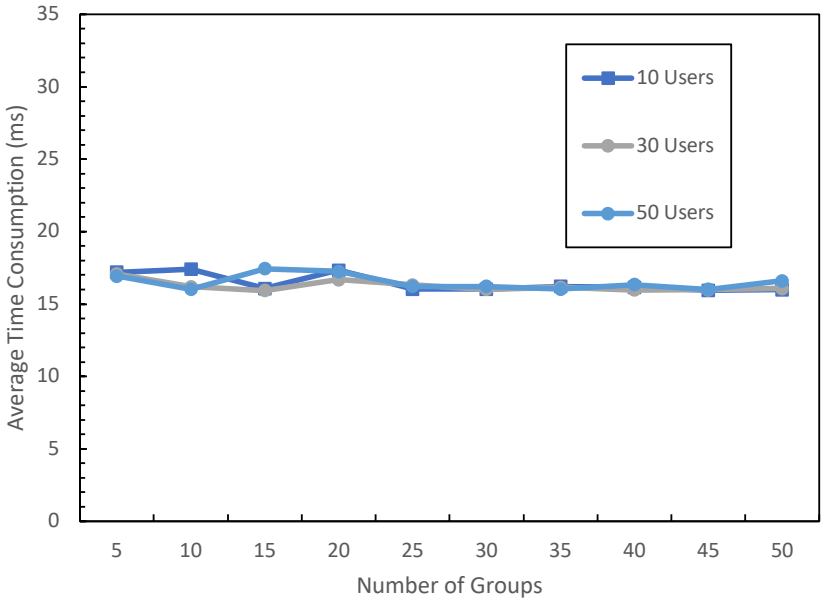

Figure 5.3: The average time consumption for the Array Recovery Phase for different number of users

linear to the number of groups. However, as shown in Figure 5.3, the time consumption will not be affected by the number of the users and the number of groups. This is mainly because, according to the Paillier decryption operation, recovering the array with Algorithm 1 is run on plaintexts and its computational cost is negligible.

Chapter 6

Conclusions and Future Work

In this chapter, we first draw our conclusions, which comprise summarizing our scheme and the contribution of this thesis. After that, we discuss our future research work.

6.1 Conclusions

In this thesis, we propose our Continuous Privacy-preserving Histogram Query (CPHQ) scheme. In the proposed CPHQ scheme, the nearly real-time electricity usage data of residential users are aggregated into histogram data in control center, and the data privacy for each user is preserved. Specifically, based on the Paillier Cryptosystem, our proposed scheme employs the gateway to homomorphically aggregate encrypted user reports for the control

center. In such a way, the control center can obtain aggregated histogram data. Moreover, both the gateway and the control center cannot know any specific user's electricity usage data. We provided security analysis showing that the users' electricity usage is indeed privacy-preserving. In addition, theoretical analysis and experimental results show that our CPHQ scheme works efficiently in terms of computation and communication.

6.2 Future Work

In our future work, we will improve our CPHQ scheme to support the case where each residential user's electricity usage data are float numbers. Currently, in our CPHQ scheme, each residential user's electricity usage data is only considered to be integer. As a result, supporting the feature on float numbers will make our CPHQ scheme more practical. Moreover, we will add features to support dynamically adding or removing residential users. That is, due to some issues, a residential user may not be able to upload his/her report timely, and the control center can still get a partial result of the two Histogram charts. This improvement can boost the robustness of our scheme.

Bibliography

[1] Abbas Acar, Hidayet Aksu, A Selcuk Uluagac, and Mauro Conti, *A survey on homomorphic encryption schemes: Theory and implementation*, ACM Computing Surveys (CSUR) **51** (2018), no. 4, 1–35.

[2] Roya Ahmadiahangar, Argo Rosin, Ivo Palu, and Aydin Azizi, *Challenges of smart grids implementation*, Demand-side Flexibility in Smart Grid, Springer, 2020, pp. 1–15.

[3] Massoud Amin, *Challenges in reliability, security, efficiency, and resilience of energy infrastructure: Toward smart self-healing electric power grid*, 2008 IEEE Power and energy society general meeting-conversion and delivery of electrical energy in the 21st century, IEEE, 2008, pp. 1–5.

[4] Ryndel V Amorado, Ariel M Sison, and Ruji P Medina, *Enhanced data encryption standard (des) algorithm based on filtering and striding techniques*, Proceedings of the 2019 2nd International Conference on Information Science and Systems, 2019, pp. 252–256.

[5] Seyed Ali Arefifar, Yasser Abdel-Rady I Mohamed, and Tarek HM El-Fouly, *Comprehensive operational planning framework for self-healing control actions in smart distribution grids*, IEEE Transactions on Power Systems **28** (2013), no. 4, 4192–4200.

[6] Atefeh Dehghani Ashkezari, Nasser Hosseinzadeh, Ayoub Chebli, and Mahammed Albadi, *Development of an enterprise geographic information system (gis) integrated with smart grid*, Sustainable Energy, Grids and Networks **14** (2018), 25–34.

[7] Canadian Electricity Association et al., *The smart grid: a pragmatic approach*, 2012.

[8] Danielly B Avancini, Joel JPC Rodrigues, Simion GB Martins, Ricardo AL Rabêlo, Jalal Al-Muhtadi, and Petar Solic, *Energy meters evolution in smart grids: A review*, Journal of Cleaner Production **217** (2019), 702–715.

[9] Amam Hossain Bagdadee and Li Zhang, *Renewable energy based self-healing scheme in smart grid*, Energy Reports **6** (2020), 166–172.

[10] Andrea Bartoli, Juan Hernandez-Serrano, M Soriano, Mischa Dohler, Apostolous Kountouris, and Dominique Barthel, *Secure lossless aggregation for smart grid m2m networks*, 2010 first IEEE international conference on smart grid communications, IEEE, 2010, pp. 333–338.

[11] Coalton Bennett and Steven B Wicker, *Decreased time delay and security enhancement recommendations for ami smart meter networks*, 2010 Innovative Smart Grid Technologies (ISGT), IEEE, 2010, pp. 1–6.

[12] Stuart A Boyer, *Scada: supervisory control and data acquisition*, International Society of Automation, 2009.

[13] Richard E Brown, *Impact of smart grid on distribution system design*, 2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, IEEE, 2008, pp. 1–4.

[14] William R Cassel, *Distribution management systems: Functions and payback*, IEEE Transactions on Power Systems **8** (1993), no. 3, 796–801.

[15] T.-H. Hubert Chan, Elaine Shi, and Dawn Song, *Privacy-preserving stream aggregation with fault tolerance*, Financial Cryptography and Data Security - 16th International Conference, FC 2012, Kralendijk, Bonaire, Februray 27-March 2, 2012, Revised Selected Papers, Lecture Notes in Computer Science, vol. 7397, Springer, 2012, pp. 200–214.

[16] Supriyo Chatterjea and Paul Havinga, *A dynamic data aggregation scheme for wireless sensor networks*, Proc. Program for Research on Integrated Systems and Circuits, Veldhoven, The Netherlands (2003), 924–935.

[17] Changsong Chen, Shanxu Duan, Tong Cai, Bangyin Liu, and Gangwei Hu, *Smart energy management system for optimal microgrid economic operation*, IET renewable power generation **5** (2011), no. 3, 258–267.

[18] Po-Yu Chen, Shusen Yang, Julie A McCann, Jie Lin, and Xinyu Yang, *Detection of false data injection attacks in smart-grid systems*, IEEE Communications Magazine **53** (2015), no. 2, 206–213.

[19] Gordon Clarke, Deon Reynders, and Edwin Wright, *Practical modern scada protocols: Dnp3, 60870.5 and related systems*, Newnes, 2004.

[20] Frances M Cleveland, *Cyber security issues for advanced metering infrasttructure (ami)*, 2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, IEEE, 2008, pp. 1–5.

[21] Axel Daneels and Wayne Salter, *What is scada?*, (1999).

[22] Soma Shekara Sreenadh Reddy Depuru, Lingfeng Wang, Vijay Devabhaktuni, and Nikhil Gudi, *Smart meters for power grid—challenges, issues, advantages and status*, 2011 IEEE/PES Power Systems Conference and Exposition, IEEE, 2011, pp. 1–7.

[23] G. Dileep, *A survey on smart grid technologies and applications*, Renew. Energy **146** (2020), 2589–2625.

[24] Aris Dimeas, Stefan Drenkard, Nikos Hatziargyriou, Stamatis Karnouskos, Koen Kok, Jan Ringelstein, and Anke Weidlich, *Smart*

houses in the smart grid: Developing an interactive network., IEEE Electrification Magazine **2** (2014), no. 1, 81–93.

[25] Alfredo D'elia, Fabio Viola, Federico Montori, Marco Di Felice, Luca Bedogni, Luciano Bononi, Alberto Borghetti, Paolo Azzoni, Paolo Bellavista, Daniele Tarchi, et al., *Impact of interdisciplinary research on planning, running, and managing electromobility as a smart grid extension*, IEEE Access **3** (2015), 2281–2305.

[26] Aaron E Earle, *Wireless security handbook*, CRC Press, 2005.

[27] Samuel East, Jonathan Butts, Mauricio Papa, and Sujeet Shenoi, *A taxonomy of attacks on the dnp3 protocol*, International Conference on Critical Infrastructure Protection, Springer, 2009, pp. 67–81.

[28] Costas Efthymiou and Georgios Kalogridis, *Smart grid privacy via anonymization of smart metering data*, 2010 first IEEE international conference on smart grid communications, IEEE, 2010, pp. 238–243.

[29] Janaka B Ekanayake, Nick Jenkins, Kithsiri Liyanage, Jianzhong Wu, and Akihiko Yokoyama, *Smart grid: technology and applications*, John Wiley & Sons, 2012.

[30] Zekeriya Erkin and Gene Tsudik, *Private computation of spatial and temporal power consumption with smart meters*, Applied Cryptography and Network Security - 10th International Conference, ACNS 2012, Sin-

gapore, June 26-29, 2012. Proceedings, Lecture Notes in Computer Science, vol. 7341, Springer, 2012, pp. 561–577.

[31] NIST Framework, *Roadmap for smart grid interoperability standards, release 2.0*, NIST special publication 1108R2 (2012), 1–225.

[32] Steffen Fries, Hans Joachim Hof, and Maik Seewald, *Enhancing iec 62351 to improve security for energy automation in smart grid environments*, 2010 Fifth International Conference on Internet and Web Applications and Services, IEEE, 2010, pp. 135–142.

[33] Amrita Ghosal and Mauro Conti, *Key management systems for smart grid advanced metering infrastructure: A survey*, IEEE Communications Surveys & Tutorials **21** (2019), no. 3, 2831–2848.

[34] NIST Smart Grid, *Guidelines for smart grid cyber security: Vol. 2, privacy and the smart grid*, Guideline, Aug **6** (2010).

[35] Smart Grid, *European technology platform for electricity networks of the future*, European Commission (2005).

[36] Vehbi C Gungor, Dilan Sahin, Taskin Kocak, Salih Ergut, Concettina Buccella, Carlo Cecati, and Gerhard P Hancke, *Smart grid and smart homes: Key players and pilot projects*, IEEE Industrial Electronics Magazine **6** (2012), no. 4, 18–34.

[37] BB Gupta and Tafseer Akhtar, *A survey on smart power grid: frameworks, tools, security issues, and solutions*, Annals of Telecommunications **72** (2017), no. 9-10, 517–549.

[38] Aaron Hansen, Jason Staggs, and Sujeet Shenoi, *Security analysis of an advanced metering infrastructure*, International Journal of Critical Infrastructure Protection **18** (2017), 3–19.

[39] Vinay M Igure, Sean A Laughter, and Ronald D Williams, *Security issues in scada networks*, computers & security **25** (2006), no. 7, 498–506.

[40] Luca Lena Jansen, Nikoleta Andreadou, Ioulia Papaioannou, and Antonios Marinopoulos, *Smart grid lab research in europe and beyond*, International Journal of Energy Research **44** (2020), no. 3, 1307–1336.

[41] N Jenkins, JB Ekanayake, and G Strbac, *Distributed generation. the institution of engineering and technology*, London, United Kingdom (2010), 293.

[42] Marc Joye and Benoît Libert, *A scalable scheme for privacy-preserving aggregation of time-series data*, Financial Cryptography and Data Security - 17th International Conference, FC 2013, Okinawa, Japan, April 1-5, 2013, Revised Selected Papers, Lecture Notes in Computer Science, vol. 7859, Springer, 2013, pp. 111–125.

[43] Sebastian Käbisch, Anton Schmitt, Martin Winter, and Jorg Heuer, *Interconnections and communications of electric vehicles and smart grids*, 2010 First IEEE International Conference on Smart Grid Communications, IEEE, 2010, pp. 161–166.

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

[45] Alan Kaminsky, Michael Kurdziel, and Stanisław Radziszowski, *An overview of cryptanalysis research for the advanced encryption standard*, 2010-MILCOM 2010 MILITARY COMMUNICATIONS CONFERENCE, IEEE, 2010, pp. 1310–1316.

[46] Himanshu Khurana, Mark Hadley, Ning Lu, and Deborah A Frincke, *Smart-grid security issues*, IEEE Security & Privacy **8** (2010), no. 1, 81–85.

[47] Ovunc Kocabas and Tolga Soyata, *Towards privacy-preserving medical cloud computing using homomorphic encryption*, Virtual and Mobile Healthcare: Breakthroughs in Research and Practice, IGI Global, 2020, pp. 93–125.

[48] Wei Kong, Jian Shen, Pandi Vijayakumar, Youngju Cho, and Victor Chang, *A practical group blind signature scheme for privacy protec-*

tion in smart grid, Journal of Parallel and Distributed Computing **136** (2020), 29–39.

[49] Bhaskar Krishnamachari, Deborah Estrin, and Stephen B. Wicker, *The impact of data aggregation in wireless sensor networks*, 22nd International Conference on Distributed Computing Systems, Workshops (ICDCSW '02) July 2-5, 2002, Vienna, Austria, Proceedings, IEEE Computer Society, 2002, pp. 575–578.

[50] Joanna Kulik, Wendi Rabiner Heinzelman, and Hari Balakrishnan, *Negotiation-based protocols for disseminating information in wireless sensor networks*, Wirel. Networks **8** (2002), no. 2-3, 169–185.

[51] Anil Lamba, *A through analysis on protecting cyber threats and attacks on cps embedded subsystems*, Available at SSRN 3517474 (2020).

[52] Fengjun Li, Bo Luo, and Peng Liu, *Secure information aggregation for smart grids using homomorphic encryption*, 2010 first IEEE international conference on smart grid communications, IEEE, 2010, pp. 327–332.

[53] Chen-Chun Lin, Chia-Han Yang, and Joseph Z Shyua, *A comparison of innovation policy in the smart grid industry across the pacific: China and the usa*, Energy Policy **57** (2013), 119–132.

[54] Hui Lin, Chen Chen, Jianhui Wang, Junjian Qi, Dong Jin, Zbigniew T Kalbarczyk, and Ravishankar K Iyer, *Self-healing attack-resilient pmu*

network for power system operation, IEEE Transactions on Smart Grid **9** (2016), no. 3, 1551–1565.

[55] Stephanie Lindsey, Cauligi S. Raghavendra, and Krishna M. Sivalingam, *Data gathering algorithms in sensor networks using energy metrics*, IEEE Trans. Parallel Distributed Syst. **13** (2002), no. 9, 924–935.

[56] 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** (2012), no. 4, 981–997.

[57] Zhi Liu, Toshitaka Tsuda, Hiroshi Watanabe, Satoko Ryuo, and Nagateru Iwasawa, *Data driven cyber-physical system for landslide detection*, Mobile Networks and Applications **24** (2019), no. 3, 991–1002.

[58] Rongxing Lu, *Privacy-enhancing aggregation techniques for smart grid communications*, Springer, 2016.

[59] Ralph E Mackiewicz, *Overview of iec 61850 and benefits*, 2006 IEEE Power Engineering Society General Meeting, IEEE, 2006, pp. 8–pp.

[60] Ali Maetouq, Salwani Mohd Daud, Noor Azurati Ahmad, Nurazean Maarop, Nilam Nur Amir Sjarif, and Hafiza Abas, *Comparison of hash function algorithms against attacks: A review*, Int. J. Adv. Comput. Sci. Appl. **9** (2018), no. 8, 98–103.

[61] Daphne Ngar-yin Mah, Johannes Marinus van der Vleuten, Jasper Chiman Ip, and Peter Ronald Hills, *Governing the transition of socio-*

technical systems: a case study of the development of smart grids in korea, Energy Policy **45** (2012), 133–141.

[62] Bilal Masood, M Arif Khan, Sobia Baig, Guobing Song, Ateeq Ur Rehman, Saif Ur Rehman, Rao M Asif, and Muhammad Babar Rasheed, *Investigation of deterministic, statistical and parametric nb-plc channel modeling techniques for advanced metering infrastructure*, Energies **13** (2020), no. 12, 3098.

[63] Andreas V Meier, *The elgamal cryptosystem*, Joint Advanced Students Seminar, 2005.

[64] Anthony R Metke and Randy L Ekl, *Smart grid security technology*, 2010 Innovative Smart Grid Technologies (ISGT), IEEE, 2010, pp. 1–7.

[65] Kevin Mets, Juan Aparicio Ojea, and Chris Develder, *Combining power and communication network simulation for cost-effective smart grid analysis*, IEEE Communications Surveys & Tutorials **16** (2014), no. 3, 1771–1796.

[66] Craig Howard Miller, *Optimized energy management system*, September 25 2007, US Patent 7,274,975.

[67] Ramyar Rashed Mohassel, Alan Fung, Farah Mohammadi, and Kaamran Raahemifar, *A survey on advanced metering infrastructure*, International Journal of Electrical Power & Energy Systems **63** (2014), 473–484.

[68] Alaa Mohd, Egon Ortjohann, Andreas Schmelter, Nedzad Hamsic, and Danny Morton, *Challenges in integrating distributed energy storage systems into future smart grid*, 2008 IEEE international symposium on industrial electronics, IEEE, 2008, pp. 1627–1632.

[69] Lin Mu, Enjin Zhao, Yuewei Wang, and Albert Zomaya, *Buoy sensor cyberattack detection in offshore petroleum cyber-physical systems*, IEEE Transactions on Services Computing (2020).

[70] Pascal Paillier, *Public-key cryptosystems based on composite degree residuosity classes*, Advances in Cryptology - EUROCRYPT '99, International Conference on the Theory and Application of Cryptographic Techniques, Prague, Czech Republic, May 2-6, 1999, Proceeding (Jacques Stern, ed.), Lecture Notes in Computer Science, vol. 1592, Springer, 1999, pp. 223–238.

[71] Emiliano Pallotti and Federica Mangiatordi, *Smart grid cyber security requirements*, 2011 10th International conference on environment and electrical engineering, IEEE, 2011, pp. 1–4.

[72] Robert G Pratt, Patrick J Balducci, Clint Gerkensmeyer, Srinivas Katipamula, Michael CW Kintner-Meyer, Thomas F Sanquist, Kevin P Schneider, and Thomas J Secrest, *The smart grid: an estimation of the energy and co2 benefits*, Tech. report, Pacific Northwest National Lab.(PNNL), Richland, WA (United States), 2010.

[73] Bartosz Przydatek, Dawn Xiaodong Song, and Adrian Perrig, *SIA: secure information aggregation in sensor networks*, Proceedings of the 1st International Conference on Embedded Networked Sensor Systems, SenSys 2003, Los Angeles, California, USA, November 5-7, 2003, ACM, 2003, pp. 255–265.

[74] Farrokh Rahimi and Ali Ipakchi, *Demand response as a market resource under the smart grid paradigm*, IEEE Transactions on smart grid **1** (2010), no. 1, 82–88.

[75] Md Golam Rahman, M Fahad Bin Ramim Chowdhury, Md Abdulla Al Mamun, Md Rakib Hasan, and Sayeed Mahfuz, *Summary of smart grid: Benefits and issues*, International Journal of Scientific & Engineering Research **4** (2013), no. 3, 1–5.

[76] Bradford P Roberts and Chet Sandberg, *The role of energy storage in development of smart grids*, Proceedings of the IEEE **99** (2011), no. 6, 1139–1144.

[77] Sebastian Rohjans, Christian Dänekas, and Mathias Uslar, *Requirements for smart grid ict-architectures*, 2012 3rd IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe), IEEE, 2012, pp. 1–8.

[78] Javier Rodríguez Roncero, *Integration is key to smart grid management*, CIRED Seminar 2008: SmartGrids for Distribution, IET, 2008, pp. 1–4.

[79] Elaine Shi, T.-H. Hubert Chan, Eleanor Gilbert Rieffel, Richard Chow, and Dawn Song, *Privacy-preserving aggregation of time-series data*, Proceedings of the Network and Distributed System Security Symposium, NDSS 2011, San Diego, California, USA, 6th February - 9th February 2011, The Internet Society, 2011.

[80] Jing Shi, Rui Zhang, Yunzhong Liu, and Yanchao Zhang, *Prisense: Privacy-preserving data aggregation in people-centric urban sensing systems*, INFOCOM 2010. 29th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies, 15-19 March 2010, San Diego, CA, USA, IEEE, 2010, pp. 758–766.

[81] C-L Su, C-N Lu, and M-C Lin, *Wide area network performance study of a distribution management system*, International Journal of Electrical Power & Energy Systems **22** (2000), no. 1, 9–14.

[82] Hanqi Tang, Qifu Tyler Sun, Xiaolong Yang, and Keping Long, *A network coding and des based dynamic encryption scheme for moving target defense*, IEEE Access **6** (2018), 26059–26068.

[83] Wayes Tushar, Bo Chai, Chau Yuen, David B Smith, Kristin L Wood, Zaiyue Yang, and H Vincent Poor, *Three-party energy management with distributed energy resources in smart grid*, IEEE Transactions on Industrial Electronics **62** (2014), no. 4, 2487–2498.

[84] Darshana Upadhyay and Srinivas Sampalli, *Scada (supervisory control and data acquisition) systems: Vulnerability assessment and security recommendations*, Computers & Security **89** (2020), 101666.

[85] David P Varodayan and Grace Xingxin Gao, *Redundant metering for integrity with information-theoretic confidentiality*, 2010 first IEEE international conference on smart grid communications, IEEE, 2010, pp. 345–349.

[86] Om Prakash Verma, Ritu Agarwal, Dhiraj Dafouti, and Shobha Tyagi, *Notice of violation of ieee publication principles: Peformance analysis of data encryption algorithms*, 2011 3rd International Conference on Electronics Computer Technology, vol. 5, IEEE, 2011, pp. 399–403.

[87] Maarten Wolsink, *The research agenda on social acceptance of distributed generation in smart grids: Renewable as common pool resources*, Renewable and Sustainable Energy Reviews **16** (2012), no. 1, 822–835.

[88] Fang-Jing Wu, Yu-Fen Kao, and Yu-Chee Tseng, *From wireless sensor networks towards cyber physical systems*, Pervasive and Mobile computing **7** (2011), no. 4, 397–413.

[89] Ye Yan, Yi Qian, Hamid Sharif, and David Tipper, *A survey on cyber security for smart grid communications*, IEEE Communications Surveys & Tutorials **14** (2012), no. 4, 998–1010.

[90] Ali Zangeneh and Mohammad Moradzadeh, *Self-healing: Definition, requirements, challenges and methods*, Microgrid Architectures, Control and Protection Methods, Springer, 2020, pp. 509–525.

[91] Xin Zhou and Xiaofei Tang, *Research and implementation of rsa algorithm for encryption and decryption*, Proceedings of 2011 6th international forum on strategic technology, vol. 2, IEEE, 2011, pp. 1118–1121.

Appendix A

Source Code

A.1 System Model

```
import java.io.FileOutputStream;
import java.io.PrintStream;
import java.math.BigInteger;
import java.security.NoSuchAlgorithmException;
import java.util.Scanner;
public class SystemModel {
  public static int N;
  public static final int alpha=100;
  public static final int beta=10;
  public static final int delta = alpha/beta;

public static void main(String[] args) throws NoSuchAlgorithmException,
    ↪ Exception {
```

```
Scanner input = new Scanner(System.in);
        System.out.println("Enter the number of Residential Users");
        N = input.nextInt();

       TrustAuthority trustObject = new TrustAuthority();
       Pub pubObject = trustObject.systemInit();

       BigInteger[] S = trustObject.getS();
       Paillier.PrivateKey Prikey = trustObject.getPriKey();

       ControlCenter ccobject = new ControlCenter();
       //Sets variable pub in Control Center with pubObject(Distributes
           ↪ pubObject to CC)
       ccobject.init(pubObject);
       // Distributes Snot and private key (lambda,mu) to the CC.
       ccobject.setKeys(S[0], Prikey);

       ResidentialGateway rgObject = new ResidentialGateway();
       rgObject.init(pubObject);

       ResidentialUser[] users = new ResidentialUser[N];
       BigInteger[][] data = new BigInteger[N][];
       for (int i = 0; i < N; i++) {
           users[i] = new ResidentialUser();
           users[i].initialize(rgObject, pubObject, S[i + 1]);
           data[i] = users[i].reportGeneration();
       }

        for (int i = 0; i < N; i++) {
```

```
            rgObject.putReport(data[i][0], data[i][1]);
        }

        BigInteger A = rgObject.getE_A();
        BigInteger B = rgObject.getE_B();
        ccobject.setCumAandB(A, B);
        ccobject.readReportAandB();
        ccobject.recoverAandB();
        int ARecArray[] = ccobject.getA();
        int BRecArray[] = ccobject.getB();

        System.out.println("The␣Recovered␣Arrays␣Are␣Presented␣Below");
        for (int i = 0; i < SystemModel.beta; i++) {

            System.out.println("A[" + i + "]␣is:␣" + ARecArray[i] + "␣␣␣B[" + i
                ↪  + "]␣is:␣" + BRecArray[i]);
        }
    }
}
```

A.2 Pub

```
import java.math.BigInteger;
public class Pub {

    int alpha;
```

```
    int beta;
    int delta;

    Paillier.PublicKey PubKey;
    BigInteger n ;
    BigInteger g ;

    BigInteger[] a ;

    public Pub(int alpha,int beta,int delta,Paillier.PublicKey PubKey,
        ↪ BigInteger n, BigInteger g, BigInteger[] a){

        this.alpha = alpha;
        this.beta = beta;
        this.delta = delta;
        this.PubKey = PubKey;
        this.n = n;
        this.g = g;
        this.a = a;
    }
}
```

A.3 Trust Authority

```
import java.math.BigInteger;
import java.nio.charset.StandardCharsets;
```

```
import java.security.MessageDigest;
import java.security.NoSuchAlgorithmException;
import java.security.SecureRandom;
public class TrustAuthority {
 private Paillier paillier ;

    private Paillier.PublicKey pubkey ;
    private Paillier.PrivateKey prikey;

    BigInteger Hash1;
    BigInteger Hash2;



    public static final String H1= "This␣is␣Hash␣function␣1";
    public static final String H2= "This␣is␣Hash␣function␣2";

    BigInteger[] S;



    public void keyGeneration(){
        paillier = new Paillier();
        paillier.keyGeneration(512);

       pubkey = paillier.getPubkey();
       prikey = paillier.getPrikey();

    }

    public Paillier.PublicKey getPubKey(){
```

```
    return pubkey;
}

public Paillier.PrivateKey getPriKey(){
    return prikey;
}

public BigInteger[] getS(){
    return S;
}

//Creates an array of random numbers S
public BigInteger[] elementsS(int N){
    BigInteger[] Selements = new BigInteger[N+1];
    BigInteger sum = BigInteger.ZERO;

    BigInteger n = pubkey.getN();
    for(int i=0;i<N;i++){
        Selements[i] = randomZN(n);
        sum = sum.add(Selements[i]).mod(n);
    }
    Selements[N] = n.subtract(sum);
    return Selements;
}

//Generates a random no. in Z*n
public static BigInteger randomZN(BigInteger n) {
    BigInteger r;
    do {
```

```
        r = new BigInteger(n.bitLength(), new SecureRandom());
    } while (r.compareTo(n) >= 0 );
    return r;
}

//Creates a hash function
public static BigInteger hashValues(String originalString , String h,
    ↪ BigInteger n) throws NoSuchAlgorithmException{
    MessageDigest digest = MessageDigest.getInstance("SHA-256");
    byte[] encodedhash = digest.digest((originalString+h).getBytes(
        ↪ StandardCharsets.UTF_8));
    BigInteger hashValue = new BigInteger(encodedhash).mod (n);
    return hashValue;
}

//Initializes the whole system and follows the steps in 3.1
public Pub systemInit(){

     keyGeneration();
     int alpha = SystemModel.alpha;
     int beta = SystemModel.beta;
     int delta = SystemModel.delta;
     Paillier.PublicKey Pubkey = getPubKey();
     BigInteger n = Pubkey.getN();
     BigInteger g = Pubkey.getG();
     BigInteger[] a = ControlCenter.createSuperIncreasingSequence(
         ↪ SystemModel.N, SystemModel.beta, SystemModel.delta);
       S = elementsS(SystemModel.N);
```

```
        Pub pubobject = new Pub(alpha,beta,delta,Pubkey,n,g,a);
       return pubobject;

   }
}
```

A.4 Control Center

```
import java.math.BigInteger;
import java.security.NoSuchAlgorithmException;
public class ControlCenter {
 int[] A = new int[SystemModel.beta]; //Stores sum of data for each class
   int[] B = new int[SystemModel.beta]; //Stores count of data for each class

   BigInteger Snot;
   Paillier.PrivateKey Prikey;
   BigInteger cumA,cumB;
   int Tp = 0;
   private Pub pub;

   public void init(Pub pub) {

       this.pub = pub;

     }
```

```
public static BigInteger[] createSuperIncreasingSequence(int N,int beta,int
    ↪  delta){

    BigInteger[] a = new BigInteger[beta];
    BigInteger NmulDelta = BigInteger.valueOf(N).multiply( BigInteger.
        ↪ valueOf(delta));
    BigInteger lowerbound = NmulDelta;
    a[0] = BigInteger.ONE;
    for (int i = 1; i < beta; i++) {

            a[i]=lowerbound.nextProbablePrime();
            lowerbound = lowerbound.add( a[i].multiply(NmulDelta));

    }
    return a;
}

public void setKeys(BigInteger Snot, Paillier.PrivateKey Prikey){
    this.Snot = Snot;
    this.Prikey = Prikey;

}

public void setCumAandB(BigInteger A, BigInteger B){
    this.cumA = A;
    this.cumB = B;
}

public int[] getA(){
```

```
    return A;
}

public int[] getB(){
    return B;
}

//This method updates E[A] , E[B] with Snot.
public void readReportAandB() throws NoSuchAlgorithmException{
    int Tp = this.Tp++;
    BigInteger nSquared = pub.n.pow(2);
    cumA = cumA.multiply(TrustAuthority.hashValues(String.valueOf(Tp),
           TrustAuthority.H1, pub.n).modPow(Snot,nSquared)).mod(nSquared);

    cumB = cumB.multiply(TrustAuthority.hashValues(String.valueOf(Tp),
           TrustAuthority.H2, pub.n).modPow(Snot,nSquared)).mod(nSquared);
}

public static int[] recoverArray(BigInteger[] a,BigInteger D){
    BigInteger prevX = D;
    BigInteger prevXminus1 = BigInteger.ZERO;
    int[] ret = new int[SystemModel.beta];
    for (int i = (SystemModel.beta-1); i >= 1; i --) {
           prevXminus1 = prevX.mod(a[i]);
           ret[i] = ((prevX.subtract(prevXminus1)).divide(a[i])).intValue()
               ↪ ;
           prevX = prevXminus1;

    }
```

```
        ret[0] = prevXminus1.intValue();
        return ret;
    }

    public void recoverAandB() throws Exception{
        A = recoverArray(pub.a,Paillier.decrypt(cumA, pub.PubKey, Prikey));
        B = recoverArray(pub.a,Paillier.decrypt(cumB, pub.PubKey, Prikey));

        for(int i=0;i<SystemModel.beta;i++){
            A[i] = A[i] + i * SystemModel.delta * B[i];
        }
    }
}
```

A.5 Residential Gateway

```
import java.math.BigInteger;
public class ResidentialGateway {

    private BigInteger E_A = BigInteger.ONE;
    private BigInteger E_B = BigInteger.ONE;

      private Pub pub;

   public void init(Pub pub) {
```

```
        this.pub = pub;

     }

       //Aggregates reports from all users into E[A] and E[B]
    public void putReport(BigInteger E_Ai, BigInteger E_Bi) {

         BigInteger nSquared = pub.n.pow(2);

        E_A = E_A.multiply(E_Ai).mod(nSquared);

        E_B = E_B.multiply(E_Bi).mod(nSquared);

    }

    public BigInteger getE_A(){
        BigInteger ret = E_A;
        E_A = BigInteger.ONE;
         return ret;
    }

    public BigInteger getE_B(){
        BigInteger ret = E_B;
        E_B = BigInteger.ONE;
         return ret;
    }
}
```

A.6 Residential User

```
import java.math.BigInteger;
import java.security.NoSuchAlgorithmException;
import java.util.Random;
public class ResidentialUser {

    private BigInteger S;
    ResidentialGateway gw;
    Pub pubObject;
    int Tp=0;

    public void initialize(ResidentialGateway gateway,Pub pub, BigInteger s){
        this.gw = gateway;
        this.pubObject = pub;
        this.S = s;
    }

    private int getM(){
        int mi = new Random().nextInt(SystemModel.alpha);
        return mi;
    }



    public BigInteger[] reportGeneration() throws NoSuchAlgorithmException{
        int mi= getM();
        int Ai = mi % SystemModel.delta;
        int Bi = (mi-Ai) / SystemModel.delta;
        int Tp = this.Tp++;
```

```
        BigInteger nSquared = pubObject.n.pow(2);
        BigInteger aMulAi = pubObject.a[Bi].multiply(BigInteger.valueOf(Ai)) ;
        BigInteger E_Ai = pubObject.g.modPow(aMulAi,nSquared).multiply(
            ↪ TrustAuthority.hashValues(String.valueOf(Tp),
               TrustAuthority.H1, pubObject.n).modPow(S,nSquared)).mod(nSquared
                   ↪ );
        BigInteger E_Bi = pubObject.g.modPow(pubObject.a[Bi],nSquared).multiply
            ↪ (
               TrustAuthority.hashValues(String.valueOf(Tp), TrustAuthority.H2,
                   ↪  pubObject.n).
                     modPow(S,nSquared)).mod(nSquared);
        return new BigInteger[]{E_Ai, E_Bi};
    }
}
```

Vita

Candidate's full name: Kingsley Kwame Baah Larbi
Bachelor of Science in Computer Science, Kwame Nkrumah University of Science and Technology, Kumasi, 2014
Master of Science in Computer Science, University of New Brunswick, Fredericton, Canada, 2020

Publication:

[1] K. Larbi and R. Lu, Achieving Continuous Privacy-Preserving Histogram Query in Smart Grid Communications, in preparation.