A Framework for Privacy-Preserving Data Sharing in Smart Grid

by

Khalid Alharbi

A thesis
presented to the University of Ontario Institute of Technology
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Computer Science

University of Ontario Institute of Technology
Oshawa, Ontario, Canada, 2016

I hereby declare that I am the sole author of this thesis. Any published (or unpublished) ideas and/or techniques from the work of others are fully acknowledged in accordance with the standard referencing practices.

I understand that my thesis may be made electronically available to the public.

Khalid Alharbi
2016

Abstract

While smart grid introduces a lot of enhancements to the traditional power grid and improves managing and controlling consumers demands, it also introduces security and privacy issues. Therefore, failure to address them will hinder the flourish of smart grid.

In this thesis, we propose a novel framework for privacy-preserving data sharing in smart grid using a combination of homomorphic encryption and proxy re-encryption. The proposed framework allows distributed energy resources to be able to analyze the consumers data while preserving the consumers privacy. To the best of our knowledge, the proposed framework is first attempt to consider an important problem concerning data sharing in smart grid. Furthermore, in order to effectively collect consumer (or household) electricity consumption data, we also propose an efficient lightweight privacy- preserving data aggregation scheme, called ELPDA, for smart grid. The proposed scheme aims at resolving the power consumption data security and residential consumer privacy by employing one-time masking technique to protect consumers privacy while achieving lightweight data aggregation.

Moreover, we study the situation in which gateways aggregating consumers' data become malicious. Then, we propose a security-enhanced data aggregation scheme for smart grid communications from a homomorphic cryptosystem, trapdoor hash functions and homomorphic authenticators. The distinctive feature of our scheme achieves data confidentiality and integrity against the malicious aggregator (e.g. gateway), meaning that the aggregator is not able to learn the privacy of users or corrupt the power consumption reports during the aggregation process.

In addition to the above schemes for smart grid upnlink communications, we propose an efficient and privacy-preserving scheme in order to protect smart grid in downlink communications. Specifically, we propose an efficient identity based signcryption, called EIBSC, providing privacy preservation in downlink communication for smart grids. The proposed scheme is characterized by employing the concealing destination technique on the tree-based network to protect consumer privacy in downlink communication. Furthermore, the proposed scheme employs identity based signcryption to efficiently achieve downlink message source authentication, data integrity and confidentiality. Additionally, compared to other identity-based signcryption schemes, the proposed scheme is more efficient in regards to computational overhead and ciphertext size. Furthermore, security analysis demonstrates that the proposed scheme is resilient against various security threats to smart grids.

Acknowledgements

I would like to thank all the people who made this possible. This thesis would not have been possible without the help and support of my supervisor, my thesis committee members, and my colleagues in Information Forensics and Security Laboratory (IFS Lab). During my PhD research I learned many new things, and without the people surrounding me I could not enjoy from this period of my life.

First of all, I gratefully acknowledge my supervisor, Professor Xiaodong Lin. He made available his support and aid in a number of ways. He always does care about his students, and I had this opportunity to discuss the obstacles encountered me in my study and research openly with him. He not only helps me to develop the academic skills, but also guides me to strive for excellence. I would also like to thank Prof. Kui Ren from University at Buffalo, The State University of New York for serving as my thesis external examiner and sharing his invaluable insight on computer and communication security with me. I would also like to extend my appreciation to the other members of my examining committee, Dr. Ying Zhu, Dr. Jing Ren, Dr. Atef Mohany, and Dr. Shahram Heydar, for the time and efforts to read my thesis. In spite of their busy schedules, all have been readily available for advice, reading and encouragement.

Grateful acknowledgements are made for Northern Border University and the Saudi Arabian Cultural Bureau in Ottawa for providing me with a full scholarship during my Ph.D study.

I would never get this far without the support of my father, uncles, and brothers. I think them for always believing in me and supporting me. Their love and encouragement have been and will always be a great source of inspiration in my life.

Finally, my special thanks go to my wife, sons and daughter for the loving support and patience they have for me to fulfill my career goals.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Smart grid is a next generation power system that delivers electricity from the utility to its customers. Essentially, the concept of smart grid consolidates various technologies, i.e., advanced sensing, remotely control centers, information and communication technologies, into traditional power systems. In smart grid, through a variety of sensors and smart electric meters installed on the power grid particularly on each household, the power grid can be monitored effectively to be more reliable and electricity companies can also control energy consumption through real-time pricing, especially, higher prices at peak times due to higher demand. Furthermore, consumers can benefit from it, for example, reducing their electricity bills by lowering their power consumption at peak times. Despite the advantages of smart grid for both power companies and its customers, security and privacy are still critical challenging issues in smart grid. Failure to address them will hinder the flourish of smart grid.

While there are different components and technologies in smart grid systems such as data networks, advanced metering infrastructure (AMI) and supervisory control and data acquisition (SCADA) systems, we could expect various attacks in these components. For instance, it has been reported that there was a power outage affected around 50 million people in eight American states and Ontario, respectively [1, 2]. The reason for the power outage was a race condition software vulnerability in one of the component in the UNIX-based energy management system. Other design flaws within an unnamed smart meter have shown that an attacker can take control about 15,000 of home smart meters out of 22,000 in a simulation environment in one day as shown by Mike Davis in the 2009 Black Hat conference [3]. Furthermore, another attack against energy and global oil named Night Dragon by McAfee allowed attackers to compromise the

company's intranet and extranet by exploiting SQL injection attacks and uploading several tools that compromise web servers [4]. Therefore, it is crucial to secure smart grid systems.

In smart grid, managing remotely-located components and monitoring the power grid systems in a timely and effectively manner can prevent overloading in the power grid from happening. For instance, a power outage in a small area if not real-timely monitored can be the cause of overloading the power grid and what happened in 2003 in US and part of Canada is a perfect example of a cascading failure that caused around 50 million people to live in the dark for days [5].

Accordingly, consumers and neighborhood privacy need to be protected in the aggregation of consumers consumption data by employing various cryptographic protocols and technologies. The strategy for detecting and monitoring the power grid requires aggregation of significant data from various distributed energy resources and consumers, including consumption of power. However, consumers may be reluctant to contribute their data if the privacy is not protected. Thus, applying some cryptographic mechanisms such as symmetric/asymmetric cryptographic protocols is considered one of the effective means used for secure communication and privacy-preserving while achieving data aggregation. As a result, adopting classical cryptographic solutions might be not applicable and be infeasible in low-cost smart grid devices or smart meters due to several factors of speed and sophisticated encryption/decryption operations used in the deployed components and distributed energy resources. Hence, a combination of various encryption protocols and lightweight cryptographic techniques in particular can help in security and consumers privacy.

Therefore, the main motivation of this study is to propose a new framework for privacy-preserving data sharing in smart grid shown in Fig 1.1, as well as privacy-preserving data aggregation scheme for smart grid, which have some unique properties to defend against various security and privacy issues when collecting consumers' power consumption data.

The proposed scheme uses a combination of cryptographic protocols i.e., a homomorphic encryption and proxy re-encryption techniques in addition to bilinear pairing as the bases of the proposed scheme. The proposed schemes help in protecting user privacy and satisfying security requirements in smart grid.

1.2 Objective and Contributions

Our objective is to design a privacy-preserving data sharing framework for smart grid, including an efficient lightweight privacy-preserving data aggregation scheme aiming at resolving the power usage data security and residential consumer privacy. Furthermore, a security-enhanced

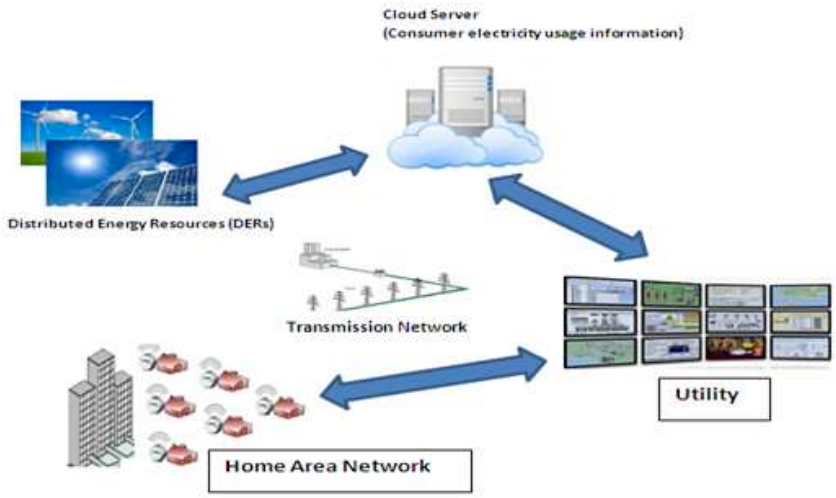

Figure 1.1: Smart grid system

data aggregation scheme based on a homomorphic cryptosystem, trapdoor hash functions and homomorphic authenticators is proposed in order to achieve data confidentiality and integrity against the malicious aggregator (e.g. gateway). An efficient and privacy-preserving downlink communication scheme for smart grid based on identity-based signcryption is also proposed in order to protect consumers privacy. The main contributions of this thesis include:

- We propose a novel data sharing framework for the smart grid [6], where we combine the two popular infrastructure: the smart grid and cloud computing together. In particular, we allow the electricity consumption reports generated in the smart grid to be stored in the cloud, and the distributed energy resources can obtain the statistics and analysis results from the cloud computing. Hence, our proposed framework can take advantage of cloud computing for the smart grid. The proposed framework makes use of the homomorphic encryption technique to facilitate the statistics and analysis on the encrypted electricity consumption reports, and the proxy re-encryption technique to keep the statistics and analysis results secret from the cloud.

- By considering residential user privacy and efficiency issues in data aggregation in a Residential Area Network (RAN) of smart meter devices, we propose ELPDA, an efficient lightweight privacy-preserving data aggregation scheme to address security and privacy challenges [7]. In ELPDA, based on one-time masking technique, each smart meter's data can be efficiently encrypted and aggregated. Compared with popular Paillier Cryptosystem based aggregation (PCBA) algorithm applied in smart grid [8–10], the proposed ELPDA

is much more efficient, reducing the aggregation delay in the whole RAN. Extensive simulations demonstrate the efficiency of the proposed ELPDA scheme. ELPDA outperforms the PCBA algorithms in terms of average aggregation delay in smart grid.

- Inspired by the facts that gateways may be corrupted, we present a security-enhanced data aggregation scheme from trapdoor hash functions, Pailliar encryption and homomorphic authenticators [11]. To the best of our knowledge our proposed scheme is the first one against malicious gateways and a successful attempt to construct authentication schemes from trapdoor hash functions with key exposure. Security analysis id given to show the proposed scheme is secure. Detailed performance analysis demonstrates that the proposed scheme is indeed significantly more efficient than the existing schemes in terms of both communication and computational overheads.

- We also propose an identity-based signcryption scheme implemented concealing technique for downlink communication in smart gird in order to protect consumer privacy, authenticity and data integrity [12]. The proposed scheme (EIBSC) is based on a tree-based network in which downlink communication can be more efficient using minimum spanning trees and privacy preservation is provided using the concealing destination technique. The proposed scheme (EIBSC) is much more efficient in terms of computational costs and ciphertext size compared to other signcryption schemes and outperforms existing competing schemes.

1.3 Organization of Thesis

This thesis is organized as follows. Chapter 2 introduces an overview on smart grid, associated security issues, applied cryptography to smart grid and related work. In Chapter 3, we discuss an efficient lightweight privacy- preserving data aggregation scheme for smart grid as well as algorithms used in a data aggregation tree and hop-by-hop authentication and forwarding, followed by security analysis and evaluation. In Chapter 4, we present a security-enhanced data aggregation scheme against malicious gateways for smart grid and provide security analysis and performance evaluation. In Chapter 5, we present a framework for privacy-preserving data sharing in smart grid, followed by security analysis and evaluation. Chapter 6 discusses an efficient and privacy-preserving smart grid downlink communication scheme based on identity based signcryption as well as the concealing destination technique, followed by a comparison to other identity-based signcryption schemes. In Chapter 7, we draw our conclusion and future work.

Chapter 2

Background and Literature Review

2.1 An Overview of Smart Grid

Smart grid is the next generation power system that delivers electricity from the utility to its customers. Smart grid is more efficient and reliable electricity system than existing power systems. Essentially, smart grid integrates the network of generation, transmission, distribution lines and advanced sensing technologies, into traditional power systems as shown in Fig. 2.1.

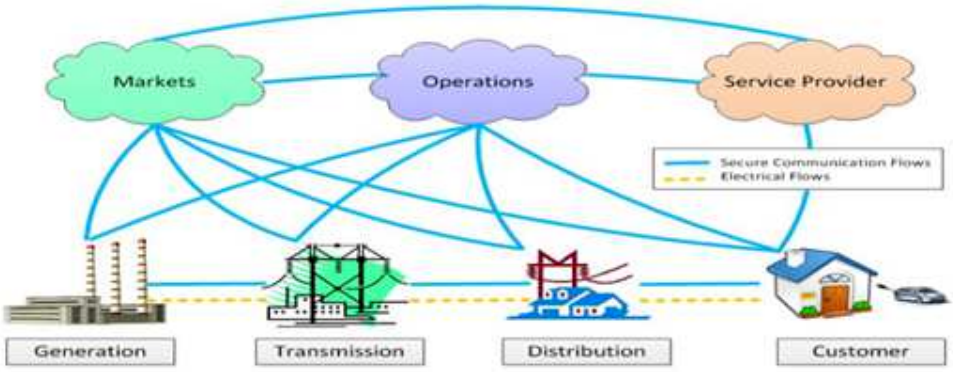

Figure 2.1: The conceptual smart grid: next generation power grid equipped with advanced sensing, control, information and communication technologies (Courtesy of [13])

The main difference between smart grid and the traditional power grid is the information flows. In the traditional power grid, there only exist the one-way electrical flows, i.e., electricity utilities only deliver power to consumers. While in smart grid, there additionally exists a

two-way information flow communication. As shown in Fig. 2.2, in smart grid, the two-way information flows are almost parallel to the one-way power flows, except that the control center is also involved in the information flows. The control center can adjust the power grid, such as the amount of power generation in the next period, based on the statistics and analysis of electricity consumption reports from the consumers. It is fair to say that the collected electricity consumption reports are the key to allow the smart grid to be truly smart.

There are a lot of benefits associated with smart grid for both consumers and power utilities [14], including the efficient transmission of power, quick restoration of power after serious power disturbances, reduction of peak demand and giving consumers control. The latter is very important in terms of saving consumers money by giving consumers information and tools that enable them to make decisions about their power use. Also, consumers in smart grid participate more than in traditional power systems due to the two-way data communications of smart grid for control and monitoring. More precisely, consumers no longer wait for their monthly bill they can view their power consumption in a timely manner and accurate data at will.

2.2 Smart Grid Security and Privacy

As mentioned earlier, smart grid is composed of various technologies, components, and a wide range of distributed energy resources, i.e., renewable energy and electric vehicles. Of these, advanced metering infrastructure (AMI) is a vital subsystem of the smart grid. Smart meters are one of critical components of AMI. Smart meters are responsible for reading consumer energy usages and send them to collectors before they reach the utility center. In fact, consumers energy data sent through smart meters is vulnerable to be intercepted or overhearing by an adversary in various ways. For example, the consumer usage transmitted over a wireless communication channel between a smart meter and utility could be monitored by another party. Not only attackers can take advantage of violating consumer's data, but also third parties that are not under the control of the government might misuse consumer privacy especially in the distributed energy resources. In addition, smart meters data could be read and compromised before it is being sent to the utility disclosing sensitive information such as consumers' activities and lifestyles which might be sold to other third parties violating customers privacy [15]. For example, privacy is not violated due to disclosing sensitive data to unwanted entities, but also sniffing and analyzing collected data from smart meters among several sessions can detect and reveal the presence of people at their homes [16], which might be used for burglaries in the worst case.

Another indirectly privacy violation is when utilities provide third parties with information from smart meters beyond billing and their ordinary supplying purposes. Consequently, this information could be exploited by the third parties for detecting and extracting power signatures

of intelligent devices and types of appliances being used in homes, and then they sell this information to appliances companies in order to make any profit. Later on we will show why and how third-party companies cannot also be treated as trustworthy entities. Obviously, this is a good example on how third-party companies use consumers' power consumption for commercial purposes.

On the other hand, intercepting and manipulating consumers' power readings sent by smart meters are not necessary to violate privacy and bad for malicious consumers. For example, malicious consumers can manipulate their power usage or fabricate collected data using different tools that modify their usages, especially in the absence of cryptographic protocols or in the presence of employing weak encryption protocols. Accordingly, consumer scams and fraud in smart grid can cause providers lots of losses; it has been estimated to $6 billion just in the US alone [17]. Clearly, the protection of power consumption in data aggregation in smart grid is a must not only for consumers privacy protection, but also for the utilities' and service providers' rights.

The collected electricity consumption reports should be of encryption form when they are transmitted over the smart grid since they are the part of consumers privacy [11]. Otherwise, security and privacy concerns from electricity customers could become a major barrier to adopting such a great technology, which can make our power grid smarter and more reliable, resilient, and environmentally friendly. When there only an energy source exists, this problem can be handled by using traditional encryption schemes, since the only one control center is usually controlled by the government and it can be considered as a trusted party. However, the real situation is not so simple, especially the distributed energy resources, featured with small-scale power generation technologies and renewable energy sources, have been considered as a necessary supplement for smart grid [18, 19]. It is also known as "micro-grid".

Furthermore, not all the distributed energy resources are under the control of the government, hence it is not acceptable that all of them are trusted. Now, we are in a dilemma that the electricity consumption reports should be analyzed by the energy resources while the individual data can be still protected. One trivial solution is to anonymize the data before being sent to the energy resources for analysis. However, it will significantly increase the communication cost since the massive anonymized data should be sent to every resource and there could be also a lot of resources in an area. A possible improvement is to let some third party to do the analysis instead of the energy resources, and only the analysis results are sent to the energy resources. Nevertheless, the third party would also know the analysis results, which is not desired for the energy resources due to some business reasons. To the best of our knowledge, there is no efficient approach so far to this problem in the context of privacy-preserving smart grid.

Although smart grid can assist in transforming the traditional power industry, playing a piv-

otal role in maintaining high levels of reliability, efficiency, and manageability with two-way communications, it also introduces cyber vulnerabilities into the grid [20–23]. Failure to address these security problems will hinder the modernization of the aging power grid. For example, it is reported that an attacker only with $500 in equipment and a basic electrical background could seize command of smart grid's two-way communication system for sabotage. Once the two-way communication system were penetrated, the attacker could cause a blackout by either gaining control of possibly millions of meters on the grid and simultaneously shutting them down or disrupting the load balance of the local system by suddenly decreasing or increasing the demand for power [24].

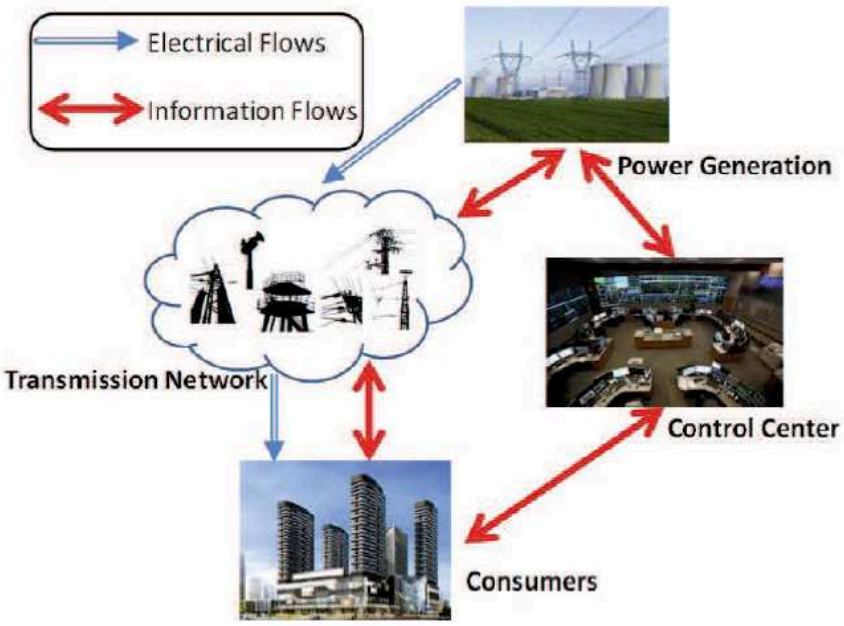

Figure 2.2: Communication architecture for the smart grid

2.3 Applied Cryptography to Smart Grid

In this section, we give an overview of some cryptographic techniques widely used in many standards, technologies and academic work recommended for smart grid. As a matter of fact, in smart grid different family cryptographic protocols have solved significant security emerging issues, as well as privacy protection. For instance, some cryptographic protocols are committed to providing mechanisms to protect consumers privacy while achieving data aggregation in smart

meters. Apart from consumer privacy preservation, a combination of cryptographic protocols can help in solving other security and privacy issues. For example, we can combine a homomorphic encryption and proxy re-encryption techniques in order to protect consumers privacy and collected data among distributed energy resources.

In fact, we take advantage of some specific cryptographic techniques such as symmetric/asymmetric cryptosystems, proxy re-encryption and homomorphic cryptosystem to achieve privacy-preserving data aggregation and privacy-preserving data sharing. In the following subsections we will introduce the basic concepts of the aforementioned techniques used in this study and they will be examined in more detail in several chapters later.

2.3.1 Symmetric and Asymmetric Cryptosystems

Cryptographic systems can be classified based on several criteria [25], but are not limited, the type of operations employed on transforming inputs into ciphertext, the way in which inputs are processed, and the number of keys used. Among these criteria, we are interested in the latter, namely encryption keys. The system that uses the same key i.e., for encryption and decryption is referred to as symmetric-key cryptography, while public-key or asymmetric is used to refer to the system applying different keys e.g., one key is for encryption and the other is for decryption. Both of them have advantages and disadvantages and they together can have a number of complementary advantages. Menezes *et al.* [26] outline some advantages and disadvantages of employing symmetric and asymmetric cryptography, and compare between them in terms of 1) key length as an important security parameter, 2) being used to build hash functions, 3) computational performance, and 5) digital signature requirements.

In addition to that, we consider the aforementioned cryptographic protocols implemented in smart grid in regard to security and privacy, network load, system robustness and scalability, and the total computation. In smart grid, various schemes have been proposed on the base of asymmetric encryption for a variety of security purposes e.g., authentication, batch verification and privacy-preserving schemes. Most of them are based on RSA, Diffie-Hellman key exchange, and Paillier cryptosystems. The Paillier cryptosystem is important to the current research due to the additive homomorphic encryption property over which computations such as multiplication on encrypted data can be carried out.

2.3.2 Proxy Re-encryption

The main goal of proxy re-encryption is to securely enable the re-encryption of cipher-texts (e.g., encrypted power consumption) under one public key to another public key, without relying on a

trusted party. It allows distributed energy resources to obtain the statistics and analysis without revealing the private key of the trusted authority (TA) to the distributed energy resources. For example, let assume Alice want to delegate her email to Bob, but she does not want him to know her secret key used for decrypting messages. This is where proxy re-encryption helps Alice in this case and we will use Blaze's *et al.* scheme [27] to solve Alice's email delegation. Blaze's *et al.* scheme is based on the ElGamal cryptosystem, and uses four algorithms as follows:

1) Key Generation

Let $\mathbb{G}$ be a cyclic group of order p, and g be a generator of $\mathbb{G}$. Alice chooses an x randomly as her private key where $x \in [1, p-1]$. Alice's public key is g^x. Similarly, Bob's private key is y and his public key is g^y. The re-encryption key $RK_{A \to B}$ is $y/x = y.x^{-1}(\mathrm{mod} p)$, where x^{-1} is the inverse of x.

2) Encryption

Let m be a message $\in \mathbb{G}$, and r is selected randomly from $\mathbb{Z}_p^*$. Any message to Alice can be encrypted using her public key g^x and r as in the following formula:

$$Enc(m, r, g^x) = (g^r.m, g^{xr}).$$

3) Decryption

To decrypt a message given $(g^r.m, g^{xr})$, we use the following formula:

$$m = \frac{g^r.m}{(g^{xr})^{\frac{1}{x}}}$$

4) Proxy Re-encryption

When the proxy re-encryption algorithm receives encrypted messages to Alice, it re-encrypts them using the re-encryption key $RK_{A \to B}$, and then sends them to Bob. The following table illustrates the proxy re-encryption process.

Table 2.1: Proxy re-encryption algorithm

Alice	**Proxy**	**Bob**
$(g^r.m, g^{xr})$	$(g^r.m, (g^{xr})^{RK_{A\to B}})$ $= (g^r.m, (g^{xr})^{RK_{A\to B}})$ $= (g^r.m, (g^{xr})^{y/x})$ $= (g^r.m, (g^{yr}))$	$(g^r.m, g^{yr})$

2.3.3 Homomorphic Encryption

Homomorphic encryption is a unique form of encryption that allows doing arithmetic (i.e., multiplication or addition, or even both) on ciphertext values, then decrypting the result will give the same result when doing arithmetic on plaintext values. Moreover, the homomorphic encryption technique is used to facilitate the statistics and analysis on the encrypted power consumption reports. For example, given n consumers and their power consumption $(m_1, \cdots, m_n)$ where m_1 denotes power consumption of consumer 1 and m_2 denotes power consumption of consumer 2, and so forth. If we encrypt $m_1, \cdots, m_n$, we get $c_1, \cdots, c_n$ as corresponding encrypted values, respectively. Accordingly, the Paillier cryptosystem, as a homomorphic encryption,has the following desirable property:

$$\sum_{k=1}^{n} m_k \equiv DEC_{sk}(\prod_{j=1}^{n} c_j (\mathrm{mod} p)),$$

where DEC is the decryption algorithm and sk is the private key.In other words, the decryption of multiplied encrypted values is equal to the sum of plaintext values. By doing so, we calculate the sum of power consumption from multiplication of the encrypted power values.

2.3.4 Identity Based Encryption (IBE)

An identity based encryption (IBE) technology is a type of public-key systems, which allows a user or entity to calculate their or other public key from a unique information such as an email address or a serial number of a device. An IBE system shares several aspects with traditional public-key systems, however, it is quite different in others. While a traditional public-key system contains all information and parameters needed to use the system, a trusted third party(e.g., a private key generator called PKG) publishes a set of public parameters needed to use an IBE system. Particularly, an entity can use those public parameters to calculate other entities' public

key for encryption. Table 2.2 shows a comparison of similarities and differences of IBE and traditional public-key systems

The process of generating a public-private key pair in traditional public-key systems compared to an IBE system is another significant aspect. In a traditional public-key system ,most often a user or an application is used on behalf of to generate a public-private key pair. After they are created, the public key and the identity of the owner key are then digitally signed by a certificate authority (CA)to create a digital certificate [28]. Contrary to traditional public-key systems, a PKG in IBE systems calculates an IBE private key of an entity by using the entity's identity as well as a system parameter called a master secret. It is clear that an entity that uses IBE does not need to obtain a recipient's certificate when sending an encrypted message. As a result, IBE provides an unprecedented practical solution to several security issues related to public-key systems such as key distribution, key validation and certificate management, which are complex and difficult problems.

IBE, like other cryptographic techniques, has its own blemishes. For example, since a PKG is responsible for generating a master secret, which in turn can be used to generate an entity's private key, all public-private pairs generated by that PKG are compromised if the hosting server is compromised. This inherent problem is known as the key escrow which can be solved by implementing certificateless public-key cryptography [29]. Additionally, not all security goals can be supported by IBE [28]. For instance, while confidentiality, integrity,authentication and non-repudiation as security goals are provided in various public-key cryptosystems, IBE provides only confidentiality. Therefore, we need to modify IBE in order to support these security goals as well as privacy preservation for smart grid as we will see in Chapter 3 and Chapter 6.

2.3.5 Fuzzy Identity-based Encryption(Fuzzy-IBE)

In identity-based encryption (IBE) a public key can be any unique information about the entity's identity such as a user's email address, and a message is encrypted using the receiver's public key. The receiver then will use their IBE private key that associates to the IBE's public-key to decrypt messages. On the other hand, in the fuzzy identity-based encryption (fuzzy-IBE) [30] the user's identity ω can be defined as a set of attributes instead an atomic identity as described in IBE. This set of attributes is considered a user's public key. In a fuzzy-IBE scheme once a user's identity ω has been decided, the user's private key is computed and the user will be able to decrypt encrypted messages with their identity ω using their private key. A further attraction is the ability to decrypt encrypted messages with other's identity if ω and ω' are close to each other by minimal set overlap or if $\mid \omega \bigcap \omega' \mid \geq d$, where d is the error tolerance [30]. Table 2.3 shows a simple example on a fuzzy-IBE scheme and we have borrowed the main idea of the table from

Table 2.2: Comparison between IBE and traditional public-key systems

Comparing Aspect	**IBE**	**Traditional public-key systems**	**Notes**
Private key generation(SK)	Trusted authority	Entity	(1*)
Public key generation(PK)	Entity's identity	Trusted authority	(2*)
Key lifetime	Shorter	Longer	(3*)
Security Goals	Only confidentiality	Most security goals	(4*)
Certificate	Certificateless	Certificate-based	(5*)

Comment ID	**Description**
(1*)	IBE's SK is calculated by PKG master secret while SK in traditional systems is calculated by an entity and kept secret.
(2*)	IBE's PK is computed by using the entity's identity while in traditional systems PK is computed from a private key (e.g., a user or agent).
(3*)	IBE's keys are usually valid for short periods while in traditional systems are typically valid for long periods.
(4*)	Most security goals are provided in traditional systems while in IBE only confidentiality is provided.
(5*)	In many traditional public-key systems the binding between the entity's identity and the public key is done via a digital signature to create a digital certificate. As a result, the drawback of certificate-based public-key systems is certificate management. On the other hand, IBE systems avoid using certificates and simplify certificate management, but still have the key escrow problem as a major drawback of IBE systems.

Prosunjit Biswas [31] .

On the contrary, in attribute-based encryption or ABE [32] an identity ω can also be viewed as a set of descriptive attributes as in the fuzzy-IBE scheme, but different in several aspects. First, in ABE a user's private key is computed after deciding key policy that is decided from the user's attributes. Second, while in the fuzzy-IBE scheme the encryption is performed based on the user's identity ω, ABE performs encryption on a set of attributes γ.

Another important aspect of the fuzzy-IBE scheme is that users can decrypt encrypted messages with their private keys while in ABE messages can be decrypted either if a user's key policy has been satisfied with γ, that is, $\Gamma(\gamma) = 1$, or polices defined over a set or subset of attributes have been satisfied (see Goyal *et al.* [32] for more details). Potential applications for fuzzy-IBE and ABE can be seen clearly in the access structures due to nice features inherent in fuzzy-IBE and ABE, for example, while in some traditional identity-based encryption schemes it is essential to compute a user's privet key before data has been encrypted, in ABE that is not necessarily essential because only users who satisfy an associated policy are able to decrypt even thought their private key has not been created yet. In other words, the associated policy is related to a set of attributes not to a set of users.

Table 2.3: Simple example of a fuzzy-IBE scheme

Name	**Fuzzy- identity**	**Error Tolerance or** d	**Comments**
Dan	ω = {'UOIT', 'Instructor', 'Computer science'}	2	Dan can decrypt encrypted messages sent by others. Also he can decrypt any encrypted message to Sam and Chris because $\mid \omega \bigcap \omega' \mid \geq 2$ and $\mid \omega \bigcap \omega'' \mid \geq 2$.
Sam	ω' = {'UOIT', 'Student', 'International','Computer science'}	3	Sam can only decrypt encrypted messages sent by others. Also he can decrypt any encrypted message to Chris, $\mid \omega' \bigcap \omega'' \mid \geq 3$, but not to Dan because $\mid \omega' \bigcap \omega \mid = 2$.
Chris	ω'' = {'UOIT', 'Student','Post Grad', 'Computer science','Canadian'}	5	Chris can decrypt encrypted messages sent by others to himself, but not others' messages.

2.4 Related Work

Various researchers have tried to tackle the challenges arising out of security requirements, legal issues, and technologies in smart grid. Confidentiality, integrity and availability specified by [8, 33], as high-level security requirements, are the most attractive topics in smart grid. Other researchers have tried to address the challenges of aggregating encrypted data, and protecting consumers privacy while achieving data aggregation in smart grid. Therefore, the main focus of related work will be on security and privacy, and existing approaches in smart grid.

Recently, there have appeared several research works on data aggregation in smart grid [9,10] which are closely related to our proposed framework. Saputro and Akkaya [9] investigate the overhead of using homomorphic encryption in smart grid. In specific, they compare the latency and data size of end-to-end and hop-by-hop homomorphic encryption within a network of smart meters.

Lu *et al.* [10] propose an efficient and privacy-reserving aggregation scheme, named EPPA, for smart grid communications, which uses a super-increasing sequence to structure multi-dimensional data and encrypt the structured data by the homomorphic Paillier cryptosystem technique [34], and has significantly less computation and communication overhead during the aggregation process. Although the above works are promising, all of them belong to the Pailliar Cryptosystem based Aggregation (PCBA). Different form the above works, the proposed ELPDA does not rely on Pailliar Cryptosystem, but a lightweight one-time masking, which makes it particularly efficient, as shown in Fig. 3.7.

Several authors [8,35,36] propose employing homomorphic encryption schemes for protecting encrypted data while aggregation and performing analysis. These proposed protocols are based on the Paillier cryptosystem [34].

Li *et al.* [8] propose a homomorphic encryption technique as well as an in-network data aggregation scheme for smart grids. In their scheme, an aggregation tree is constructed to achieve better aggregation in terms of communication and computational cost saving and breadth-first traversal of the graph is used to short the height of the aggregation tree by reducing the maximum hops for the longest path. Particularly, the aggregation is performed in a distributed manner in accordance with the aggregation tree, where each node collects data from its children, aggregates them with its own data, and sends the intermediate result to the parent node. Homomorphic encryption is employed to protect the privacy of the electricity use data, so that inputs and intermediate results are not revealed to smart meters on the aggregation path, while the aggregation is still correctly performed. Despite the fact that the proposed scheme achieves the privacy goal as participating smart meters cannot retrieve meaningful information from the consumption of the others, it has not addressed any failure that can happen to some parent so close to the collector

at level-1of the aggregation tree. Hence, the close parent receives data from its children but not able to send them to the collector unit.

Erkin *et al.* [35] propose a homomorphic encryption protocol based on a modified version of the Paillier cryptosystem, and on sharing secret keys. Essentially, the modification allows them to propose three schemes for computing the power usage of an individual, a set of consumers at a specific time and interval, respectively. The third scheme is used for the total amount of electricity in a neighborhood. While the proposed scheme theoretically preserves privacy in aggregation, it still practically requires a lot of computations for performing homomorphic operations, randomization, and using hash functions, especially when compared Li's *et al.* scheme [8]. Also, their proposed protocol requires secret sharing that might be another overhead cost in terms of the calculation of users consumption, and the number of operations.

Garcia and Jacobs [36] propose the No-Leakage Protocol used for achieving security and privacy goals, especially for aggregated consumption in a neighborhood or block. The proposed protocol combines secret sharing and an additive property in Paillier homomorphic cryptosystem for computing aggregated energy consumptions in smart meters. Concretely, this protocol requires the existence of a Trusted Certification Authority (TCA) which is responsible for issuing public key certificates for smart meters and the collector that sends issued certs to all smart meters in a neighborhood. Also, each message or consumers consumption data are split into shares, where the sum of shares is equal to the measure of a smart meter. Then, each smart meter encrypts its shares with other's public key in a smart meter network, except its own public key and for one share that is kept without encryption to be a secret and used later in homomorphic encryption. The encrypted shares are sent to the collector which in turn, accumulates encrypted data intended to each smart meter and then multiplies accumulated cipher texts. Since their protocol employs the Paillier homomorphic cryptosystem, the result of the product of cipher texts is sent to each smart meter in accordance with the cumulated cipher texts.

Accordingly, each smart meter will decrypt the received cumulated ciphertext and add up the unencrypted share to the one just decrypted giving the total (not the total consumption) in plain text. Therefore, the total consumption is given when all smart meters send its decrypted total to the collector in which smart meters data are aggregated in clear text. However, this protocol suffers from security and performance issues. The most significant issue is related to the total network load, concretely, each smart meter participates in aggregation has to know all smart meters' public keys, as well as the collector's to establish a connection with the collector. As a result, storing several keys, most of which are redundant, is not efficient in terms of aggregation, and not preferable regarding network performance. Moreover, the collector needs to send (n certs) to all smart meters involved in aggregation, where (n) denotes the number of smart meters. In addition, every smart meter will send $n-1$ messages to the collector, precisely; the collector will handle $n(n-1)$ messages simultaneously.

Later, Li *et al.* [37] combined homomorphic encryption and key evolution technique to propose a privacy-preserving demand response scheme, which supports adaptive key evolution and forward secrecy of users' session keys. Consequently, to prevent users' privacy from being disclosed to internal attackers (e.g., electricity suppliers), Fan *et al.* [38] proposed a privacy-enhanced data aggregation scheme by injecting blinding factors into the data. Unfortunately, the aggregated data lay a big obstacle on the billing, since the individual consumption data are required to compute electricity bills. To solve billing issues, Ohara *et al.* [39] adopted commitment technique and lifted Elgamal encryption as tools. It requires the aggregator to compute a sum of real-time power consumption over a time period and a sum of electricity consumption in a RA at a time unit, respectively.

The data aggregation schemes from distributed random noises enjoy prominent advantages on computational performance. Dimitriou [40] proposed a secure and scalable aggregation scheme for fine-gained consumption monitoring, which allows users to securely aggregate their measurements in a way that preserves their privacy against honest-but-curious adversaries. Lin *et al.* [41] designed a smart metering system that simultaneously supports billing and road monitoring applications with an embedded trusted platform module chip for generating random values. This scheme is based on the layered meter model, and provides an efficient and feasible solution to privacy-preserving smart meter systems.

Unfortunately, all the aforementioned schemes except [39] are under the assumption that the aggregator (e.g. gateway) executes the protocol correctly but tries to learn as much as possible. This means that these schemes can no longer be secure if the aggregator (e.g. gateway) is compromised. In [39], the scheme only can aggregate the measurements, and thus the communication overhead between the gateway and the operation center is still linear with the number of the users. Therefore, it can only be viewed as a semi-aggregation scheme.

Kursawe *et al.* [42] propose two protocols to compute the total power consumption in smart meters network with computation and hardware constraints. The first protocol, called aggregation protocols, is where smart meters' data are masked in such a way that an adversary cannot retrieve meters' data, and the sum of the masking values is set to a zero or any other value. More precisely, inputs from all meters are collected and summed, while the masking values cancel out each other i.e., set to zero and the collector can obtain the total consumption without knowing user's usages of power and without disclosing consumer sensitive information. For example, let $c_{i,j}$ be the consumption of meter (j) with a reading (i), as well as $x_{i,j}$ is the mask being used. Hence, the output of employing the aforementioned mask is $x_{i,j} + c_{i,j}$. After aggregation, the collector adds up the received masking values, thus, the masking values would cancel out each other giving $\sum c_{i,j}$. The latter, named comparison protocols, is based on an assumption that the collector knows the approximate sum of meters readings.

Ye *et al.* [43] studied communication security for AMI downlink transmission in smart grid. They propose a hierarchical identity-based signature scheme (HIBaSS) for the downlink transmission in smart grid in order to help in integrity and the authenticity of control message in the downlink transmission, that is, from the collector/data aggregate point (DAP) to a smart meter. The proposed scheme suggests that every smart meter receives a group signature from all data aggregate points with a certificate from the authentication server (AS) to prevent forgery or repudiation. More importantly, in the scheme smart meters do not trust any signature from DAP even though it is the original one due to the absence of directly connection between the AS and smart meters.

Our proposed privacy-preserving data sharing framework is also related to "access control of outsourced data". Hence, we will also review some works in this field. Yu *et al.* [44] proposed a secure, scalable and fine-grained data access control system on the cloud server by using key-policy attribute-based encryption, proxy re-encryption, and lazy re-encryption. Wang *et al.* [45] propose a searchable and privacy-preserving data access control system over outsourced cloud data by using searchable encryption. Recently, Shao *et al.* [46] propose a secure, scalable, searchable and fine-grained data access control system in cloud computing for mobile devices by using ciphertext-policy anonymous attribute-based encryption, proxy re-encryption, lazy re-encryption, and transformation key. However, none of the above systems supports evaluation on the cipher-texts.

In smart grid, collection and aggregation techniques have gained much attention in academic research due to its importance 1) for solving some networking issues (i.e., bandwidth, energy efficiency and computations costs) and 2) for achieving statistical information such as average, maximum and minimum values. In terms of solving networking issues, data aggregation algorithms are concerned not only with collecting data from nodes and sending them to a collector, but also with achieving that in a very efficient method such as the efficient shortest path or minimum spanning trees. In addition, in wireless sensor networks (WSNs) and some smart grids topology, it is unavoidable and necessary to adopt the decentralized data aggregation techniques when the flow of data in a network is only done by hop-by-hop transport, especially when data are collected and aggregated from far away nodes.

Aside from statistical information and network issues, data aggregation is promising to save communication and computational costs when collecting users' electricity consumption data in smart grid. Most data aggregation schemes such as [10, 51] consider a scenario where smart meters installed in homes and businesses communicate directly with a collector, which aggregates the consumption data received and then forward the results to the utility. Data transmitted wirelessly on route to the collector is protected and further aggregated so that individual user data can be protected and the privacy of individual users can be preserved. Lu *et al.* [10] propose an efficient and privacy preserving aggregation scheme, named EPPA, for smart grid communi-

cations, which uses a super-increasing sequence to structure multi-dimensional data and encrypt the structured data by the homomorphic Paillier cryptosystem technique, and has significantly less computation and communication overhead during the aggregation process. However, data are still transmitted directly from each smart meter to the collector. Zhang *et al.* [51] propose an efficient and privacy-preserving aggregation scheme with adaptive key management, named PARK, for smart grid communications, which uses an adaptive key management mechanism with effective revocation allowing automatically updating to users' encryption keys. PARK, in terms of performance, is significantly less costly during the aggregation process. Nevertheless, the traditional aggregation technique is still implemented in PARK.

While such kind of data aggregation works well in certain situation, for example, in neighborhoods or suburbs where houses are scattered over a big area and far away from each other, it may not be efficient in other situations, for example, a neighborhood in a highly dense city (Fig. 3.2(a)) or a high-rise apartment building (Fig. 3.2(b)). This is because it doesnt take into consideration geographic characteristics of communication networks formed by smart meters. To improve the effectiveness of data aggregation in smart grid, several approaches have been proposed [8, 52–54]. D. Li *et al.* [53] propose an efficient authentication scheme for power consumption data aggregation in Neighborhood Area Network (NAN). They constructed a spanning tree to be utilized for data aggregation. They also proposed a mechanism for fault tolerance in which the collector re-executes the proposed MST algorithm within itself and associates a faulty smart meter with a new parent. They assumed faulty meters have the ability to directly communicate to the collector. However, this approach has not addressed faulty smart meters that are far away from a collector and are not able to directly communicate to the collector.

F. Li *et al.* [8, 52], and Y. Lei *et al.* in [54] introduced the concept of aggregation tree in smart gird. They built a network model based on geographic locations of homes and businesses, and constructed a data aggregation tree that minimizes the total communication and computational costs for data collection in a smart grid. In an aggregation tree, each smart meter is considered as a tree node, and sends consumption data to a collector periodically by following a unique and shortest path. Data are aggregated when being forwarded to the collector in a multi- hop mode. As a result, the total cost of communication and computation of data collection can be minimized. This is particularly very useful in these neighborhoods in highly dense areas because of the following reasons: First, short-range wireless technologies, such as WiFi, can be used, and therefore, data transmission between smart meters can happen at free of cost. Second, communication and computational costs for data collection can be minimized.

However, tree-based data aggregation also introduces additional challenges for data aggregation and has incidentally introduced several network dilemmas into smart meter networks. The failure of a single smart meter can be catastrophic to the data collection or aggregation. For example, if a single smart meter fails, a neighborhood or some adjacent smart meters might not be

reachable (either the area contains faulty or intact meters) and as a result, they could be isolated from smart grid. In the worst of cases, the entire area will not be reachable if critical smart meters have failed. One of the reasons for unreachability and isolation in the data aggregation is the use of aggregation trees that enforces selecting a specific path (i.e., the least cost path) during the aggregation, thereby preventing some smart meters in the network from selecting another link as shown in Fig 3.5.

In spite of the fact that privacy preservation has been extensively studied in uplink communications, we also investigate privacy preservation for smart grid downlink communications. More precisely, we implement signcryption as a significant cryptographic primitive in which both encryption and digital signing on a message can simultaneously be performed [47]. This type of cryptosystems is intended to provide essential security services as well as reduce computational cost and communication overhead.

Recently, there have appeared several schemes on secure data and communication in smart grid, which are closely related to our proposed scheme. Hayden *et al.* [48] proposed a scheme based on identity-based signcryption cryptography in order to provide authenticity and confidentiality for end-to-end communications in uplink transmission. They employed Tate pairing and AES for authentication and encryption of data packet, respectively. However, their scheme has not addressed privacy preservation in the uplink and downlink transmission. Libert and Quisquater [49] presented an identity based signcryption scheme based on pairings over elliptic curves. Their scheme can provide both encryption and signature, and is provably secure in the random oracle model. Lal *et al.* [50] also proposed an identity based generalized signcryption scheme based on bilinear pairing. Their proposed scheme can provide ciphertext authentication and message confidentiality. However, the aforementioned two schemes have not addressed privacy preservation in communications. Moreover, since the proposed scheme is based pairings cryptography; our proposed scheme is particularly efficient in terms of privacy preservation, signcryption and computational overhead.

Chapter 3

Efficient Lightweight Privacy-preserving Data Aggregation Scheme for Smart Grid

In this chapter, we present a novel efficient lightweight privacy-preserving data aggregation scheme, called ELPDA, for improving security and privacy in smart grid. The proposed ELPDA integrates one-time masking technique with tree network topology to protect user power consumption data while achieving efficient lightweight data aggregation. However, with the introduction of tree network (or aggregation tree), a smart meter failure could cause catastrophic disruption of electricity consumption data collection and cause the smart meters that the faulty meter links to the collector to be disconnected. In order to address the issue, we further propose a faulty smart meter detection scheme to locate faulty smart meters, and then rebuild the aggregation tree, with removal of these faulty meters. Moreover, the proposed ELPDA resists various security and privacy threats, and preserves user privacy. Through extensive analysis, we demonstrate that the proposed scheme has significantly less computation and communication overhead compared with existing competing schemes.

3.1 Introduction

Smart grid is a next generation power system that delivers power from the utility to its customers. Essentially, smart grid integrates various technologies, i.e., advanced sensing, remote control centers, information and communication technologies, with traditional power systems. In smart grid, through various smart meters installed on houses or buildings, the power grid

can be monitored effectively to be more reliable and power companies can also control power consumption through real-time pricing, especially, higher prices at peak times due to higher demand from consumers. Furthermore, not only power utilities can benefit from smart grid, but also consumers can benefit from it, for example, reducing their monthly bills by shifting their power consumption to take advantage of the lower prices that apply during off-peak periods or weekends and statutory.

While smart grid introduces a lot of enhancements to the traditional power grid and improves managing and controlling consumers demands, it requires some changes to the components, communications and computational techniques used for data collection and aggregation and for security and privacy . For example, several data collection techniques have been proposed to automatically collect consumers' power consumption from smart meters and then send them to the utility or through an intermediate collector. Meanwhile, consumers and businesses can benefit from these techniques when they are given the control and tools to monitor their power consumption and choose the best times (e.g., off-peak) to use electricity and intelligent devices.

Fig. 3.1(A) shows several smart meters devices connected to one collector. In the traditional collecting scenario each smart meter sends directly its power readings to the utility. Since utility customers usually spread across a large area, a long-range cellular communication system is needed for both smart meters and the collector to communicate with each other. In doing so, it could be costly to both utility companies and customers. Another alternative is to use free short-range wireless communication technologies, such as Wi-Fi. As shown in Fig. 3.1(B), user electricity data have to be transmitted over multiple hops, and the same copy of an electricity data has to be forwarded many times along the path to the collector from the smart meter that the data is originating from. It becomes problematic when the residential density in an area is high, and precious wireless bandwidth is not used efficiently, and wasted on transmitting the duplicated electricity data. The traditional data collection approach is preferable in rural areas where the number of steakhouses i.e., consumers is fewer. Another important feature in the traditional collecting approach is that the whole power grid will not be affected and will still able to report directly their power consumption to the utility when one smart meter fails.

In spite of the advantages of traditional aggregation techniques, they can lead to high redundancy and communication load, and might be useless in some scenarios, especially, when a single smart meter that receives other aggregated results fails. As a result, the entire aggregated data will be affected. In addition, security issues such as confidentiality, integrity and authentication should be provided while data aggregation is performed, otherwise malicious users can pretend to be someone else' smart meter and get free power usage by exploiting the aggregation process Therefore, we need a very different efficiently mechanism for data aggregation in smart grid to be reliable and taking into user privacy preservation and security issues consideration. Therefore, different data aggregation schemes have been suggested to tackle aforementioned issues in

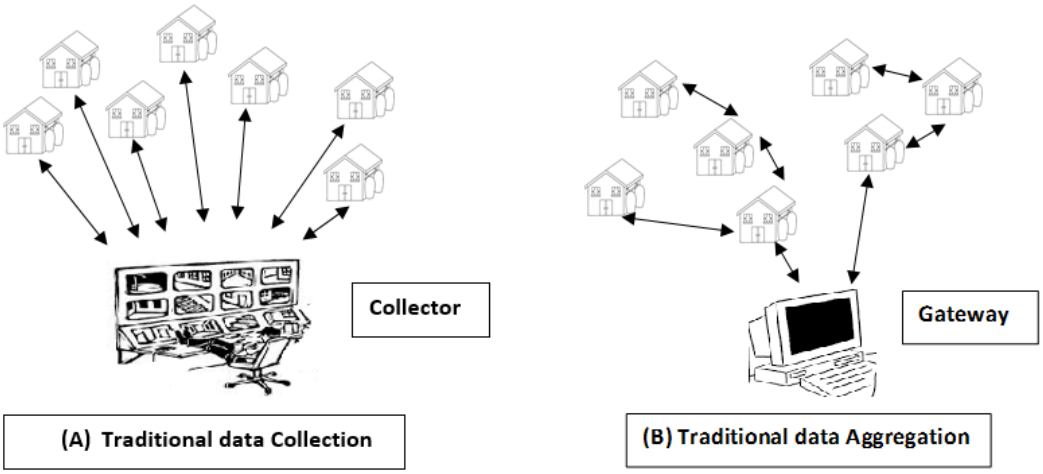

Figure 3.1: Traditional data collection and aggregation

traditional data aggregation schemes by adopting in-network aggregation techniques.

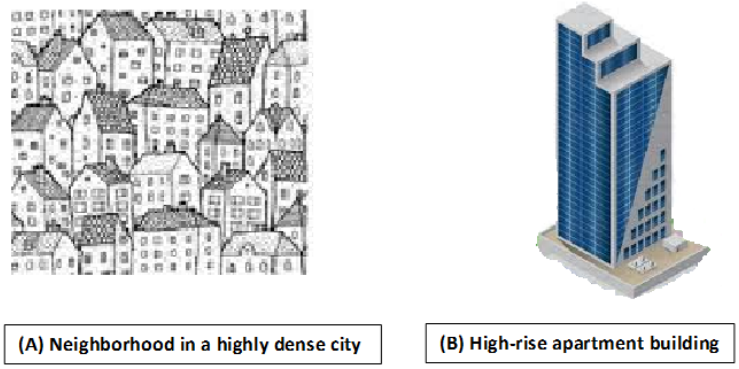

Figure 3.2: Example of neighborhoods in highly dense areas.

In Fig. 3.1(B) each smart meter establishes a connection with one neighboring smart meter, which in turn, connects with another smart meter and so for. Then, data are sent from one smart meter to another till the utility receives all information from the smart meters network. Furthermore, this approach, hop-by-hop transport, is used not only for collecting data from smart meters, but also can be used for aggregation. Concretely, in an in-network approach, we use distributed processing instead of central processing. For instance, the sum of power consumption in a building or neighborhood can be gathered and processed in each smart meter instead of the

collector and the final result will be sent to the collector and then to the utility. By doing so, we reduce the amount of data transmitted within smart meter networks, as well as save a high percentage of network load and traffic.

In in-network data aggregation techniques, especially in wireless sensor networks (WSN), data are aggregated in a distributed manner rather than at the collector devices in order to reduce the redundant data and therefore decreasing the energy consumption. This can be done by collecting information from different sensors surrounding the event, but not all of them and then sending the useful information to the collector. However, in smart grid, we sometimes need information such as the sum of power consumption in an area for operational purposes from all smart meters in an efficient way while achieving data aggregation. Therefore, we utilize data aggregation tree algorithms that are based on aggregation tree structures.

The data aggregation tree is a tree structure which contains all smart meters and does not have any loop. In tree-based aggregation algorithms each smart meter computes its power consumption and receives other smart meters' data, then aggregates this data, based on the aggregation functions and transmits the aggregated result to another smart meter or the collector in a tree or hierarchical structure. More precisely, data aggregation in smart grid mainly focuses on aggregate functions such as minimum, maximum, average, and sum that can be used in operational purposes such as balancing loads and monitoring. Usually, consumers power consumption needs to be aggregated and sent to the utility periodically i.e., every 15-30 minutes. Therefore, several data aggregation schemes have been proposed to route all or a set of smart meters readings.

In fact, data aggregation tree algorithms are of paramount importance in smart grid and have exploited several mechanisms from WSN regarding maximizing the shared path and redundancy omitting. However, finding the shortest path and constructing spanning tree while data aggregation is performed are the main objectives of these algorithms. Thus, in order to achieve the reliable multi-hop communications between gateway and smart meters, the data aggregation techniques should be efficient in terms of communications overhead and computational costs.

Unfortunately, after data aggregation techniques are adopted in smart grid, which help in efficiently aggregating power consumption data, they also have incidentally introduced several network dilemmas into smart meter networks. For example, when the data aggregation based on aggregation trees is implemented in smart grid and if some smart meter fails, a subarea or adjacent smart meters might not be reachable (either they are faulty or intact smart meters) and could be isolated from smart grid. In the worst of cases, the entire area will not be reachable if critical smart meters have failed. One of the reasons for unreachability and isolation in smart meters is the use of traditional aggregation trees that enforces selecting a specific path (i.e., the least cost path) during the aggregation, thereby preventing some smart meters in the network from selecting another link as shown in Fig. 8. It is obvious that smart meters 8 and 9 are

unreachable while smart meters 4, 5, 7, and 10 are faulty. Since smart meters are logically organized in an aggregation tree and based on a spanning tree, we cannot detect which of smart meters is faulty or unreachable.

Since data aggregation tree algorithms are used to aggregate consumers consumption, we need to improve the protection of the privacy of consumers by protecting users' data from malicious users, insiders or outsiders, and by preventing any entity or third-party companies from associating specific information to specific consumers. Accordingly, if the privacy is not protected, residential consumers may be reluctant to contribute their data to make the success of smart grid. Thus, providing an efficient security scheme for protecting aggregated data in smart grid without taking account of data aggregation tree issues might degrade any proposed privacy-preserving aggregation scheme and might involve inaccurate or unexpected aggregated results from smart meters.

In this chapter, we propose an efficient lightweight privacy-preserving data aggregation scheme for smart meters using efficient data aggregation tree algorithm, which is based on geographic distribution of smart meters at houses or buildings for improving security and privacy while taking faulty meters detection into consideration. A preliminary version of this work has been reported in [7]. The proposed scheme integrates one-time masking technique with tree network topology to protect user power consumption data while achieving efficient lightweight data aggregation. The main contributions of this scheme include:

- We propose an efficient lightweight privacy-preserving data aggregation scheme, called ELPDA, for improving security and privacy in smart grid. The ELPDA consolidates the one-time masking technique with tree-based network topology in order to protect user power consumption data while performing lightweight aggregation. Compared with popular Paillier Cryptosystem based aggregation (PCBA) algorithm applied in smart grid [8, 10, 34], the proposed ELPDA is much more efficient, reducing the aggregation delay in the whole residential area network.
- We further propose a faulty smart meter detection scheme to locate faulty smart meters and then dynamically rebuild the aggregation tree with removal of these faulty smart meters.
- The proposed scheme resists various security and privacy threats. In addition to that, it preserves user privacy while achieving lightweight aggregation.

The rest of this chapter is organized as follows. Section 3.2 introduces our system model, security model and design goal. In Section 3.3, we review the bilinear pairing technique and data aggregation techniques as preliminaries. Then, we present our ELPDA in Section 3.4, followed by analysis in Section 3.5. Finally, we draw our conclusions in Section 3.6.

3.2 System Model, Security Model and Design Goals

In this section, we formalize the system model, security model, and identify our design goals.

3.2.1 System Model

In our system model, we mainly consider a typical residential area network (RAN) of smart meter devices in a smart grid system, which consists of a control center (CC), a RAN gateway (R-Gateway), and a number of smart meters (SMs) $\{SM_1, SM_2, \cdots\}$ within the RAN, as shown in Fig. 3.3.

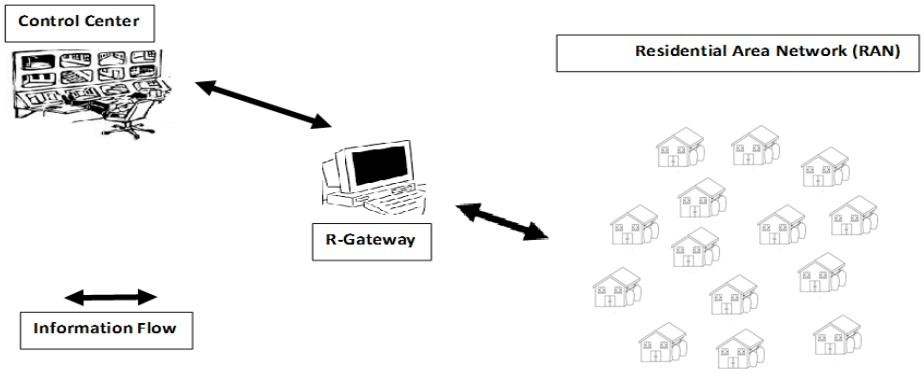

Figure 3.3: System model under consideration: a residential area network (RAN) with a number of smart meters (SMs) and RAN gateway.

- Control Center (CC): CC is a trusted entity, whose duties include initializing the whole system, and is responsible for collecting the data from RAN, and real-time monitoring the electricity consumption at RAN in smart grid systems.
- RAN Gateway (R-Gateway): R-Gateway is a network entity which serves as a relay between the CC and smart meters to help exchanging request and response data. In the data aggregation applications, R-Gateway also helps the CC aggregate the electricity consumption data in the whole residential area.
- Smart Meters Network (SMs) $\{SM_1, SM_2, \cdots\}$: Smart meter is an important component that can electrically record the nearly real-time data about each home's electricity consumption. Smart meters not only report the real-time data to the R-Gateway, but also receive the requests from the latter.

Communication model. In the residential area network (RAN), the communication between each SMi and the R-Gateway is through relatively cheap WiFi technology. However, when a smart meters network is huge or the distance between them is far away, it is impossible for some SMi to directly communicate with R-Gateway. In this case, multi-hop communication will be formed in RAN. On the other hand, the communication between R-Gateway and CC is dependent upon the high-bandwidth, low delay, reliable and secure channel, which can guarantee smart grid's two-way communication, facilitating for demand response, dynamic pricing, and system monitoring in smart grid systems.

3.2.2 Security Model

In our security model, the focus is not only on how to provide the security of data exchanged between the R-Gateway and SMs, but also on how to protect users privacy. Specifically, the following security requirements should be satisfied.

- *Request command should be authorized*. Only the authorized request command can be accepted by the SMs with correct response. That is, if a command is not from an authorized entity, the request command won't be executed by the SMs.
- *Hop-by-Hop exchange should be authenticated*. When SMs report their accurate and nearly real-time data, the data integrity should be provided, i.e., any bogus data inserted or omitted by an adversary should be detected. Otherwise, it will manipulate the data received at CC, making CC take a wrong action/decision.
- *Consumer Privacy should be protected*. While security is crucial for the success of secure smart grid communications, privacy is very sensitive to users. When a smart meter reports its accurate and nearly real-time data, the data may disclose the users privacy. Therefore, privacy-preserving data aggregation algorithm is expected to not only meet smart grid's application requirements, but also ensure user privacy preservation.

3.2.3 Design Goals

Our design goals are to propose an efficient lightweight privacy-preserving data aggregation (ELPDA) scheme by integrating an efficient aggregation tree algorithm with LPDA [7] to improve data aggregation while taking faulty meters detection into consideration and to satisfy the above security requirements. Specifically, the following desirable goals should be achieved:

- *Security and privacy preservation*: The proposed ELPDA scheme should take advantage of the real-time data security and the user privacy to satisfy the security requirements above.
- *Efficiency*: The proposed ELPDA scheme should also be efficient, i.e., compared with previously reported ones, the computational and communication costs of the proposed LPDA should be minimized.
- *Building efficient aggregation trees for smart meters*, based on geographic distribution of smart meters at a residential area (such as houses or buildings).
- *Designing and developing an efficient faulty smart meters detection algorithm.*
- *Reconstructing data aggregation trees* when failures occur to smart meters by removing all faulty meters from the running aggregation tree and rebuilding a new data aggregation tree excluding the faulty smart meters.

3.3 Preliminary

In this section, we give the essential features of the bilinear pairing technique, which serves as the basis of the proposed ELPDA scheme, and review the data aggregation techniques, which help in constructing data aggregation trees.

3.3.1 Bilinear Pairing

Let $\mathbb{G}_1$ and $\mathbb{G}_2$ be two cyclic groups of the same large prime order p, and $P_1 \in \mathbb{G}_1$ be the generator of $\mathbb{G}_1$. An *admissible* bilinear pairing $\hat{e} : \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$ is a map with the following properties: i) *Bilinearity:* For all $P, Q \in \mathbb{G}_1$ and any $a, b \in \mathbb{Z}$, we have $\hat{e}(aP, bQ) = \hat{e}(P, Q)^{ab}$; ii) *Non-degeneracy:* $\hat{e}(P_1, P_1) \neq 1_{\mathbb{G}_2}$; and iii) *Computability:* There is an efficient algorithm to compute $\hat{e}(P, Q)$ for all $P, Q \in \mathbb{G}_1$. Such an *admissible* bilinear pairing $\hat{e} : \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$ can be implemented by the modified Weil/Tate pairings over elliptic curves [55].

Definition 1 (Bilinear Parameter Generator)**.** *: A bilinear parameter generator $\mathcal{G}en$ is a probabilistic algorithm that takes a security parameter k as input, and outputs a 5-tuple $(p, P_1, \mathbb{G}_1, \mathbb{G}_2, \hat{e})$ where p is a k-bit prime number, $\mathbb{G}_1, \mathbb{G}_2$ are two groups with order p, $P_1 \in \mathbb{G}_1$ is a generator, and $\hat{e} : \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$ is a non-degenerated and efficiently computable bilinear map.*

3.3.2 Data Aggregation Techniques

In fact, not all aggregation construction algorithms used in the past (e.g., in WSNs) can be ideal for data aggregation in smart grid because of several reasons including, but not limited to, computation and communication constraints, number of messages and time required in those algorithms, network topology, or fault detection and recovery. While the aforementioned constraints and requirements are common in smart grid and WSNs, smart grid networks have unique requirements are not identical to WSNs. For example, aggregation trees algorithms used in WSNs can be constructed without including all sensors in order to save energy and only sensors surrounding the event will report collected data to the sink (i.e., aggregator) as shown in Fig. 3.4.

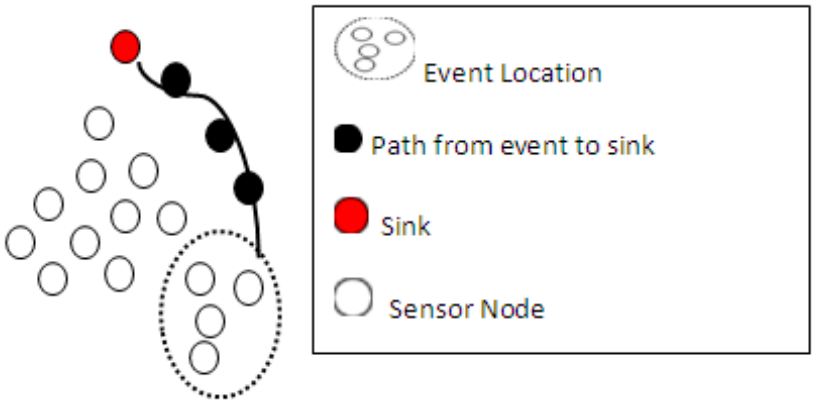

Figure 3.4: Traditional aggregation in WSN.

On the contrary, in smart grid, we are interested in gathering data collected from all nodes (i.e., smart meters). Furthermore, in WSNs, if a node surrounding the event has failed, another node might be chosen or data could be aggregated without any effect from the faulty node on the total aggregation. As a result, the detection of faulty smart meters in smart grid is not necessary to be the same act of faulty sensor nodes detection in WSNs due two reasons: 1) the uniqueness of shortest paths or minimum spanning trees limits constructing data aggregation trees and allows the aggregation tree to use specific paths among meters, and 2) It is essential to include all smart meters for aggregation in a subarea or neighbor.

3.4 Proposed ELPDA Scheme

In this section, we will present our efficient lightweight privacy-preserving data aggregation (ELPDA) scheme. Before proceeding to the scheme's details, we first introduce an overview on the ELPDA scheme.

3.4.1 Overview of ELPDA Scheme

In a typical residential area network of smart meters based on geographic locations of homes and businesses, it is not efficient for some smart meters at different locations to directly communicate with R-Gateway. Therefore, it is unavoidable to adopt decentralized data communication techniques when the flow of data in RAN is only done by hop-by-hop transport, especially when data are collected and aggregated from far away smart meters. The proposed ELPDA scheme takes advantage of in-network processing techniques to achieve reliable multi-home meter data communication in RAN. Fig. 3.5 shows such a multi-hop topology, where the R-Gateway serves as the root, and smart meters represent other nodes. In order to achieve the reliable multi-hop communication between R-Gateway and all smart meters, each smart meter parent forwards the REQUEST from R-Gateway or from other smart meters to its subsequent smart meter(s), and also aggregates the RESPONSE from its subsequent node(s) and returns to its predecessor smart meter(s), finally to the root (R-Gateway). In addition, in order to achieve privacy-preserving aggregation in meter periodic data report scenario, the ELPDA also employs the one-time masking technique together with hop-by-hop authentication to guarantee the transmission's efficiency and integrity protection. Moreover, in order to achieve detecting faulty smart meters in smart grid, the ELPDA makes use of data aggregation trees and spanning table. In the next subsection, we describe the ELPDA scheme in details.

3.4.2 Description of ELPDA Scheme

The ELPDA scheme consists of five phases: system initialization phase, aggregation tree construction, aggregation request phase, aggregation response phase, and fault detection phase.

1) System Initialization

In the system initialization phase, the control center (CC) is in charge of system parameter configuration. Specifically, given the security parameter k, CC first generates the bilinear parameter $(p, P_1, \mathbb{G}_1, \mathbb{G}_2, \hat{e})$ by running $\mathcal{G}en(k)$, and chooses two secure cryptographic hash functions

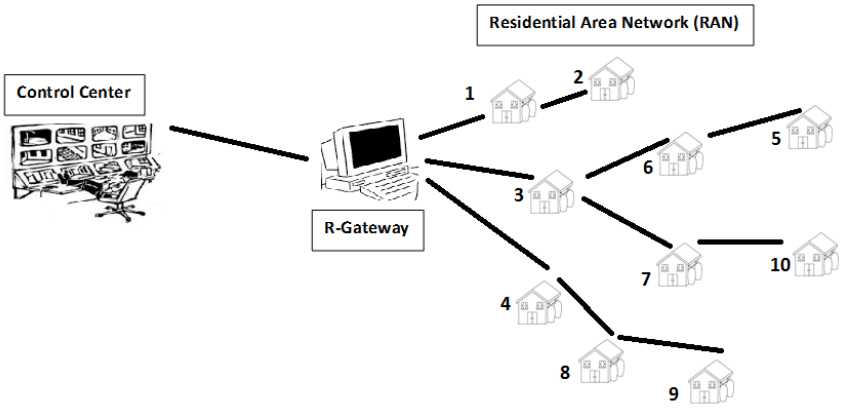

Figure 3.5: Multi-hop communication in a smart meters network

H_1, H_2, where $H_1 : \{0,1\}^* \to \mathbb{G}_1$ and $H_2 : \{0,1\}^* \to \mathbb{Z}_p^*$. Then, CC chooses a random number $s \in \mathbb{Z}_p^*$, and computes $P_{pub} = sP_1$. With these settings, CC keeps s as the master keys secretly, and publishes the public parameters $Pubs = (p, P_1, \mathbb{G}_1, \mathbb{G}_2, \hat{e}, H_1, H_2, P_{pub})$.

When the gateway makes the registration, it submits its identity $\texttt{R-Gateway}$ to CC. CC uses the master key s to compute R-Gateway's private key as $S_B = sH_1(\texttt{R-Gateway})$, and returns S_B back to the R-Gateway in a secure channel. Likewise, each smart meter $SM_i \in \{SM_1, SM_2, \cdots\}$ submits its identity $\texttt{SM}_i$ to CC for registration, and also gets its private key $S_i = sH_1(\texttt{SM}_i)$ from CC via a secure channel.

Non-interactive static key establishment: With the private key $S_B = sH(\texttt{R-Gateway})$, the R-Gateway can establish a static key with each $SM_i \in \{SM_1, SM_2, \cdots, SM_n\}$ in a non-interactive way, i.e., computing $K_{bi} = \hat{e}(S_B, H_1(\texttt{SM}_i))$. SM_i can also establish the corresponding key as $K_{ib} = \hat{e}(S_i, H_1(\texttt{R-Gateway}))$. The correctness is as follows:

$$\begin{aligned} K_{bi} &= \hat{e}(S_B, H_1(\texttt{SM}_i)) = \hat{e}(sH_1(\texttt{R-Gateway}), H_1(\texttt{SM}_i)) \\ &= \hat{e}(H_1(\texttt{R-Gateway}), sH_1(\texttt{SM}_i)) \\ &= \hat{e}(H_1(\texttt{R-Gateway}), S_i) = K_{ib} \end{aligned}$$

In addition, any two smart meters can also establish their static key in the same way. For example, if SM_i and SM_j are two neighboring smart meters in the topology, they can calculate their static key as the following:

$$K_{ij} = K_{ji} = \hat{e}(H_1(\texttt{SM}_j), H_1(\texttt{SM}_i))^s.$$

2) Construction of Aggregation Tree

In aggregation tree construction phase, we consider a typical smart meter network in a residential area (RAN) as a connected, directed graph in which 1) the gateway is represented by the root C_0 , 2) smart meters are represented by a set of vertices $\{SM_1, SM_2, \cdots, SM_N\}$ and 3) the available wireless connection is represented by a set of edges. The gateway is connected with smart grid operation center at the utility, and with a large number of residential consumers (i.e., smart meters). Before constructing a minimum spanning tree that is used for data aggregation, we need to build a spanning table. This table contains significant information about all smart meters in a RAN such as the weights of edges and the nearest connected smart meters to each smart meter as shown in Table 3.1. For example, SM_1, SM_2 and SM_3 are connected with the gateway C_0 with the weights of v_i, v_j, v_k, respectively. Also, they could be connected to other smart meters. While smart meter SM_1 can only connect to the collector, SM_1 and SM_3, there is no link between SM_1 and SM_4; and we denote that by ϕ. Finally, we denote the self-loop by (–) to indicate there is no link within a smart meter or gateway and itself. v_{i1}, v_{i2}, v_{i3}, and v_{j1}, v_{j2} *etc*. are values used to refer to the communication costs between smart meters. In other words, these values are the shortest paths.

Table 3.1: Example of a spanning table

	C_0	SM_1	SM_2	SM_3	SM_4	SM_i	$\cdots$	SM_N
C_0	–	v_{i1}	v_{j1}	v_{k1}	ϕ	ϕ	ϕ	ϕ
SM_1		–	v_{i2}	v_{j1}	ϕ	ϕ	ϕ	ϕ
SM_2			–	v_{i2}	v_{j2}	v_{k1}	ϕ	ϕ
SM_3				–	v_{i3}	v_{j3}	ϕ	ϕ
SM_4					–	v_{i4}	v_{m1}	ϕ
SM_i						–	v_{m2}	v_{n1}
$\cdots$							–	$\cdots$
SM_N								–

Data Aggregation Tree Algorithm:Data Aggregation Tree algorithm (DAT) is responsible for forming a minimum spanning tree used in data aggregation. The main objective of DAT algorithm is to find the minimum spanning tree that contains all smart meters constructed from the spanning table with the smallest total cost (e.g., it is less costly in terms of communications). For instance, when DAT algorithm is executed on the spanning table as shown in Table 3.2, it will form the following minimum spanning tree:<C_0-SM_1>, <C_0-SM_3>, <C_0-SM_4>,

$<C_0$-$SM_6>$, $<SM_1$-$SM_2>$, $<SM_5$-$SM_6>$, $<SM_3$-$SM_7>$, $<SM_7$-$SM_{10}>$, $<SM_4$-$SM_8>$, and $<SM_8$-$SM_{10}>$. In this case, the gateway is directly connected to smart meters SM_1, SM_3, SM_4, and SM_6 while SM_{10} is indirectly connected to the gateway via SM_8 and SM_4, respectively. Also, SM_2 can be directly connected to the gateway, but with higher cost, therefore, it can be connected to SM_1 with lower cost. The lower cost is preferable when it is possible in a spanning tree, otherwise the higher cost. Similarly, other smart meters also are connected to SM_3 and SM_4, and so forth to form a spanning tree as in Fig. 3.5. As a result, when a spanning tree is constructed, we will use it as a prominent part for performing data aggregation.

Algorithm 1 Data Aggregation Tree Algorithm (DAT)

INPUT: A set of all mart meters and spanning table.
OUTPUT: Efficient Data Aggregation Tree.

1: **procedure** DATA AGGREGATION TREE ALGORITHM (DAT)
2: Set C_0 as the gateway.
3: Set *MSP*=ϕ ▷ MSP is a minimum spanning tree set
4: Set *RS*= C_0 ▷ *RS* is the current smart meter
5: Set SM_i = Select cheapest unused smart meter connected to *RS*.
6: *MSP* =$\{<$*RS*–$SM_i>\}$
7: **repeat**
8: SM_i=(Select the next cheapest smart meter connected with *RS*
9: **until** *NO* smart meters connect to *RS*.
10: **for** i=1 To N **do**
11: Set *RS*= SM_i
12: **run** steps [5–7]
13: i= $i+1$
14: **if** i =$N+1$ **then** GOTO Line 17
15: **end if**
16: **end for**
17: **outputs** *MSP*
18: **end procedure**

3) Aggregation Request Phase

In order to achieve more accurate and nearly real-time electricity consumption data in the RAN, the R-Gateway will send an aggregation request command REQUEST to all SMs in the RAN

every 15-30 minutes [10]. Particularly, the R-Gateway first sets REQUEST which includes a monotone increasing timestamp, and then runs the following steps:

Step 1: choose a random number $r \in \mathbb{Z}_p^*$, and compute $U = rH_1(\texttt{R-Gateway})$, $V = (r + H_2(\texttt{REQUEST}||U))S_B$.

Step 2: send REQUEST together with (U, V) to all its subsequent nodes as shown in Fig. 3.6.

After receiving $\texttt{REQUEST}||(U, V)$ from the R-Gateway, each subsequent node first checks the timestamp in REQUEST to resist a potential replay attack. If the timestamp is valid, the following equation will be performed:

$$\hat{e}(P_1, V) \stackrel{?}{=} \hat{e}(P_{pub}, U + H_2(\texttt{REQUEST}||U)H_1(\texttt{R-Gateway}))$$

If it does hold, the REQUEST is authenticated, and $\texttt{REQUEST}||(U, V)$ will be further forwarded to the subsequent nodes. Otherwise, $\texttt{REQUEST}||(U, V)$ is invalid and will be rejected. The correctness is as follows:

$$\begin{aligned}\hat{e}(P_1, V) &= \hat{e}(P_1, (r + H_2(\texttt{REQUEST}||U))S_B) \\ &= \hat{e}(sP_1, (r + H_2(\texttt{REQUEST}||U))H_1(\texttt{R-Gateway})) \\ &= \hat{e}(P_{pub}, U + H_2(\texttt{REQUEST}||U)H_1(\texttt{R-Gateway}))\end{aligned}$$

Other nodes use the same way to verify-forward $\texttt{REQUEST}||(U, V)$ to their subsequent smart meters. Finally, the valid REQUEST from the R-Gateway will be received by all smart meters in the residential area network.

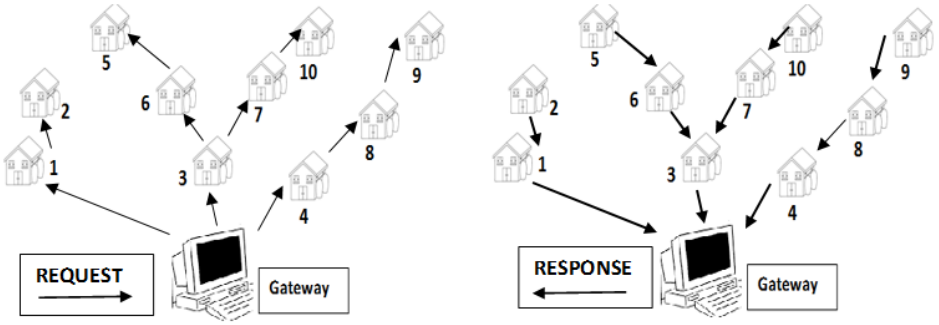

Figure 3.6: Aggregation REQUEST and RESPONSE in RAN

4) Aggregation Response Phase

In response to REQUEST, each smart meter $SM_i \in \{SM_1, SM_2, \cdots, SM_n\}$ collects its electricity consumption data m_i and performs the following steps:

Step 1: With the one-time masking technique, SM_i uses the static key k_{ib} shared with the R-Gateway to compute

$$c_i = m_i + H_1(\texttt{REQUEST}||k_{ib}) \bmod p$$

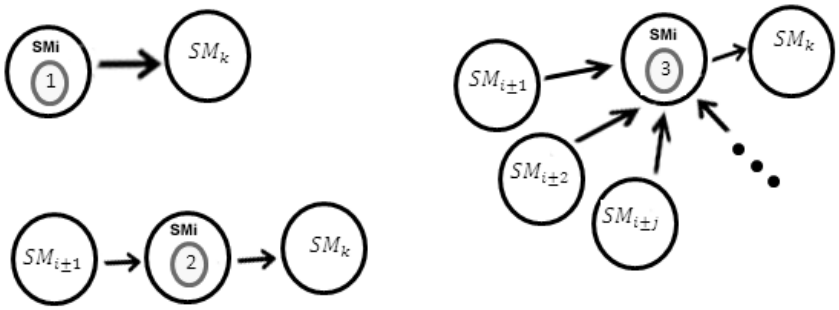

Figure 3.7: Categories of smart meters

Step 2: In the multi-hop topology in the RAN, each SM_i can be categorized into three types, namely i) node without subsequent node, ii) node with one subsequent node and iii) node with two or more subsequent nodes, as shown in Fig. 3.7. Then, based on the different type, SM_i runs the Algorithm 2 to perform the hop-by-hop authentication and forwarding. Assume there are total n smart meters in the RAN. Then the aggregated data $C = \sum_{i=1}^{n} c_i$ will finally arrive at the R-Gateway in a multi-hop way, as shown in Fig. 3.6.

Upon receiving $C = \sum_{i=1}^{n} c_i$, the R-Gateway calculates $C - \sum_{i=1}^{n} H_2(\texttt{REQUEST}||k_{bi}) \bmod p$ to recover the aggregate data $M = \sum_{i=1}^{n} m_i$. Since the electricity consumption data within 15-30 minutes should be small in size, it is reasonable to assume $\sum_{i=1}^{n} m_i < p$. Then, the correctness

Algorithm 2 Hop-By-Hop Authentication and Forwarding

procedure HOP-BY-HOP AUTHENTICATION AND FORWARDING
  switch SM_i **do**
    case SM_i is Type 1
      set $C_i = c_i$
      run the hop-by-hop authentication with SM_k and forward C_i to SM_k, as shown in Fig. 3.7
    case SM_i is Type 2
      after receiving $C_{i\pm1}$ from $SM_{i\pm1}$, set $C_i = c_i + C_{i\pm1}$
      run the hop-by-hop authentication with SM_k and forward C_i to SM_k
    case SM_i is Type 3
      after receiving $C_{i\pm1}$, $C_{i\pm2}$,$\cdots$,$C_{i\pm j}$ from $SM_{i\pm1}$, $SM_{i\pm2}$,$\cdots$, $SM_{i\pm j}$, set $C_i = c_i + C_{i\pm1} + C_{i\pm2} + \cdots + C_{i\pm j}$
      run the hop-by-hop authentication with SM_k and forward C_i to SM_k
end procedure

is as follows:

$$
\begin{aligned}
& C - \sum_{i=1}^{n} H_2(\texttt{REQUEST}||k_{bi}) \bmod p \\
=& \sum_{i=1}^{n} c_i - \sum_{i=1}^{n} H_2(\texttt{REQUEST}||k_{bi}) \bmod p \\
=& \sum_{i=1}^{n} m_i + H_2(\texttt{REQUEST}||k_{ib}) - H_2(\texttt{REQUEST}||k_{bi}) \bmod p \\
=& \sum_{i=1}^{n} m_i \bmod p = \sum_{i=1}^{n} m_i = M \quad (\because k_{ib} = k_{bi}, \sum_{i=1}^{n} m_i < p)
\end{aligned}
$$

With the received aggregate data $M = \sum_{i=1}^{n} m_i$, the R-Gateway chooses a random number $\bar{r} \in \mathbb{Z}_p^*$, computes $\bar{U} = \bar{r}H_1(\texttt{R-Gateway})$, $\bar{V} = (\bar{r} + H_2(M||\bar{U}))S_B$, and then reports $M||\bar{U}||\bar{V}$ to the control center via a secure channel. The control center can verify the aggregated data M by checking $\hat{e}(P_1, \bar{V}) = \hat{e}(P_{pub}, \bar{U} + H_2(M||\bar{U})H_1(\texttt{R-Gateway}))$ for demand response, dynamic pricing and real-time system monitoring in smart grid systems.

SM_i	predecessor SM_{i-1}
static shared key $K_{i(i-1)} = K_{(i-1)i}$, REQUEST	
1) compute $sk = H_2(\texttt{REQUEST}\|\|K_{i(i-1)})$	
$MAC = H_2(C_i\|\|sk)$	
	$\xrightarrow{C_i, MAC}$
2)	compute $sk' = H_2(\texttt{REQUEST}\|\|K_{(i-1)i})$
	$MAC \stackrel{?}{=} H_2(C_i\|\|sk')$
	if it holds, C_i is authenticated and accepted by SM_{i-1}

Figure 3.8: Anonymous authentication protocol (AAP) for group members

5) Fault Detection Phase

It is worth noting that the proposed lightweight response aggregation mechanism will not work if there is any smart meter(s) that malfunction. For example, in Fig. 3.9, we can see that smart meters 4, 5, 7 and 10 are faulty and smart meter 4 is a parent of non-faulty or unreachable smart meters. For example, when performing data aggregation, the smart meter 4 will not send its data nor will smart meter 8 and 9. The gateway also does not know if they are faulty or intact. Thus, the gateway is not sure how many smart meters are faulty or intact behind the faulty smart meter(s). In some situation, it could be one smart meter has failed after the faulty one and the rest are intact, or all could be intact or fail. Therefore, we need to detect faulty smart meter(s) in order to successfully recover the aggregated data which are received from all reachable smart meters. In other words, the detected faulty meter(s) and their successor nodes are considered unreachable and should be removed from the current response aggregation procedure. Afterwords, all the smart meters that are working properly will be re-organized to form another aggregation tree.

In this phase, we describe the faulty smart meters detection algorithm while achieving aggregation in more detail.

We propose some assumptions before going any further and deciding whether a smart meter is faulty or not. First, we consider any parent or leaf in a smart meter network to be faulty if it has predecessor paths (i.e., it is reachable) and has not responded to the gateway, e.g., smart meter 5 in Fig. 3.9. As a result, smart meter 5 is considered to be faulty since it has a predecessor path via smart meter 6, but has not responded to the gateway. In the case that leaves or parents were faulty, they will be excluded from the re-construction aggregation tree and marked as dead or faulty smart meters while its children are not necessarily faulty. In fact, a faulty parent is not necessary to be followed by faulty children smart meters and they will be marked as unreachable, but not

faulty. If this case happens, the algorithm will re-construct the aggregation tree and associate children smart meters with a new reachable parent when it is possible. Sometimes the algorithm associates children with a predecessor or successor parent even though we do not know the parent status provided that it is not being unreachable. The second assumption is that we consider the smart meters network isolated if the gateway is not connected with any smart meter even though they are connected with each other except the gateway. The third assumption is that we assume each smart meter has at least two different paths for communications with its neighbors. The last assumption may guarantee an alternate path when the failure happens to the other path, and will help in re-constructing the aggregation tree.

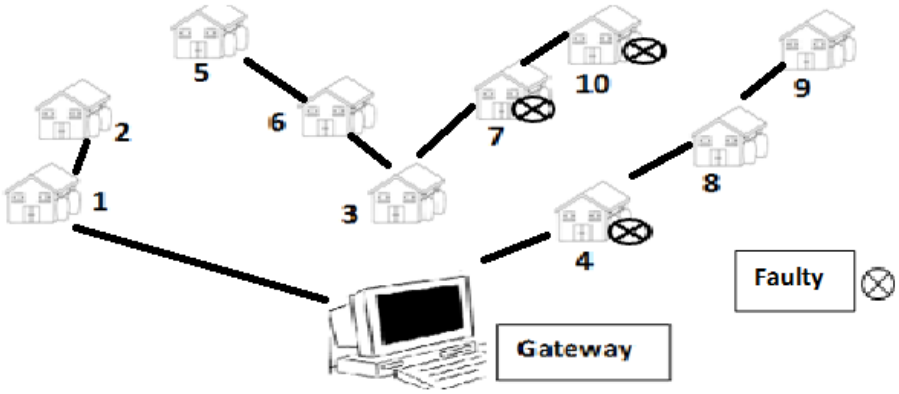

Figure 3.9: Faulty smart meters in smart grid

In this phase, the re-constructing aggregation tree algorithm and fault detection algorithm have to construct sets of smart meters from which we can detect faulty meters. In the first round of re-construction, the re-construction algorithm will construct three sets; $\mathcal{S}$ is a set of all smart meters (faulty and running smart meters) and $\mathcal{F}$ denotes a set of only faulty parent(s) and only leaves that are considered faulty as mentioned in assumption 1 above. However, non-faulty or unreachable smart meters are not included in $\mathcal{F}$. The third set is $\mathcal{CS}$ which is a set containing the difference set between $\mathcal{S}$ and $\mathcal{F}$, the set of only non-faulty and/or only unreachable nodes. More precisely, when $\mathcal{F}$ is subtracted from $\mathcal{S}$, it would be identical to the difference between $\mathcal{S}$ and $\mathcal{F}$, that is, the result of $\mathcal{S} - \mathcal{F}$ is put into the $\mathcal{CS}$ set. In this procedure, we exclude smart meters satisfying the first assumption and can track other smart meters that have not responded to the gateway by associating them with a new reachable parent meters when it is possible. For example, SM_{10} will be associated with SM_6, but not with SM_5. Also, SM_8 and SM_9 together will be linked to SM_3 according to communications and spanning table in Table 3.2.

At this time, we will execute the proposed data aggregation tree algorithm `Algorithm 1`

to build a new spanning tree. However, smart meters passed or used in `Algorithm` 1 would not be the same smart meters used in the first time when the algorithm was invoked. The reason for that is some smart meters have been excluded due to being faulty. Moreover, smart meters in Table 3.2 that have been marked as faulty will not be used as parents or any intermediate smart meters in `Algorithm` 1.

After several rounds of execution of re-constructing aggregation trees, the gateway should have received some responses from some smart meter(s) or non-faulty meters, otherwise the non-faulty one(s) might be isolated (e.g., where we could not check from being faulty or running), making it difficult to the gateway to make sure about their status. If the gateway still has not received responses from some smart meters after associating unreachable meters, the set $\mathcal{CS}$ is then considered to be either unreachable or isolated while the set $\mathcal{S}$ is considered the set containing only faulty smart meters.

In the meanwhile, the aforementioned steps are used as well in the fault detection algorithm for several rounds till the algorithm detects all faulty meters as in the set $\mathcal{F}$ or deduces isolated and/or unreachable ones as in the set $\mathcal{CS}$.

Concretely, the fault detection algorithm detects fault parents or leaves (when no responses) by looking at the set $\mathcal{F}$ while it marks smart meters as unreachable or isolated as in the set $\mathcal{CS}$.

Table 3.2: A numerical example of spanning table

	C_0	SM_1	SM_2	SM_3	SM_4	SM_5	SM_6	SM_7	SM_8	SM_9	SM_{10}
C_0	–	5	16	6	7	ϕ	12	ϕ	ϕ	ϕ	ϕ
SM_1		–	3	14	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ
SM_2			–	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ
SM_3				–	ϕ	ϕ	ϕ	4	15	ϕ	ϕ
SM_4					–	ϕ	ϕ	ϕ	2	ϕ	ϕ
SM_5						–	4	8	ϕ	ϕ	9
SM_6							–	8	ϕ	ϕ	9
SM_7								–	4	ϕ	3
SM_8									–	2	ϕ
SM_9										–	ϕ
SM_{10}											–

3.5 Analysis and Evaluation

3.5.1 Security Analysis

In this section, we will discuss the security of the proposed ELPDA scheme, i.e., the request command's authentication, hop-by-hop authentication, and user data privacy preservation in aggregation.

• *The proposed ELPDA scheme can provide the request command's authentication.* When the R-Gateway sends a REQUEST to all HANs in the building, it also attaches the signature (U, V), where $U = rH_1(\texttt{R-Gateway})$, $V = (r + H_2(\texttt{REQUEST}||U))S_B$. Since the signature (U, V) is provably secure in the random oracle model [61], it can resist deliberate forgery attacks. In other words, the proposed ELPDA scheme can provide the request command's authentication. In addition, since REQUEST includes a monotone increasing timestamp, which can resist the possible replay attack as well.

• *The proposed ELPDA scheme provides the hop-by-hop authentication.* Because of the unique timestamp in REQUEST, the session key $sk = H_2(\texttt{REQUEST}||K_{i(i-1)})$ for SM_i and SM_{i-1} in Fig. 3.8 secure and distinct each other. As a result, with the fresh sk in $MAC = H_1(C_i||sk)$, HAN_{i-1} can authenticate C_i really comes from its subsequent node HAN_i. Suppose an adversary wants to insert a bogus data to pollute the result $C = \sum_{i=1}^{n} c_i$. In order to make it successful, the adversary must guess at least one correct value of MAC in total n hop-by-hop forwarding. Since the hash function $H_1 : \{0,1\}^* \rightarrow \mathbb{Z}_p^*$ has possible q outputs, the probability that the adversary can pollute the aggregated result $C = \sum_{i=1}^{n} c_i$ is $1-(1-\frac{1}{p})^n$. When q is large enough, i.e., $\frac{n}{p} \rightarrow 0$, we know

$$1-(1-\frac{1}{p})^n \approx 1-e^{-\frac{n}{p}} \rightarrow 0$$

which shows that the proposed ELPDA scheme can provide the hop-by-hop authentication and resist any bogus data inserting attack during the aggregation phase.

• *The proposed ELPDA scheme provides user data privacy preservation in aggregation.* In the proposed ELPDA scheme, each SM_i's data m_i is masked by $H_2(\texttt{REQUEST}||k_{ib})$ in $c_i = m_i + H_2(\texttt{REQUEST}||k_{ib}) \bmod p$. Since $H_2(\texttt{REQUEST}||k_{ib})$ won't be used more than once, the data m_i is secure. In addition, when the aggregated data

$$C = \sum_{i=1}^{n} m_i + H_2(\texttt{REQUEST}||k_{ib}) \bmod p$$

is reported to the R-Gateway, each m_i is still hidden. Due to these reasons, the proposed ELPDA scheme provides user data privacy preservation in aggregation.

Summarizing the above analysis, we can clearly see that the proposed ELPDA scheme can provide request command authentication, hop-by-hop authentication, and user data privacy in data aggregation.

3.5.2 Average Aggregation Delay(AAD)

In this section, we evaluate the performance of the proposed ELPDA scheme in terms of the average aggregation delay and then compare the average aggregation delay and the security strength of the proposed scheme with Paillier Cryptosystem Based Aggregation (PCBA) [8].

The AAD is defined as the average time between the furthest smart meter in the aggregation based tree and the residential gateway (R-Gateway) when the furthest smart meter begins calculating the electricity consumption data and when the R-Gateway successfully recovers the aggregated data.

While the proposed ELPDA scheme and PCBA scheme lower the communication overhead by implementing aggregation trees, the proposed ELPDA scheme efficiently reduces the average aggregation delay (AAD) compared to the PCBA scheme. Since the proposed scheme performs aggregation based on the modular addititon in $\mathbb{Z}_n$, the AAD is considered negligible [56] compared to the PCBA scheme. The AAD in the proposed scheme is dominated by the hop-by-hop communication delay while the PCBA scheme is dominated by the hop-by-hop communication delay as well as the computational delay inherent in the Paillier Cryptosystem (e.g., a single modular multiplication in $\mathbb{Z}_{n^2}$ costs 4 modular multiplication operations in $\mathbb{Z}_n$. [57])

In order to compare the AAD in both schemes, we construct a perfect rooted k-ary tree as shown in Fig 3.10 and compute the AAD as given in Table 3.3. The detailed parameter settings are summarized in Table 3.4.

Table 3.3: Computation of average aggregation delay

Scheme	**Average Aggregation Delay**
PCBA	$T_{pcba} = (T_{agg} \times [(\sum_{i=1}^{h-1} \lvert k^i \rvert) + 1]) + (h \times T_{comm})$
ELPDA	$T_{elpda} = h \times T_{comm}$

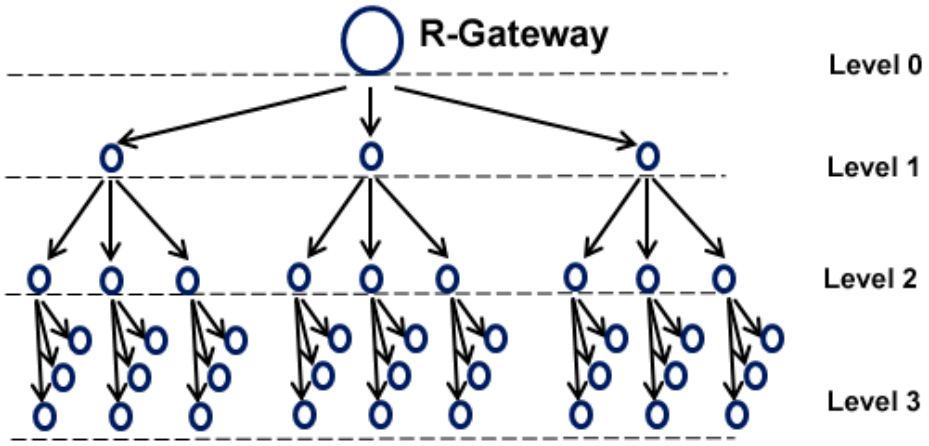

Figure 3.10: Perfect rooted k-ary tree

3.5.3 Simulation

To simulate a tree-based residential area network, we consider a perfect rooted k-ary aggregation tree in which all nodes are at the same depth having the degree of k. The height of the tree is denoted by h. The transmission radius of each smart meter-gateway is $r = 30$ m, the hop-by-hop communication delay T_{comm} is set to 300 ms and the computational delay of PCBA T_{agg} is set to 0.8851 ms [57, 58]. T_{agg} is computed from: i) the modular multiplication in $\mathbb{Z}_n$ which costs 0.2951 ms for a single modular multiplication operation. ii) a single modular multiplication operation in $\mathbb{Z}_{n^2}$ costs 4 modular multiplication operations in $\mathbb{Z}_n$. iii) as the computational delay in PCBA is non-negligible and affects data aggregation, each aggregating node will face a delay of $k \times 0.2951$.

The number of intermediate nodes that aggregate data is given by $(\sum_{i=1}^{h-1} |k^i|) + 1$, where h>1 and $|k^i|$ denotes the number of nodes in level i . If $h = 1$ then the number of aggregating nodes is only k.

In order to show the efficiency of the proposed ELPDA, we compare the proposed ELPDA scheme with Paillier cryptosystem based aggregation (PCBA) with modulus 2048 [8]. The detailed parameter settings are summarized in Table 3.4.

In terms of the security strength, the ELPDA with 224-bit p is secure enough for the aggregation application in smart grid, and is equivalent to 2048-modulus size. This is because we implement the modified Weil/Tate pairings over elliptic curves [55]. More specifically, for a

Table 3.4: Simulation Settings

Parameter	Setting
perfect rooted k-ary tree of height h	$h = [1, 2, \cdots, 5, 6], k = 3$
transmission radius	$r = 30$ m
hop-by-hop communication delay	$T_{comm} = 300$ ms
computational delay of PCBA	$T_{agg} = 0.9$ ms
length of p in ELPDA	$\|p\| = 224$ bits
modulus of Paillier cryptosystem	$\|n\| = [2048]$ bits

given level of security, an elliptic curve requires shorter key lengths which in turn requires fewer memory and CPU resources to implement the proposed scheme. While PCBA has to use large modulus to guarantee its security, unfortunately the large modulus will cause the long communication delay.

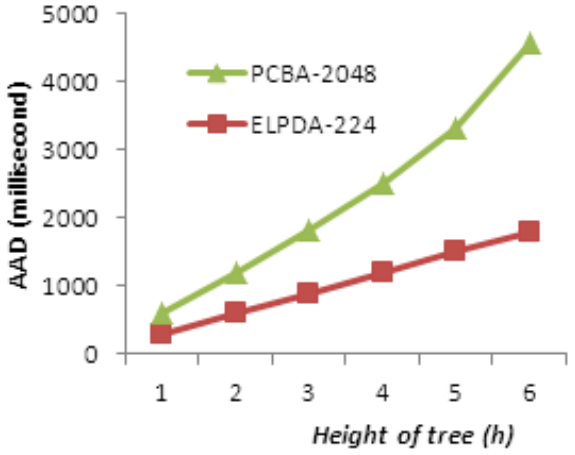

Figure 3.11: Average aggregation delay (AAD)

In Fig. 3.11, we compare the average aggregation delay (AAD) in the proposed ELPDA with 224 modulus and PCBA with 2048 modulus when the parameter h increases from 1 to 6. From the figure, we can see, with the increase of h, AAD in all schemes increase. However, the AAD of the proposed ELPDA is lower than that of the PCBA. The reason is that the computational delay cost can be almost negligible in the proposed ELPDA scheme, and the AAD is mainly dominated by the hop-by-hop communication delay.

3.6 Concluding Remarks

In this chapter, we have proposed an efficient lightweight privacy-preserving data aggregation scheme (ELPDA), including an efficient faulty smart meters detection algorithm for smart grid. Compared with the previously data aggregation schemes, the proposed scheme and algorithms can efficiently achieve privacy-preserving electricity consumption at a residential area. Meanwhile, the proposed scheme can detect faulty smart meters and dynamically rebuild data aggregation trees using the proposed algorithms.

Chapter 4

Security-Enhanced Data Aggregation against Malicious Gateways in Smart Grid

In smart grid, to monitor, predict and control the power consumption in real time, energy usage data have to be periodically collected through publicly accessible communication channels, and are stored in a centralized operation center. However, electricity consumption data may disclose the privacy information of users. Therefore, protecting privacy of users and validity of power usage reports becomes a crucial security issue. In this chapter, we propose a security-enhanced data aggregation scheme for smart grid communications based on homomorphic cryptosystem, trapdoor hash functions and homomorphic authenticators. The proposed scheme can achieve data confidentiality and integrity against the malicious aggregator (e.g. gateway), meaning that the aggregator is not able to access users' private information or corrupt the power consumption reports during the aggregation process. Through extensive analysis, we demonstrate that our scheme can resist potential threats and be proved secure under cryptographic hard assumptions. It has less computational and communication overheads than existing approaches.

4.1 Introduction

The world is undergoing the development of smart grid technology. The smart grid integrates the traditional grid with information and communication technologies, such as network communication, control systems and computation facilities, to achieve two-way electricity and information exchange between utilities and users while making the grid more reliable, efficient, secure and greener [59]. The smart grid, with these appealing advantages, is able to significantly avoid

the occurrence of electrical blackout, and automatically and quickly identify the disturbance in the electrical distribution so that repair crews can be immediately dispatched to the problem area [60].

Electric power systems are complex physical networks that control the electricity generation, transmission, distribution and consumption. Combining with the communication networks, the management and dispatch of the electricity become much smarter. The smart grid communication network deployed in parallel to the hierarchical power grid. Smart meters, critical two-way wireless or wired communication devices, are deployed at customers premise and used to collect the real-time power consumption periodically. The electricity usage and detailed information about the transmission links are centralized at the operation center (OA) that allow OA to manage energy consumption in real time and lower the need of customers in peak hours through adjusting the electrical price dramatically. OA also exposes the customers' detailed real-time electrical usage information to the power plants, which may help them to adjust the energy production and then reduce the need to fire up the costly and secondary power plans. The customers can access their own real-time usage information and decrease their electricity costs by shifting the uninterrupted activities from peak time to non-peak time.

While smart grid provides a numerous amount of appealing benefits toward both users and operation centers, it is still confronted with a great deal of cyber security threats [23, 37, 62–65]. It is of paramount importance to prevent the data from being eavesdropped, modified, forged and denied. Furthermore, from the perspective of users, privacy is a primary concern as it is possible to infer the users' daily activities, habits and other privacy witnessable references from the electricity data. For example, a relatively low and static daily consumption of a household may indicate that no one is at home [37]; a conspicuous drop of power consumption at midnight may indicate the house owner goes to sleep [10]; power variation every several hours throughout every night might indicate that this family has a new baby [66]. End-to-end encryption is a straightforward way to hide the communication content and preserve users' privacy, but at the same time will increase the data size and cause heavy overhead on communication channels.

To reduce the communication burden and keep data confidentiality, several data aggregation schemes [8, 10, 38, 41, 67, 68] have been proposed to compress the consumption reports in a specific residential area at a local gateway and then forward in a compact form to the operation center. These schemes can be proved secure against gateways under the honest-but-curious model, where the gateways honestly follow the communication protocols agreed upon among the ones involved, but snoop on users' electricity consumption out of curiosity. Unfortunately, in reality, a gateway may become malicious due to many reasons. For example, terrorists may insert false messages or viruses to mislead operation centers and make a lengthy blackout over an extensive region, and thereby may lead to the deaths of many people during a period of extreme weather [69]. Hackers can modify on-going traffic or communications in order to manipulate the

market price and encourage users to increase power consumption during on-peak periods [70]. As a result, malicious gateways are able to control the power dispatch and the market price; thus, ruin all the attractive benefits of smart grid. However, malicious gateways have not received any attention lately, and how to prevent malicious gateways still remains an open problem.

Therefore, there should be a mechanism to prevent both the individual usage reports and the aggregated one from being modified. While it is of great difficulty to construct an efficient universal homomorphic signature which can aggregate the signatures generated by multiple signers [71], the challenge is how to aggregate the reports and the corresponding signatures simultaneously at the gateways. To address this problem, we propose a security-enhanced data aggregation scheme that is able to avoid the attacks from compromised aggregators. By integrating the Paillier encryption [34] with the trapdoor hash function [72] and the homomorphic authenticators [73], our scheme can aggregate the authentication responses, ciphertexts and signatures at the same time, thus drastically reducing the communication and computational overheads as compared with existing protocols. Specifically, our contributions can be summarized as the following two aspects.

Firstly, inspired by the facts that the gateways may be compromised, we propose a security-enhanced data aggregation scheme from trapdoor hash functions, Paillier encryption and homomorphic authenticators. Our proposed scheme is the first one against malicious gateways and a successful attempt to construct authentication schemes from trapdoor hash functions with key exposure.

Secondly, we analyze the security strength and the performance of our proposed scheme. In particular, we employ provable security technique to reduce the security of our scheme to well-known mathematical hard problems and underlying cryptographic tools. Through the performance comparison, we demonstrate that our scheme is much more efficient than existing schemes in terms of both communication and computational overheads.

The remainder of the chapter is organized as follows. In section 4.2, we define the system model, security requirements and design goals. Then, we describe our proposed scheme in section 4.3 and provide the security analysis in section 4.4, respectively. We analyze the performance of our proposal in section 4.5. Finally, we draw our conclusion in section 4.6.

4.2 Models and Design Goals

In this section, we briefly discuss the system model, security requirements and design goals.

4.2.1 System Model

In our system model, we formalize the communications between the users and the operation center as depicted in Fig. 4.1. Specifically, a typical residential area (RA) has large amounts of residential users $U = \{U_1, U_2, \cdots, U_w\}$ deployed smart meters to collect their real-time electricity consumption. The users are required to report their usage information to the operation center (OC) every several minutes [62]. The smart meters are installed at the places under the control of the users, can remotely update the time and the cryptographic keys. Therefore, we assume the time periods of smart meters in the same RA is synchronous [67]. A local gateway (GW), which is deployed to connect with the OC and the smart meters in this area, mainly performs two functions: aggregation and relaying. When the GW receives the usage reports from the users through relatively inexpensive WiFi technologies, as suggested in [60], the aggregation component aggregates all the individual reports into a compressed one. Afterwards, the relaying component forwards it to the OC through wired network or other links that support long distance communication with low delay.

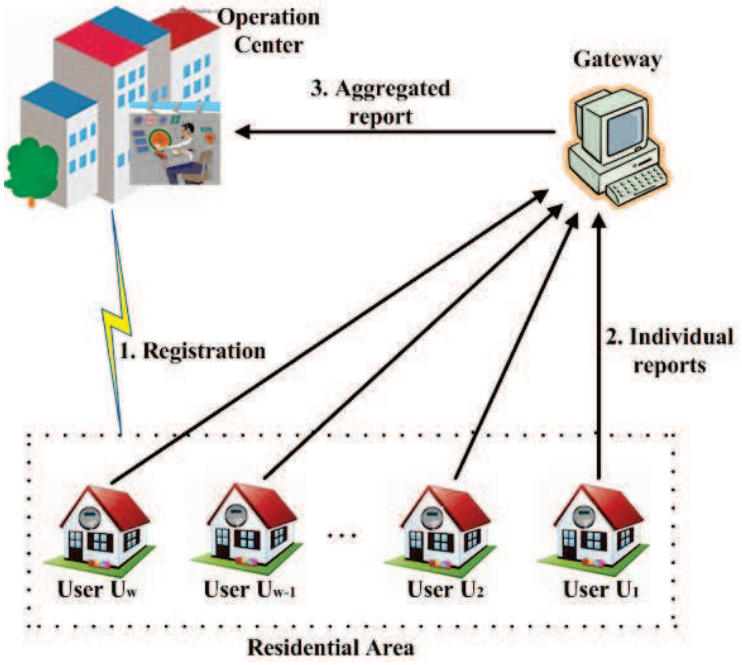

Figure 4.1: System model for data aggregation.(Courtresy of [10])

4.2.2 Security Requirements

In the security model, we assume that OC can be fully-trusted and the users $U = \{U_1, U_2, \cdots, U_w\}$ are honest as well. Smart meters are protected by the hardware, and it is difficult to recover the keys and the random numbers used. However, an attacker is able to eavesdrop on entire communication channels to capture users' usage reports. More seriously, powerful attackers, such as terrorists or hackers, can intrude the intermediate nodes on the channels, like gateways, to modify transmission data and steal individual usage reports. Therefore, in order to ensure that OC obtains valid results and preserves the privacy of users, the following secure requirements must be satisfied:

- *Authentication*: Assure that the electricity usage reports are really from the legal residential users and the users can not repudiate the existing reports. Thus, the attacker is not able to pretend to be a legal user generating an individual report.
- *Confidentiality*: Protect the users' consumption data from the attacker, even if the attacker can eavesdrop WiFi communication channel and corrupt the local gateways. In such way, the privacy of users and the contents of usage reports will not be disclosed during the transmission.
- *Integrity*: Prevent the usage reports from being modified by the attacker when they are transmitting on links, which means that any alternation on the reports, either the individual one or the compressed one, must be detected by OC when reading the report. Therefore, OC receives the legitimate result and ignores all the ill reports.

4.2.3 Design Goals

To enable security-enhanced data aggregation under the aforementioned system model and the security requirements, our scheme should achieve following objectives:

- *Security*: As mentioned above, three aspects of security requirements should be satisfied for our new scheme that ensures OC can receive the authentic and reliable information. The users' privacy should also not be disclosed as well. The proposed scheme should prevent the malicious gateways from modifying the usage reports, so that the overall consumption is available and trustful.
- *Efficiency*: In terms of efficiency, the computational cost is required to be low since smart meters are the devices that have limited computation power. There should be no time-consuming operations, such as pairing computation, on the users' side. In addition, the

new scheme should consider the communication-effectiveness, thus, OC can receive real-time usage reports in short delay.

4.3 Proposed scheme

In this section, we present our security-enhanced data aggregation scheme for smart grid communications by utilizing the trapdoor hash function [72], the Paillier cryptosystem [34] and the homomorphic authenticators [73], which mainly consists of five phases: system initialization, user registration, report generation, report aggregation and report reading.

4.3.1 System Initialization

In a single-authority smart grid system, OC acts as a trusted authority to bootstrap the whole system. In the system initialization phase, given the security parameters (κ, κ_1), OC firstly chooses three distinct larger primes (q_1, p, q) randomly, where $|q_1| = \kappa$ and $|p| = |q| = \kappa_1$. Then, OC computes the RSA modulus $n = pq$, Carmichael's function $\lambda = lcm(p-1, q-1)$ and chooses a generator $g \in \mathbb{Z}^*_{n^2}$. OC also generates a cyclic group $\mathbb{G}$ with the prime order q_1 and chooses additional generators $g_0, u \in \mathbb{G}$. In addition, OC defines a function $L(\gamma) = \frac{\gamma-1}{n}$, a cryptographic hash function $H : \{0,1\}^* \to \mathbb{G}$ and a pseudorandom function $f : \mathbb{G} \times \mathbb{G} \to \mathbb{Z}^*_{q_1}$. Assume that the number of households in the RA is w. In the end, OC releases the system parameters as

$$\text{Params}=\{q_1, \mathbb{G}, g_0, u, H, f, n, g, w\},$$

and keeps the master key $msk = (p, q, \lambda)$ privately.

4.3.2 User Registration

When a user of RA U_i joins the smart grid system, it firstly chooses a random number $x_i \in \mathbb{Z}^*_{q_1}$ as its private key and computes the corresponding public key as $h_i = g_0^{x_i} \in \mathbb{G}$. Then U_i picks $a_i, b_i \in \mathbb{Z}^*_{q_1}$ randomly, and generates a value of trapdoor hash as $\mathcal{H}_i = g_0^{a_i} h_i^{b_i}$. Finally, U_i sends $(h_i, \mathcal{H}_i)$ and its identifier ID_i to the OC.

Upon receiving the registration message from a new user, OC chooses a random value $k_i \in \mathbb{G}$ as a secret identifier of the user and picks a number $r_i \in \mathbb{Z}^*_{q_1}$ randomly to compute $e_i = g_0^{r_i}$ and $e_i^* = (k_i||\alpha) \cdot h_i^{r_i}$, where α is a secret and unique identifier for RA chosen by OC and distributed to

all the households in the RA securely. At last, OC computes $v_0 = g_0^{\alpha}$, stores $(ID_i, h_i, \mathcal{H}_i, k_i, v_0)$ and returns (e_i, e_i^*) to U_i.

With the secret key x_i, U_i can securely receive the secret identifier k_i and RA's secret identifier α by decrypting (e_i, e_i^*) as $k_i||\alpha = e_i^* \cdot (e_i)^{-x_i}$.

4.3.3 Report Generation

In order to achieve nearly real-time dispatch, a smart meter is deployed at the user's side to collect and report the electricity consumption every ρ minutes, e.g., $\rho = 15$ minutes. The smart meter collects the usage data m_i and performs the following steps to generate a consumption report:

- Compute $b_i' = f(k_i, t)$ and $a_i' = x_i \cdot (b_i - b_i') + a_i \ (mod\ q_1)$, where t is the synchronized system time period.
- Choose a random number $s_i \in \mathbb{Z}_{n^2}^*$ and compute

$$c_i = g^{m_i} \cdot s_i^n \bmod n^2.$$

- Then use the RA's secret identifier α to compute a tag:

$$\sigma_i = (H(ID_i||t) g_0^{a_i'} u^{m_i})^{\alpha}.$$

- Send the consumption report $P_i = ID_i||a_i'||c_i||\sigma_i||t$ to the local GW in the RA.

4.3.4 Report Aggregation

Upon receiving total w individual consumption reports $\{P_1, \cdots, P_w\}$ from the smart meters, the local GW aggregates the reports as follows:

$$a = \textstyle\sum_{i=1}^{w} a_i' \ mod\ q_1; \quad c = \prod_{i=1}^{w} c_i \ mod\ n^2; \quad \sigma = \prod_{i=1}^{w} \sigma_i.$$

Then, GW forwards the aggregated usage report $P = ID^*||a||c||\sigma||t$ to the OC, where ID^* is the identifier of the local GW.

4.3.5 Report Reading

After receiving $P = ID^*||a||c||\sigma||t$, OC performs the following steps to read the aggregated report P:

- Use each user's unique identifier k_i to compute $b_i^* = f(k_i, t)$ and then verify whether all reports are released by legitimate users by checking the following equation:

$$\prod_{i=1}^{w} \mathcal{H}_i \stackrel{?}{=} g_0^a \cdot \prod_{i=1}^{w} h_i^{b_i^*}.$$

  If it succeeds, continues to decrypt the ciphertext; Otherwise, aborts and outputs failure.

- Decrypt the aggregated ciphertext c as $m = \frac{L(c^\lambda \ mod \ n^2)}{L(g^\lambda \ mod \ n^2)} \ mod \ n$.

- Verify whether the following equation is valid:

$$\hat{e}(\sigma, g_0) \stackrel{?}{=} \hat{e}(\prod_{i=1}^{w} H(ID_i||t) g_0^a u^m, v_0).$$

If yes, accepts m, which is the sum of the electricity consumption for the households in the residential area; Otherwise, rejects m and outputs failure. If OC outputs failure in either of steps, it retrieves all individual reports to find the corrupted reports utilizing a recursive divide-and-conquer approach (binary search), as suggested by Ferrara et al. [74].

4.4 Security Analysis

This section will analyze the security properties of the proposed scheme. In particular, following the security requirements described in section 4.2, we will focus on the authentication, confidentiality and integrity.

• *Authentication*: In the new scheme, the trapdoor hash function is utilized to design the effective interaction of the authentication process. Since the trapdoor hash function is secure under the discrete logarithm assumption [72], the authenticity of the consumption reports can be achieved. Specifically, an attacker can not find a collision (a_i', b_i') of (a_i, b_i) to make $g_0^{a_i'} h_i^{b_i'} = g_0^{a_i} h_i^{b_i}$ hold without possessing the corresponding private key x_i in polynomial time with non-negligible success probability, unless the discrete logarithm problem is solvable. Therefore, the attacker's behaviors that aim at sending forged reports can be detected in the proposed scheme.

• *Confidentiality*: In order to prevent the users' electricity usage data from being revealed, we employ the Paillier cryptosystem to encrypt the data and aggregate the ciphertexts based on the property of the additive homomorphism. Since each ciphertext $c_i = g^{m_i} \cdot s_i^n \bmod n^2$ is valid and Paillier cryptosystem is semantic secure against the chosen plaintext attack [34], the confidentiality of the usage data m_i can be protected. Although an attacker eavesdrops on the communication channels, it is still unable to learn about the individual report m_i. When the GW receives all the reports from the residential users, it can not recover the reports but aggregating the ciphertexts directly and forwarding them to the OC. Thus, the GW is unable to acquire any information from both the separated reports and the aggregated one. At the OC's side, it decrypts the compressed report and obtains the sum of the power usage in a residential area rather than individual data. Therefore, the confidentiality of the usage data entirely depends on the semantic security of the Paillier encryption scheme, which can be reduced to the computational composite residuosity assumption [34].

• *Integrity*: In the user registration, OC utilizes the Elgamal encryption to distribute α to all the users in the residential area. Since the Elgamal encryption is semantic secure, only the users in RA are able to recover α. So α can be viewed as a secret key shared among the users. Then we should prove that the tags $\sigma_i (1 \leq i \leq w)$ are existentially unforgeable under an adaptive chosen message attack if the computational Diffie-Hellman (CDH) assumption holds.

Theorem 1: The tags $\sigma_i (1 \leq i \leq w)$ are existentially unforgeable under an adaptive chosen message attack, provided that the CDH assumption holds.

Proof: Assume that there is an adversary $\mathcal{A}$ who can break the existential unforgeability of the tags with a non-negligible advantage, then we can construct an algorithm $\mathcal{B}$ to solve the CDH problem.

Let g_0 be a generator of $\mathbb{G}$. Algorithm $\mathcal{B}$ is given $g_0, g_0^s, h \in \mathbb{G}$, its goal is to compute h^s. $\mathcal{B}$ simulates a challenger and interacts with the adversary $\mathcal{A}$ as follows.

- In generating a key, $\mathcal{B}$ sets the public key v_0 to g_0^s and sends it to $\mathcal{A}$.
- $\mathcal{B}$ programs the random oracle to respond the hash queries. To ensure the consistency of the responses, it maintains a lists of tuples to keep the queries. When receiving queries (ID_i, t) from $\mathcal{A}$, $\mathcal{B}$ chooses a random $x \in \mathbb{Z}_{q_1}^*$ and returns g_0^x.
- $\mathcal{B}$ also programs the tag oracle to respond the tag queries and a list of tuples are kept to maintain the interaction messages. $\mathcal{B}$ chooses random values $\beta, \gamma \in \mathbb{Z}_{q_1}^*$ to set $u = g_0^\beta h^\gamma$. Assume (ID_i, t, a_i', m_i) is the tag query issued by $\mathcal{A}$, $\mathcal{B}$ picks a random $x_i \in \mathbb{Z}_{q_1}^*$ and programs the random oracle as

$$H(ID_i||t) = g_0^{x_i}/(g_0^{a_i'+\beta m_i} \cdot h^{\gamma m_i}).$$

So $\mathcal{B}$ computes

$$\sigma_i = (H(ID_i||t) \cdot g_0^{a_i'} u^{m_i})^s = (g_0^s)^{x_i}.$$

- Eventually, $\mathcal{A}$ produces a tag σ on the data (ID, t, a, m). Assume that σ is a valid tag on m under the given public key; Otherwise, $\mathcal{B}$ reports failure and aborts. So the response satisfies the verification equation, i.e.,

  $$\hat{e}(\sigma, g_0) = \hat{e}(\prod_{i=1}^{w} H(ID||t) \cdot g_0^a u^m, v_0).$$

  Let the expected tag, which would have been obtained from the honest signers, be σ' on the data (ID, t, a', m'). The expected tag also satisfies the verification equation, i.e.,

  $$\hat{e}(\sigma', g_0) = \hat{e}(\prod_{i=1}^{w} H(ID||t) \cdot g_0^{a'} u^{m'}, v_0).$$

  If $a = a'$ and $m = m'$, then $\sigma = \sigma'$. Thus, if we define that $\Delta a = a - a'$ and $\Delta m = m - m'$, there must be the case that at least one of Δa and Δm is nonzero.

  - If $\sigma \neq \sigma'$, we divide the verification equation for σ by the equation for σ' and obtain

    $$\hat{e}(\tfrac{\sigma}{\sigma'}, g_0) = \hat{e}(g_0^{\Delta a + \beta \Delta m} h^{\gamma \Delta m}, v_0).$$

    Rearranging the equation yields

    $$\hat{e}(\tfrac{\sigma}{\sigma'} \cdot v_0^{\Delta a + \beta \Delta m}, g_0) = \hat{e}(h^{\gamma \Delta m}, v_0).$$

    Since $v_0 = g_0^s$, we have

    $$h^s = (\tfrac{\sigma}{\sigma'} \cdot v_0^{\Delta a + \beta \Delta m})^{\frac{1}{\gamma \Delta m}},$$

    which is the solution to the CDH problem.
  - Otherwise, we get $g_0^a u^m = g_0^{a'} u^{m'}$, and that

    $$1 = g_0^{\Delta a + \beta \Delta m} h^{\gamma \Delta m}.$$

    So we can solve the discrete logarithm problem:

    $$h = g_0^{-\frac{\Delta a + \beta \Delta m}{\gamma \Delta m}}.$$

  Since the hardness of the CDH problem implies the hardness of the discrete logarithm problem, the existential unforgeability can be reduced to the CDH problem.

In summary, the adversary, even the malicious GW, is not able to modify the electricity usage reports when they are transmitted on the communication channels.

4.5 Performance Evaluation

In this section, we evaluate the performance of our proposed scheme in terms of the computational performance and the communication overhead.

4.5.1 Computational Performance

When a residential user U_i registers for the smart grid system, it requires three exponentiation operations in $\mathbb{G}$ to generate its public key and the trapdoor hash value. After receiving the registration message, OC needs three exponentiation operations in $\mathbb{G}$ to encrypt k_i and α and calculate v_0. To decrypt the ciphertext, the U_i should perform one exponentiation operation in $\mathbb{G}$. When the users in the residential area generate the electricity consumption reports, two exponentiation operations in $\mathbb{Z}_{n^2}$ and 3 exponentiation operations in $\mathbb{G}$ are required for each smart meter. To aggregate the reports, GW only needs to perform inexpensive multiplication operations to aggregate the reports and forward the result to the OC. Before getting the sum of the consumption, the OC verifies the validation of the aggregated tag, which needs to execute 2 exponentiation operations in $\mathbb{G}$ and two pairing operations. Then, it performs $w+1$ exponentiation operations in $\mathbb{G}$ and two exponentiation operations in $\mathbb{Z}_{n^2}$ to check the collision of the trapdoor hash functions and decrypt the aggregated ciphertext of the usage data.

We present the computational performance comparison of our security-enhanced data aggregation (SEDA) scheme and three schemes, namely, EPPA (Lu *et al.* [10]), Ohara14 (ohara *et al.* [39]) and Fan14 (Fan *et al.* [38]), which are designed from additive homomorphic encryption schemes as well. The implementation is conducted on a notebook with Intel Core i5-4200U CPU @1.6 GHz, 2.29GHz and the memory is 4.00 GB. We use the MIRACL library to implement number-theoretic based methods of cryptography. The RSA modulus n is approximately 1024 bits and the parameter q_1 is 160 bits. The time costs of four algorithms are shown in Table 4.1.

Table 4.1: Comparison of Time Costs (Unit:ms)

Matric	SEDA	EPPA [10]	Fan14 [38]	Ohara14 [39]
Rep. Gene. Time	23.7	26.2	7.3	96.2
Rep. Aggr. Time	5.3	4906.3	6737.6	4.1
Rep. Read. Time	421.8	97.7	890.5	943.1

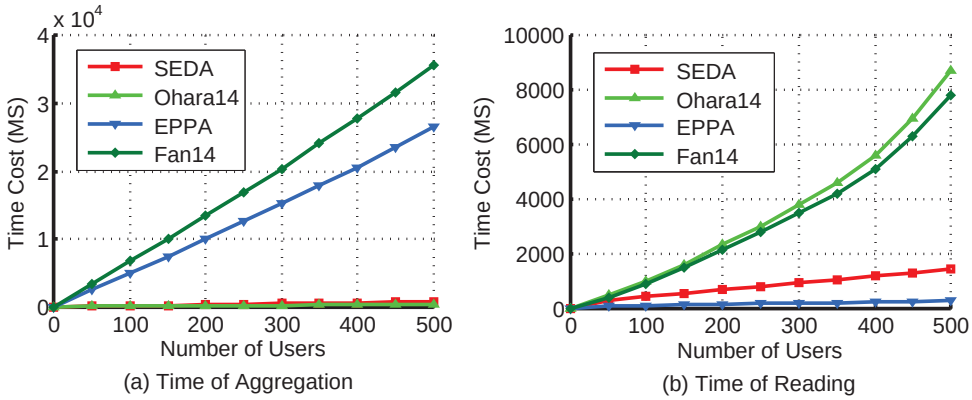

Figure 4.2: Computational overhead comparison

Fig. 4.2(a) shows the comparison results of four schemes on the aggregation computational costs. From the figure, we can see that our scheme is much more efficient than Lu's *et al.* EPPA scheme [10] and Fan's *et al.* scheme [38], since the time costs of the aggregation algorithms in their schemes increase significantly as the number of users increases. Ohara's *et al.* scheme [39] costs less time than ours, but its report reading algorithm needs the longest time in four schemes.

Fig. 4.2(b) shows the comparison results in terms of report reading computational cost. Among these schemes, Fan's *et al.* scheme [38] and Ohara's *et al.* scheme [39] cost plenty of time to decrypt the aggregated report when the number of the users is large. Although Lu's *et al.* scheme [10] is slightly more efficient than our scheme in the report reading process, their scheme poses a risk of report forging attack from malicious GW.

4.5.2 Communication Overhead

The communication of the scheme composes of user-to-GW communication and GW-to-OC communication. In the user-to-GW communication, each user generates its electricity usage data $P_i = ID_i||a_i'||c_i||\sigma_i||t$ and sends them to the GW, which is of binary length $S_{UG} = |ID_i| + 160 + 2048 + 1024 + |t|$, if n is 1024 bits and q_1 is 160 bits. So at the GW's side, the communication overhead between users and GW is $w * S_{UG}$, assume there are w users in a specific residential area. In report aggregation process, the GW aggregates the reports to compute $P = ID^*||a||c||\sigma||t$, indicating that the communication overhead between GW and OC is significantly reduced. Specifically, the overhead of GW-to-OC communication decreases from $(|ID^*|+160+2048+1024+|t|)*w$ bits to $S_{GO} = |ID^*|+160+2048+1024+|t|$ bits. Furthermore, we plot the communication overheads of three schemes with respect to the users' number

w. Since Fan *et al.* [38] do not consider the communication between GW and OC, we only show the overhead from three schemes in Fig. 4.3. It can be seen that Ohara's *et al.* scheme [39] causes the heaviest burden on the channel and the communication overhead of our scheme is almost equivalent to that of Lu's *et al.* scheme [10].

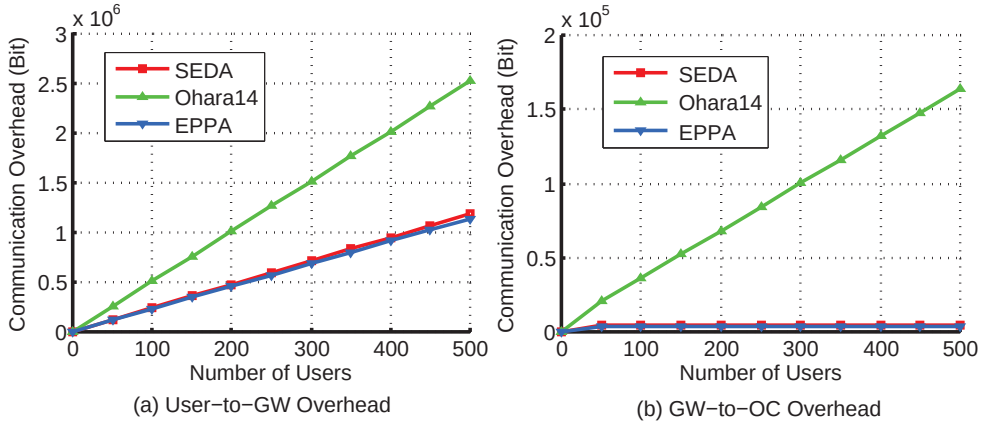

Figure 4.3: Communication overhead comparison

4.6 Concluding Remarks

In this chapter, we have proposed an efficient and secure data aggregation scheme for achieving the real-time electrical measurements collection for smart grid communications based on trap-door hash functions, Paillier cryptosystem and homomorphic authenticators that can resist the attacks from malicious gateways. We have also provided security analysis to demonstrate security strength and privacy-preserving capacity. Compared with existing data aggregation schemes for smart gird, our scheme has lighter the computational overhead and lower the communication cost. For our future work, we will study the methods to design the data aggregation schemes that are able to detect and trace the misbehaviors of legal users.

Chapter 5

A Privacy-Preserving Data Sharing Framework for Smart Grid

Distributed energy resources, featured with small-scale power generation technologies and renewable energy sources, are considered as necessary supplements for smart grid. To ensure that merged resources contribute effectively to the grid, data generated by consumer side should be shared among the energy resources. However, it also introduces challenges of the protection of consumer privacy. To address these difficulties, in this chapter we propose a new framework to share data in smart grid by leveraging new advances in homomorphic encryption and proxy re-encryption. Our proposed framework allows energy resources to analyze consumer data while ensuring consumer privacy. An additional benefit of our proposed framework is that consumer data is transmitted over the smart grid only once. Furthermore, we present a concrete scheme falling into the proposed framework. Extensive analysis shows that the concrete scheme is secure and efficient.

5.1 Introduction

There have been several instances when power grids across the globe risked catastrophic failure [2]. Oftentimes, power outages are caused by localized defects in the electricity networks. If a small defect is not dealt in a proper and timely manner, it could lead to a cascading failure of the power supply network. For example, a power outage on the east coast of the United States and Canada in 2003 was such a case. A power line was damaged by a tree in the Cleveland, USA. Making matters worse is that nearby lines became overloaded and overheated by rerouted

power and sagged from the excessive heat. This eventually tripped circuit breakers after these lines contacted trees. Approximately 50 million people in the Northeast US and part of Canada were left without power for several days [5]. Power outages can also be caused by overloaded electrical circuits. Electricity consumption is higher during hotter summer days. In some cases electrical demands may exceed power grid capacity. In such cases appliances should be turned off to conserve energy or additional resources should be added to grid to compensate demands. If left unaddressed, an overloaded power grid could fail, resulting in blackouts. It is thus crucial that we monitor power grid systems in real-time to ensure that abnormalities are dealt with promptly and effectively.

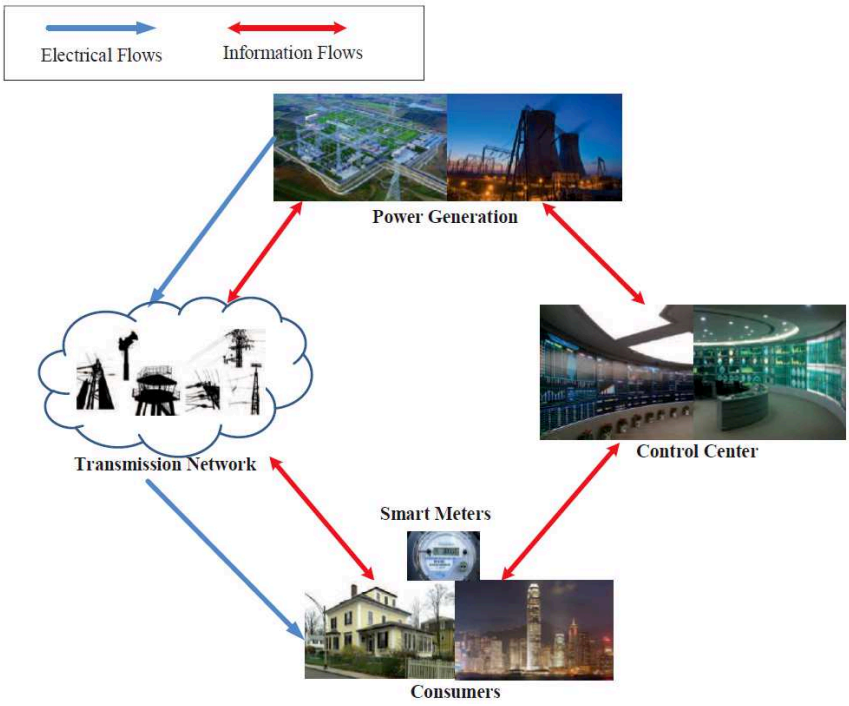

Figure 5.1: Communication architecture for smart grid

Smart grids have recently been gaining popularity. They support real-time diagnosis and can react to avoid failures and blackouts [10, 37, 68, 75, 76]. The major difference between the smart and traditional power grids is information flows. In the traditional power grid, there exists only one-way electrical flows, i.e., electricity utilities only deliver power to consumers. In contrast, smart grids allow for two-way information flow communications. As shown in Fig. 5.1, the two-way information flows in the smart grid are almost parallel to that of the one-way power flows. However, a control center is also involved in information flows. A control center collects data

with which it can decide on how to alter a grid. With electricity consumption reports a control center can analyze consumers electricity consumption data and forecast electricity consumption, and adjust power generation accordingly over a given period. Regular electricity consumption reports are key in smart grid efficacy.

On the other hand, the collected electricity consumption reports should be secured to preserve consumer privacy [15, 37]. To operate reliable and resilient smart grids it is paramount that we address security and privacy concerns through established cryptographic schemes. Traditional encryption schemes are quite suitable for single energy source configurations. In these cases, a single control center, often controlled by a government or government affiliated party can be counted on as a trusted confidant. This can become troublesome for more complex arrays where supplemental sources from small-scale generation and renewable energy projects come into play [18, 19]. Such grids are referred to as "micro-grids". Not all the distributed energy resources are under direct government control. Therefore, trusting these entities is questionable. A dilemma of balancing between consumer privacy and enabling energy resources to freely analyze records becomes apparent.

A trivial solution is to anonymize data before sending it to energy resources for analysis. However, it would significantly increase the communication costs; massive anonymized data needs to be sent to every resource. An alternative is to let some third party perform analysis instead of energy resources. Only analysis results are sent to energy resources. Nevertheless, the third party would be privy to analysis results. This is undesirable by competing businesses and privacy advocates. To the best of our knowledge, there is no efficient approach so far to this problem in the context of privacy-preserving smart grid. In this chapter, we aim to address the above challenge and propose a framework for data sharing in smart grid. The contributions of this chapter are twofold.

- First, we propose a novel data sharing framework for smart grid, where we combine the two popular infrastructures: the smart grid and cloud computing. In particular, we allow the electricity consumption reports generated in smart grid to be stored in the cloud. Distributed energy resources can obtain the statistics and analysis results from the cloud computing. Hence, our proposed framework can take advantage of cloud computing for the smart grid.

- Second, our proposed framework makes use of the homomorphic encryption technique to facilitate the statistics and analysis on the encrypted electricity consumption reports, and the proxy re-encryption technique to keep the statistics and analysis results secret from the cloud.

The remainder of the chapter is organized as follows: In Section 5.2, we present the system model, security model, and the design goal. Then, we propose the data sharing framework in Section 5.3, followed by security analysis and performance evaluation in Section 5.4. Finally, we draw our conclusion in Section 5.5.

5.2 Models and Design Goals

In this section, we present the system model, security model, and design goals.

5.2.1 System Model

In this chapter, we only focus on how the electricity consumption reports are securely shared among the distributed generation resources. In particular, we take advantage of the data-as-a-service (DaaS) model in cloud computing, where the system is composed of the following parties: the trusted authority (TA), many electricity consumers (ECs), many energy resources (ERs) and the cloud server as shown in Fig. 5.2. The TA is responsible for generating the system parameters and the certificate for the public key of each ER. The ECs produce the electricity consumption reports that are outsourced to the cloud server. To achieve the confidentiality, the electricity consumption reports should be encrypted by using the public key of the corresponding ER where the consumed electricity comes from. In order to make a smart decision on the power generation, price and others, each ER would like to do analysis on the electricity consumption reports corresponding to itself or other ERs. Before doing the analysis, the ER should obtain the analysis rights from other ERs.

5.2.2 Security Model

We assume that the cloud server is honest-but-curious as many literatures related to cloud computing [44–46]. That is to say, the cloud server will follow the proposed framework faithfully, but could launch passive attacks to get secret information as much as possible. In particular, the cloud server is interested in getting the content of electricity consumption reports or analysis results, but they won't modify the communication data with other entities or collude with other entities. The ERs want to get the analysis results from the cloud server so that they can plan more effectively and efficiently to produce an adequate supply of electricity to serve their local needs. It plays an important role in making our power grid more reliable and resilient since we must balance the grid by matching electricity supply with demand exactly to avoid power grid failure

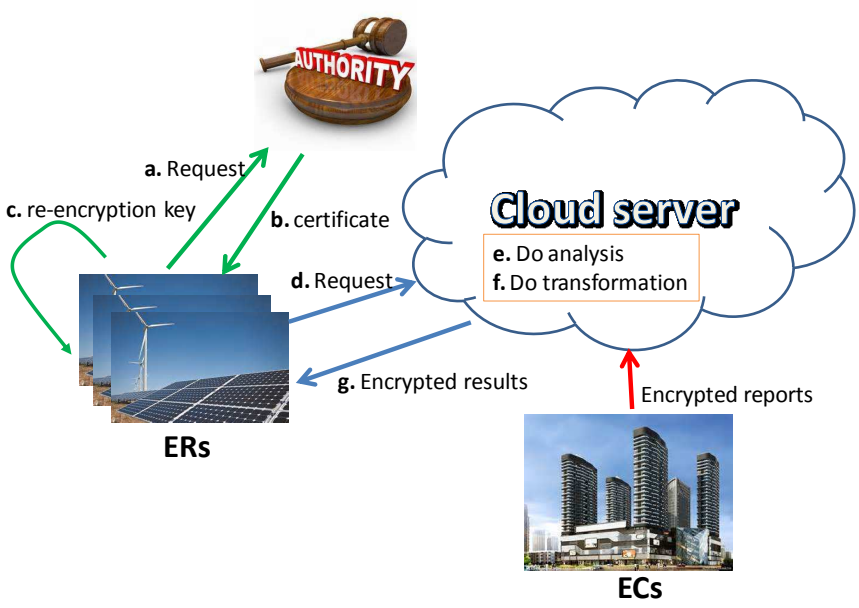

Figure 5.2: The proposed data sharing framework for smart grid

and blackout. Meanwhile, malicious ERs may try to access electricity consumption reports or get the analysis results beyond their analysis rights.

5.2.3 Design Goals

Our design goal is to develop a data sharing framework for smart grid. It has the following desirable properties.

- First, we propose a novel data sharing framework for smart grid, where we combine the two popular infrastructures: the smart grid and cloud computing. In particular, we allow the electricity consumption reports generated in smart grid to be stored in the cloud, and the distributed energy resources can obtain the statistics and analysis results from the cloud computing. Hence, our proposed framework can take advantage of cloud computing for smart grid.
- Second, our proposed framework makes use of the homomorphic encryption technique to facilitate the statistics and analysis on the encrypted electricity consumption reports, and

the proxy re-encryption technique to keep the statistics and analysis results secret from the cloud.

- *DaaS model*: The proposed framework takes advantage of the DaaS model in cloud computing, which can save a good amount of hardware and software maintenance cost for smart grid. Furthermore, the electricity consumption reports do not need to transmit over the network after the analysis request from the ER, which saves the communication cost in smart grid.
- *Privacy preservation*: The proposed framework should achieve privacy requirements of ECs. In particular, i) the electricity consumption reports stored in the cloud server cannot be revealed to anyone except the corresponding ER; and ii) the analysis results cannot be revealed to the one who has no corresponding analysis rights.

5.3 Proposed Data Sharing Framework

In this section, we present our data sharing framework, which consists of four parts: system initialization, reports creation, analysis grant, and reports analysis. Before plugging into the framework detail, we first need to review the preliminaries, including Bilinear groups, homomorphic encryption and proxy re-encryption, which will serve as the basis of our proposed framework.

5.3.1 Preliminaries

1) Bilinear Groups

Let $\mathbb{G}$ and $\mathbb{G}_T$ be two multiplicative cyclic groups of prime order q. They are equipped with an admissible bilinear map $e : \mathbb{G} \times \mathbb{G} \rightarrow \mathbb{G}_T$, such that $e(g_1^a, g_2^b) = e(g_1, g_2)^{ab}$ for all $a, b \in \mathbb{Z}_q$ and any $g_1, g_2 \in \mathbb{G}$. We denote `BSetup` as an algorithm that, on the input of security parameter λ, outputs the parameters $(\mathbb{G}, \mathbb{G}_T, q, g, e)$, where $q \in \Theta(2^\lambda)$.

2) Homomorphic Encryption

Homomorphic encryption [78] is a special form of encryption that allows anyone with ciphertexts of messages $(m_1, \cdots, m_t)$ to output a ciphertext of message $f(m_1, \cdots, m_t)$ for some desired function f without knowing the decryption key. If the function f could be any function,

then the homomorphic encryption is a fully homomorphic encryption. A concrete homomorphic encryption scheme is composed of the following four algorithms.

- HE.KeyGen: The key generation algorithm is a randomized algorithm that takes a security parameter λ as input, and outputs the public/private key pair (pk, sk).
- HE.Enc: The encryption algorithm is a randomized algorithm that takes public key pk and a message m from the message space $\mathcal{M}$ as input, and outputs the ciphertext c.
- HE.Dec: The decryption algorithm is a deterministic algorithm that takes the private key sk and a ciphertext c as input, and outputs the corresponding message m.
- HE.Eva: The evaluation algorithm is a (possibly randomized) algorithm that takes the public key pk, a set of ciphertexts on messages $(m_1, \cdots, m_t)$, and a evaluation function f as input, and outputs the ciphertext c on $f(m_1, \cdots, m_t)$.

The correctness of a homomorphic encryption scheme should satisfy the following two requirements for $\texttt{HE.KeyGen}(\lambda) \rightarrow (pk, sk)$.

$$\texttt{HE.Dec}(sk, \texttt{HE.Enc}(pk, m)) = m, \text{ and}$$

$$\texttt{HE.Dec}(sk, \texttt{HE.Eva}(pk, \{\texttt{HE.Enc}(pk, m_i)\}_{i=1}^{t}, f)) = f(m_1, \cdots, m_t).$$

3) Proxy Re-encryption

Proxy re-encryption [79] is a special kind of public key encryption, which allows a semi-trusted proxy with some information to transform a ciphertext under one public key into another ciphertext under another public key. However, the corresponding message cannot be revealed during the transformation process. A concrete proxy re-encryption scheme is composed of the following five algorithms.

- PRE.KeyGen: The key generation algorithm is a randomized algorithm that takes a security parameter λ as input, and outputs the public/private key pair (pk, sk).
- PRE.Enc: The encryption algorithm is a randomized algorithm that takes public key pk and a message m from the message space $\mathcal{M}$ as input, and outputs the ciphertext c.

- PRE.Dec: The decryption algorithm is a deterministic algorithm that takes the private key sk and a ciphertext c as input, and outputs the corresponding message m.
- PRE.ReKey: The re-encryption key algorithm is a (possibly randomized) algorithm that takes one public/private key pair (pk_1, sk_1) and another public key pk_2 as input, and outputs the corresponding re-encryption key rk.
- PRE.ReEnc: The re-encryption algorithm is a (possibly randomized) algorithm that takes a ciphertext c_1 under public key pk_1, and a re-encryption key rk corresponding to the delegation from pk_1 to pk_2 as input, and outputs the ciphertext c_2 under public key pk_2.

The correctness of a proxy re-encryption scheme should satisfy the following two requirements for any $\texttt{PRE.KeyGen}(\lambda) \rightarrow (pk, sk)$.

$$\texttt{PRE.Dec}(sk, \texttt{PRE.Enc}(pk, m)) = m, \text{ and}$$

$$\texttt{PRE.Dec}(sk_2, \texttt{PRE.ReEnc}(\texttt{PRE.ReKey}(pk_1, sk_1, pk_2), c_1)) = m,$$

where c_1 is the ciphertext corresponding to m under public key pk_1, which can be from algorithm PRE.Enc or algorithm PRE.ReEnc.

5.3.2 Main Idea

In order to protect consumers' privacy in the electricity consumption reports, these reports should be encrypted before uploading to the cloud server under the corresponding ER's public key. To respond an analysis request from an ER, the cloud server does the following steps.

- First, the cloud server makes use of homomorphic encryption technique to do the statistics and analysis on the electricity consumption reports encrypted under the *same* ER's public key without revealing the content of the electricity consumption reports or the obtained analysis result. We call the obtained analysis result as *meta-result*.
- Second, the cloud server takes advantage of proxy re-encryption technique to transform the meta-result encrypted under other ER's public key to the one encrypted under the public key of the requesting ER without revealing the content of the meta-result.
- Third, the cloud server makes use of homomorphic encryption technique once more. In particular, it does the statistics and analysis on the meta-result under the public key of the requesting ER without revealing the content of the meta-result or the final analysis result. It is easy to see that the requesting ER can obtain the content of the final analysis result by using its own private key.

From the above steps, we can see that only the encrypted final analysis is required to be sent from the cloud server to the requesting ER. This method allows the smart grid to save the network bandwidth as opposed to sending the anonymized report to each distributed energy resource, which is widely used in practice today.

5.3.3 Description of the Proposed Framework

In this subsection, we will give our framework for the smart grid based on the homomorphic encryption and proxy re-encryption. Note that we require the key space and ciphertext space of the underlying homomorphic encryption scheme and proxy re-encryption scheme be the same; otherwise, the proposal cannot work well.

1) System Initialization

In this phase, the TA generates the system parameters, including the security parameter λ, the ER generates its own public/private key pair (pk, sk) by running $\texttt{PRE.KeyGen}(\lambda)$, and the public key pk is implemented in the device (such as smart meter) of the EC who will consume the electricity generated by the ER. Furthermore, each ER should obtain the certificate $cert$ for its public key from the TA.

2) Reports Creation

Before uploading the electricity consumption reports to the cloud server, the consumer encrypts the electricity consumption reports m by running $\texttt{PRE.Enc}(pk, m) \rightarrow c$, where pk is the public key of the corresponding ER. The encrypted reports are stored on the cloud server in the format as shown in Fig. 5.3, where `ID` contains the necessary information to identify the reports, such as address and the ER's identity information.

ID	$\texttt{HE.Enc}(pk, m)$

Figure 5.3: Format of an encrypted report stored on the cloud server

3) Analysis Grant

In this phase, the requesting ER (denoted as A) will obtain the analysis rights from other ERs (denoted as $\{B_1, \cdots, B_n\}$, where n is a positive integer). Besides these entities, the cloud server will also be involved in this phase. ER A will interact with each ER in $\{B_1, \cdots, B_n\}$ as follows.

- ER A sends the request containing $\mathcal{ID}$ and evaluation function f to ER B_i ($i \in \{1, \cdots, n\}$), where $\mathcal{ID}$ shows the range of reports it wants to analyze, and the certificate of ER A's public key.
- ER B_i ($i \in \{1, \cdots, n\}$) first checks the validity of the certificate of ER A's public key. If no, ER B_i ($i \in \{1, \cdots, n\}$) aborts this phase; otherwise, it then decides whether ER A could have the analysis rights on the reports corresponding to $\mathcal{ID}$ by using the evaluation function f. If yes, ER B_i ($i \in \{1, \cdots, n\}$) generates the re-encryption key rk_i corresponding to (pk_{B_i}, pk_A) by running $\texttt{PRE.ReKey}(pk_{B_i}, sk_{B_i}, pk_A)$, and sends (rk_i, pk_{B_i}, pk_A) to the cloud server via a secure and authenticated channel. Note that pk_{B_i} ($i \in \{1, \cdots, n\}$) and pk_A are the public keys of ER A and ER B_i respectively.
- Upon receiving (rk_i, pk_{B_i}, pk_A) from ER B_i ($i \in \{1, \cdots, n\}$), the cloud server records the values with an encryption format in the re-encryption key list.

4) Reports Analysis

In this phase, the ER A obtains the statistics and analysis results from the cloud server as follows.

- The ER A sends its certificate $cert$, $\mathcal{ID}$, f and $\{B_1, \cdots, B_n\}$ to the cloud server via an authenticated channel. In the following steps, we implicitly set $i = 1, 2, \cdots, n$.
- Upon receiving the information from the ER, the cloud server checks the validity. If it is valid, the cloud server does the next step; otherwise, it aborts this phase.
- The cloud server collects the ciphertext set $\mathcal{C}_i$ corresponding to $\mathcal{ID}$ and B_i, and computes the encrypted meta-results $c_{B,i}$ by running $\texttt{HE.Eva}(pk_{B_i}, \mathcal{C}_i, f)$.
- The cloud server finds the re-encryption key rk_i corresponding to pk_{B_i} and pk_A from the re-encryption key list.
- The cloud server transforms the ciphertext $c_{B,i}$ under public key pk_{B_i} into another ciphertext $c_{A,i}$ under public key pk_A by running $\texttt{PRE.ReEnc}(rk_i, c_{B,i})$.
- The cloud server computes the encrypted final analysis $\hat{c}$ by running

$$\texttt{HE.Eva}(pk_A, \{c_{A,1}, c_{A,2}, \cdots, c_{A,n}\}, f).$$

- At last, the ER A can obtain the final analysis result from $\hat{c}$ by running $\texttt{PRE.Dec}(sk_A, \hat{c})$.

5) Discussion on the Proposed Framework

As we mentioned before, we require the key space and ciphertext space of the underlying homomorphic encryption scheme and proxy re-encryption scheme be the same; otherwise, the proposal cannot work. However, none of the existing homomorphic encryption schemes or proxy re-encryption schemes satisfy the above two conditions naturally. In this subsection, we show that homomorphic encryption schemes can be transformed to the satisfied one by an example using ElGamal-type encryption that only supports multiplicative homomorphism. This ElGamal-type encryption scheme can be found in [55].

The system parameters are $(\mathbb{G}, \mathbb{G}_T, q, g_1, g_2) \leftarrow \texttt{BSetup}(1^\lambda)$.

- HE.KeyGen (PRE.KeyGen): Randomly choose x from $\mathbb{Z}_q$, and set the key pair as $(pk, sk) = ((pk_1 = e(g_1, g_2)^x, pk_2 = g_2^{1/x}), x)$.
- HE.Enc (PRE.Enc): When inputting a message m from $\mathbb{G}_T$ and a public key $pk = (pk_1, pk_2)$, it outputs the ciphertext $c = (c_1, c_2)$ as follows.

$$c_1 = g_1^r, \;\; c_2 = pk_1^r \cdot m = e(g_1, g_2)^{x \cdot r} \cdot m$$

  where r is a random element from $\mathbb{Z}_q$.

- PRE.ReKey: When inputting one key pair $(pk_A, sk_A, pk_B) = ((pk_{A,1}, pk_{A,2}), sk_A, sk_A, (pk_{B,1}, pk_{B,2}))$, it outputs $rk = pk_{B,2}^{sk_A} = g_2^{x_A/x_B}$ as the corresponding re-encryption key.
- PRE.ReEnc: When inputting a ciphertext $c_A = (c_{A,1}, c_{A,2}) = (g^r, pk_{A,1}^r \cdot m)$ and $rk = g_2^{x_A/x_B}$ corresponding to the delegation from pk_A to pk_B, it outputs the re-encrypted ciphertext $c_B = (c_{B,1}, c_{B,2})$, where

$$\begin{aligned} c_{B,1} &= e(c_{A,1}, rk) \cdot e(g_1, g_2)^\alpha \\ &= e(g_1, g_2)^{r \cdot x_A/x_B + \alpha} \\ &= e(g_1, g_2)^{\hat{r}}, \\ c_{B,2} &= c_{A,2} \cdot pk_{B,1}^\alpha \\ &= pk_{B,1}^{r \cdot x_A/x_B + \alpha} \cdot m \\ &= pk_{B,1}^{\hat{r}} \cdot m, \end{aligned}$$

  α is a random element from $\mathbb{Z}_q$, and $\hat{r} = r \cdot x_A/x_B + \alpha \bmod q$.

- HE.Eva: When inputting two ciphertexts c and c', it outputs evaluated ciphertext $\hat{c}$ as follows.
  - If $c = (c_1, c_2) = (g_1^r, pk_1^r \cdot m)$ and $c' = (c'_1, c'_2) = (g_1^{r'}, pk_1^{r'} \cdot m')$, the evaluated ciphertext $\hat{c} = (\hat{c}_1, \hat{c}_2)$ is as follows.

$$\hat{c}_1 = c_1 \cdot c'_1 \cdot g_1^{\alpha} = g_1^{r+r'+\alpha} = g_1^{\hat{r}},$$

$$\hat{c}_2 = c_2 \cdot c'_2 \cdot pk_1^{\alpha} = pk_1^{r+r'+\alpha} \cdot m_1 \cdot m_2 = pk_1^{\hat{r}} \cdot m_1 \cdot m_2,$$

    where α is a random element from $\mathbb{Z}_q$, and $\hat{r} = r + r' + \alpha \bmod q$.
  - If $c = (c_1, c_2) = (e(g_1, g_2)^r, pk_1^r \cdot m)$ and $c' = (c'_1, c'_2) = (e(g_1, g_2)^{r'}, pk_1^{r'} \cdot m')$, the evaluated ciphertext $\hat{c} = (\hat{c}_1, \hat{c}_2)$ is as follows.

$$\hat{c}_1 = c_1 \cdot c'_1 \cdot g_1^{\alpha} = e(g_1, g_2)^{r+r'+\alpha} = e(g_1, g_2)^{\hat{r}},$$

$$\hat{c}_2 = c_2 \cdot c'_2 \cdot pk_1^{\alpha} = pk_1^{r+r'+\alpha} \cdot m_1 \cdot m_2 = pk_1^{\hat{r}} \cdot m_1 \cdot m_2,$$

    where α is a random element from $\mathbb{Z}_q$, and $\hat{r} = r + r' + \alpha \bmod q$.
- PRE.Dec: When inputting a ciphertext c under pk and a private key sk, it outputs m as follows.
  - If $c = (c_1, c_2) = (g_1^r, pk_1^r \cdot m)$, the message m is computed as $m = c_2 / e(c_1, g_2)^{sk} = (e(g_1, g_2)^{sk})^r \cdot m / e(g_1^r, g_2)^{sk}$.
  - If $c = (c_1, c_2) = (e(g_1, g_2)^r, pk_1^r \cdot m)$, the message m is computed as $m = c_2 / c_1^{sk} = (e(g_1, g_2)^{sk})^r \cdot m / (e(g_1, g_2)^r)^{sk}$.

Note that we do not need the algorithm HE.Dec or PRE.Dec with the case of $c = (c_1, c_2) = (g_1^r, pk_1^r \cdot m)$ in our framework, since the ER always do the decryption on the ciphertext with the format $c = (c_1, c_2) = (e(g_1, g_2)^r, pk_1^r \cdot m)$.

As we can see that there are two kinds of ciphertexts in the above scheme. One is the ciphertexts can be re-encrypted, the other is the ciphertexts cannot be re-encrypted any more. We respectively name them as the original ciphertexts and re-encrypted ciphertexts. Hence, we have two theorems for the security of the above scheme.

Theorem 1. *The above scheme is chosen plaintext secure under the eDBDH assumption for the original ciphertext. The eDBDH assumption assumes that given $(g, g^{b/a}, g^a, g^b, g^c) \in \mathbb{G}$ and $T \in \mathbb{G}_T$, it is hard for any polynomial probabilistic time adversary to decide whether $T = e(g, g)^{abc}$ holds or not.*

Proof. We use the following two games played between an adversary $\mathcal{A}$ and a challenger $\mathcal{C}$ to prove this theorem.

Game 0: This game models the original chosen plaintext attacks on the above scheme.

- Phase 1: In this phase, $\mathcal{A}$ can adaptively issue the following queries to the challenger $\mathcal{C}$.
  - Public key oracle $\mathcal{O}_{pk}$: When inputting an index i by $\mathcal{A}$, $\mathcal{C}$ runs $\mathtt{HE.KeyGen}(1^\lambda)$ to obtain (pk_i, sk_i), and returns pk_i to $\mathcal{A}$ but keeps sk_i secret. At last, $\mathcal{C}$ records (pk_i, sk_i) into List L_k.
  - Private key oracle $\mathcal{O}_{sk}$: When inputting a public key pk_i from List L_k by $\mathcal{A}$, $\mathcal{C}$ returns the corresponding sk_i to $\mathcal{A}$.
  - Re-encryption key generation oracle $\mathcal{O}_{rk}$: When inputting two public keys (pk_i, pk_j) from List L_k by $\mathcal{A}$, $\mathcal{C}$ returns the corresponding rk_{ij} to $\mathcal{A}$.
- Challenge Phase: At some point, $\mathcal{A}$ decides to finish Phase 1, then it sends $\mathcal{C}$ two same length messages m_0, m_1 from $\mathbb{G}_T$ and a public key pk^*. There exist three restrictions on pk^*. 1) pk^* has never been queried to $\mathcal{O}_{sk}$; 2) if (pk^*, pk) has been queried to $\mathcal{O}_{rk}$, pk cannot be queried to $\mathcal{O}_{sk}$; and 3) if pk has been queried to $\mathcal{O}_{sk}$, (pk^*, pk) cannot be queried to $\mathcal{O}_{rk}$. $\mathcal{C}$ returns $\mathtt{HE.Enc}(pk^*, m_{\mathbf{b}})$ to $\mathcal{A}$ as the challenge ciphertext c^*, where $\mathbf{b}$ is a random bit.
- Phase 2: It is almost the same as that in Phase 1, except the following restrictions.
  - Private key oracle $\mathcal{O}_{sk}$: 1) pk^* cannot be queried to this oracle; 2) if (pk^*, pk) has been queried to $\mathcal{O}_{rk}$, pk cannot be queried to $\mathcal{O}_{sk}$; and 3) if (pk^*, c^*, pk) has been queried to $\mathcal{O}_{re}$, pk cannot be queried to $\mathcal{O}_{sk}$.
  - Re-encryption key generation oracle $\mathcal{O}_{rk}$: If pk has been queried to $\mathcal{O}_{sk}$, (pk^*, pk) cannot be queried to $\mathcal{O}_{rk}$.
- Guess: $\mathcal{A}$ outputs a guess $\mathbf{b}'$ on $\mathbf{b}$. If $\mathbf{b}' = \mathbf{b}$, then $\mathcal{A}$ wins the game.

If the probability of $|\Pr[\mathbf{b}' = \mathbf{b}] - 1/2|$ is negligible, then the above scheme is chosen plaintext secure for the original ciphertext.

Game 1: In this game, we will modify Game 0 as follows. Given the input of eDBDH problem $(g, g^{b/a}, g^a, g^b, g^c) \in \mathbb{G}$ and $T \in \mathbb{G}_T$, $\mathcal{C}$ sets $g_1 = g$ and $g_2 = g^b$.

- Phase 1: In this phase, $\mathcal{A}$ can adaptively issue the following queries to the challenger $\mathcal{C}$.

– Public key oracle $\mathcal{O}_{pk}$: When inputting an index i by $\mathcal{A}$, $\mathcal{C}$ decides the value of $\theta_i \in \{0,1\}$ such that $\Pr[\theta_i = 1] = \delta$, and chooses a random x_i from $\mathbb{Z}_q$. If $\theta_i = 1$, $\mathcal{C}$ sets $pk_i = (pk_{i,1}, pk_{i,2}) = (e(g_1, g_2)^{x_i}, g_2^{1/x_i})$. If $\theta_i = 0$, $\mathcal{C}$ sets $pk_i = (pk_{i,1}, pk_{i,2}) = (e(g^a, g_2)^{x_i}, (g^{b/a})^{1/x_i})$. At last, $\mathcal{C}$ records (pk_i, x_i, θ_i) into List L_k.

– Private key oracle $\mathcal{O}_{sk}$: When inputting a public key pk_i from List L_k by $\mathcal{A}$, $\mathcal{C}$ searches (pk_i, x_i, θ_i) in List L_k. If $\theta_i = 1$, $\mathcal{C}$ returns x_i to $\mathcal{A}$; otherwise, it outputs `failure`.

– Re-encryption key generation oracle $\mathcal{O}_{rk}$: When inputting two public keys (pk_i, pk_j) from List L_k by $\mathcal{A}$, $\mathcal{C}$ searches (pk_i, x_i, θ_i) and (pk_j, x_j, θ_j) in List L_k.

  * If $\theta_i = \theta_j$, $\mathcal{C}$ returns $rk_{ij} = g_2^{x_i/x_j}$.
  * If $\theta_i = 1$ and $\theta_j = 0$, $\mathcal{C}$ returns $rk_{ij} = pk_{j,2}^{x_i/x_j}$.
  * If $\theta_i = 0$ and $\theta_j = 1$, $\mathcal{C}$ outputs `failure`.

- Challenge Phase: After receiving m_0, m_1, pk^* from $\mathcal{A}$, $\mathcal{C}$ searches (pk^*, x^*, θ^*) List L_k. If $\theta^* = 1$, $\mathcal{C}$ outputs `failure`; otherwise, it outputs the challenge ciphertext c^* as follows.

$$c_1^* = g^c,\ c_2^* = T^{x^*} \cdot m_{\mathbf{b}}.$$

  Note that if $T = e(g,g)^{abc}$, then we have $c_2^* = (e(g_1^a, g_2)^c)^{x^*} \cdot m = pk_1^{*c} \cdot m_{\mathbf{b}}$.

- Phase 2: It is almost the same as that in Phase 1 but with the restrictions as that in Game 0.

- Guess: $\mathcal{A}$ outputs a guess $\mathbf{b}'$ on $\mathbf{b}$. If $\mathbf{b}' = \mathbf{b}$, then $\mathcal{A}$ wins the game.

If $\mathcal{C}$ does not output `failure` in Game 1, the probability of $|\Pr[\mathbf{b}' = \mathbf{b}] - 1/2|$ is not larger than the probability of solving the eDBDH problem. This probability could be obtained as follows. Suppose $\mathcal{A}$ makes a total of q_{sk} private key queries and q_{rk} re-encryption generation queries. Then the probability that $\mathcal{C}$ does not output `failure` in phases 1 or 2 is $\delta^{q_{sk}} \cdot (1 - \delta \cdot (1 - \delta))^{q_{rk}} \geq \delta^{q_{sk}+2q_{rk}}$. The probability that $\mathcal{C}$ does not output `failure` during the challenge phase is $(1 - \delta)$. Therefore, the probability that $\mathcal{C}$ does not output `failure` in Game 1 is no less than $\delta^{q_{sk}+2q_{rk}}(1 - \delta)$. This value is maximized at $\theta_{opt} = 1 - 1/(q_{sk} + 2q_{rk} + 1)$. As a result, the probability that $\mathcal{C}$ does not output `failure` in Game 1 is no less than $1/e(q_{sk} + 2q_{rk} + 1)$. □

Theorem 2. *The above scheme is chosen plaintext secure under the eDDH assumption for the re-encrypted ciphertext. The eDDH assumption assumes that given $(g, g^{1/a}, g^a) \in \mathbb{G}$ and $(e(g,g)^b, T) \in \mathbb{G}_T$, it is hard for any polynomial probabilistic time adversary to decide whether $T = e(g,g)^{ab}$ holds or not.*

Proof. As that in the proof of Theorem 1, we use two games played between an adversary $\mathcal{A}$ and a challenger $\mathcal{C}$ to prove this theorem.

Game 0: This game models the original chosen plaintext attacks on the above scheme.

- Phase 1: Identical to that in the proof of Theorem 1.
- Challenge Phase: At some point, $\mathcal{A}$ decides to finish Phase 1, then it sends $\mathcal{C}$ two same length messages m_0, m_1 from $\mathbb{G}_T$ and two public keys pk, pk^*. There exists only one restriction on pk^*, i.e., pk^* has never been queried to $\mathcal{O}_{sk}$. $\mathcal{C}$ returns $\texttt{PRE.ReEnc}(\texttt{PRE.ReKey}(pk, sk, pk^*)$ and $\texttt{HE.Enc}(pk, m_{\mathbf{b}})$ to $\mathcal{A}$ as the challenge ciphertext c^*, where sk is the private key corresponding to pk, and $\mathbf{b}$ is a random bit.
- Phase 2: It is almost the same as that in Phase 1, except the following restrictions.
  - Private key oracle $\mathcal{O}_{sk}$: pk^* cannot be queried to this oracle.
- Guess: $\mathcal{A}$ outputs a guess $\mathbf{b}'$ on $\mathbf{b}$. If $\mathbf{b}' = \mathbf{b}$, then $\mathcal{A}$ wins the game.

If the probability of $|\Pr[\mathbf{b}' = \mathbf{b}] - 1/2|$ is negligible, then the above scheme is chosen plaintext secure for the re-encrypted ciphertext.

Game 1: In this game, we will modify Game 0 as follows. Given the input of eDDH problem $(g, g^{1/a}, g^a,) \in \mathbb{G}$ and $(e(g,g)^b, T) \in \mathbb{G}_T$, $\mathcal{C}$ sets $g_1 = g$ and $g_2 = g^w$, where w is a random element from $\mathbb{Z}_q$.

- Phase 1: In this phase, $\mathcal{A}$ can adaptively issue the following queries to the challenger $\mathcal{C}$.
  - Public key oracle $\mathcal{O}_{pk}$: When inputting an index i by $\mathcal{A}$, $\mathcal{C}$ decides the value of $\theta_i \in \{0,1\}$ such that $\Pr[\theta_i = 1] = \delta$, and chooses a random x_i from $\mathbb{Z}_q$. If $\theta_i = 1$, $\mathcal{C}$ sets $pk_i = (pk_{i,1}, pk_{i,2}) = (e(g_1, g_2)^{x_i}, g_2^{1/x_i})$. If $\theta_i = 0$, $\mathcal{C}$ sets $pk_i = (pk_{i,1}, pk_{i,2}) = (e(g^a, g_2)^{x_i}, (g^{w/a})^{1/x_i})$. At last, $\mathcal{C}$ records (pk_i, x_i, θ_i) into List L_k.
  - Private key oracle $\mathcal{O}_{sk}$: When inputting a public key pk_i from List L_k by $\mathcal{A}$, $\mathcal{C}$ searches (pk_i, x_i, θ_i) in List L_k. If $\theta_i = 1$, $\mathcal{C}$ returns x_i to $\mathcal{A}$; otherwise, it outputs `failure`.
  - Re-encryption key generation oracle $\mathcal{O}_{rk}$: When inputting two public keys (pk_i, pk_j) from List L_k by $\mathcal{A}$, $\mathcal{C}$ searches (pk_i, x_i, θ_i) and (pk_j, x_j, θ_j) in List L_k.
    - If $\theta_i = \theta_j$, $\mathcal{C}$ returns $rk_{ij} = g_2^{x_i/x_j}$.
    - If $\theta_i = 1$ and $\theta_j = 0$, $\mathcal{C}$ returns $rk_{ij} = pk_{j,2}^{x_i/x_j}$.

$*$ If $\theta_i = 0$ and $\theta_j = 1$, $\mathcal{C}$ returns $rk_{ij} = (g^a)^{w \cdot x_i / x_j}$.

- Challenge Phase: After receiving m_0, m_1, pk^* from $\mathcal{A}$, $\mathcal{C}$ searches (pk^*, x^*, θ^*) List L_k. If $\theta^* = 1$, $\mathcal{C}$ outputs `failure`; otherwise, it outputs the challenge ciphertext c^* as follows.

$$c_1^* = e(g,g)^b,\ c_2^* = T^{w \cdot x^*} \cdot m_{\mathbf{b}}.$$

  Note that if $T = e(g,g)^{ab}$, then we have that $c_2^* = (e(g_1^a, g_2)^b)^{x^*} \cdot m = pk_1^{*c} \cdot m_{\mathbf{b}}$.

- Phase 2: It is almost the same as that in Phase 1 but with the restrictions as that in Game 0.
- Guess: $\mathcal{A}$ outputs a guess $\mathbf{b}'$ on $\mathbf{b}$. If $\mathbf{b}' = \mathbf{b}$, then $\mathcal{A}$ wins the game.

If $\mathcal{C}$ does not output `failure` in Game 1, the probability of $|\Pr[\mathbf{b}' = \mathbf{b}] - 1/2|$ is not larger than the probability of solving the eDDH problem. This probability could be obtained as follows. Suppose $\mathcal{A}$ makes a total of q_{sk} private key queries. Then the probability that $\mathcal{C}$ does not output `failure` in phases 1 or 2 is $\delta^{q_{sk}}$. The probability that $\mathcal{C}$ does not output `failure` during the challenge phase is $(1-\delta)$. Therefore, the probability that $\mathcal{C}$ does not output `failure` in Game 1 is $\delta^{q_{sk}}(1-\delta)$. This value is maximized at $\theta_{opt} = 1 - 1/(q_{sk}+1)$. As a result, the probability that $\mathcal{C}$ does not output `failure` in Game 1 is no less than $1/e(q_{sk}+1)$. □

5.4 Analysis

In this section, we give the security analysis and performance analysis of the proposed frame.

5.4.1 Security Analysis

In the proposed framework, the reports are encrypted by the homomorphic encryption. Hence, only if the underlying homomorphic encryption scheme is secure, the consumer's privacy in the reports can be guaranteed. On the other hand, the statistics and analysis results are encrypted by the proxy re-encryption. Hence, only if the underlying proxy re-encryption scheme is secure, the confidentiality of the results can be guaranteed.

5.4.2 Performance Analysis

In our framework, the TA only needs to generate a certificate for each ER, and the communication cost is only related to this type of data. The cloud server needs to verify the certificate, and run

Table 5.1: The experimental results of the proposed scheme

Algorithm		**Time Cost**	
		in theory	in simulation (ms)
HE.KeyGen (PRE.KeyGen)		$1T_{e,\mathbb{G}} + 1T_{e,\mathbb{G}_T}$	53.69
HE.Enc (PRE.Enc)		$1T_{e,\mathbb{G}} + 1T_{e,\mathbb{G}_T} + 1T_{m,\mathbb{G}_T}$	53.75
PRE.ReKey		$1T_{e,\mathbb{G}}$	45.95
PRE.ReEnc		$1T_p + 2T_{e,\mathbb{G}_T} + 2T_{m,\mathbb{G}_T}$	66.99
HE.Eva	orig. ciphertext	$1T_{e,\mathbb{G}} + 1T_{e,\mathbb{G}_T} + (n-1)(T_{m,\mathbb{G}} + T_{m,\mathbb{G}_T})$	See Fig. 5.4
	re-encr. ciphertext	$2T_{e,\mathbb{G}_T} + 2(n-1)T_{m,\mathbb{G}_T}$	
PRE.Dec	orig. ciphertext	$1T_p + 1T_{e,\mathbb{G}_T} + 1T_{m,\mathbb{G}_T}$	59.18
	re-encr. ciphertext	$1T_{e,\mathbb{G}_T} + 1T_{m,\mathbb{G}_T}$	7.82

algorithms HE.Eva and PRE.ReEnc for every query from the ER, and the communication cost is related to these three types of data. As for the consumer side, the computation cost and communication cost is quite simple, and it is only related to HE.Enc. Regarding the ER side, the computation cost is mainly related to PRE.ReEnc and PRE.Dec.

As shown above, the performance of our proposed framework is mainly up to the underlying homomorphic encryption scheme and proxy re-encryption scheme. While the proposed framework itself is quite simple and easy to analyze.

Moreover, we implement the proposed scheme in Section 5.3.3 by using the Pairing-based Cryptography Library [80], where the parameter is the type A curve. The underlying processor is Pentium (R) Dula-Core CPU T4300 @ 2.10GHz 2.09GHz, and the operating system is Windows 7 Professional with Service Pack 1. The experimental results can be found in Table 5.1 and Fig.5.4. In Table 5.1, we denote $T_{e,\mathbb{G}}$, $T_{m,\mathbb{G}}$, $T_{e,\mathbb{G}_T}$, $T_{m,\mathbb{G}_T}$, and T_p as the timing of an exponentiation and a multiplication in $\mathbb{G}$, an exponentiation and a multiplication in $\mathbb{G}_T$, and a pairing, respectively. The time cost of HE.Eva is related to the number of ciphertexts n, and Fig. 5.4 shows such a relationship.

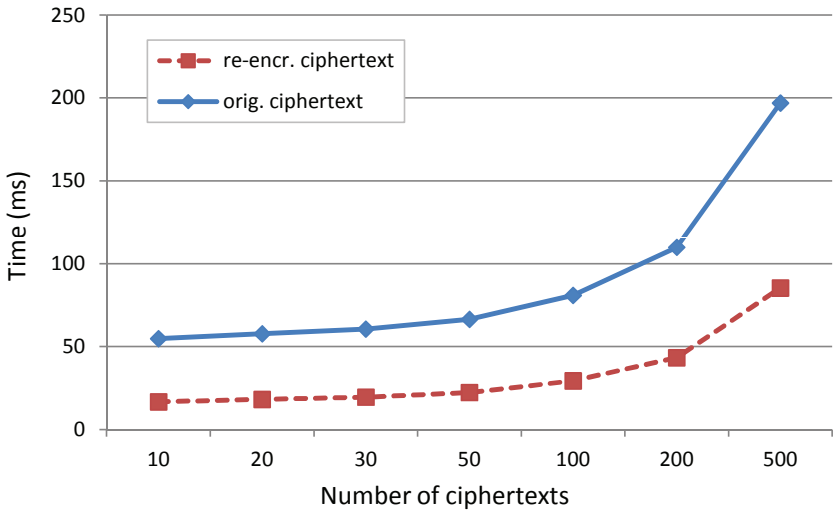

Figure 5.4: The experimental results of algorithm HE.Eva

5.5 Concluding Remarks

In this chapter, we have proposed a data sharing framework for smart grid. The proposed framework mainly studies how to keep smart grid still smart in the sense that electricity consumption reports can be analyzed by distributed energy resources, while the consumer privacy in the reports can still be protected. We also presented a concrete scheme (supporting multiplication homomorphism) falling into the proposed framework. Extensive analysis shows that the concrete scheme is secure and efficient. In the future work, we plan to design a concrete scheme supporting our requirements for the proposed framework and full homomorphism.

Chapter 6

Efficient and Privacy-preserving Smart Grid Downlink Communication Using Identity Based Signcryption

In this chapter, we propose an efficient and privacy-preserving scheme for smart grid downlink communication. Specifically, we propose an efficient identity based signcryption, called EIBSC, providing privacy preservation in downlink communication for smart grids. The proposed scheme is characterized by employing the concealing destination technique on the tree-based network to protect consumer privacy in downlink communication. Moreover, the proposed scheme employs identity based signcryption to efficiently achieve downlink message source authentication, data integrity and encryption. Additionally, compared to other identity-based signcryption schemes, the proposed scheme is more efficient in regards to computational overhead and ciphertext size. Furthermore, our security analysis illustrates that the proposed scheme is resilient against various security threats to smart grids.

6.1 Introduction

Smart grids are a combination of the electrical grid and power infrastructure together supplemented by information and communication technology (ICT). The smart grid has introduced several advantageous components over their traditional counterparts. One being the presence of smart meters, a significant part of modern day electrical infrastructure. They are installed at customer houses and are capable of computing and communicating with control centers (i.e.,

the utility companies). These smart meters can be wired or wirelessly connected to electrical appliances in consumer homes and are capable ccommunicating with other intuitive smart grid components such as gateways. This is beneficial not only to consumers but distributed energy resources/suppliers as well. Smart grids allow utility companies to monitor power generation, transmission, delivery and power consumption in real time. However, remotely control and use management rely on collecting data from smart grids, especially from smart meters in uplink and downlink transmissions.

In uplink transmissions, collected data is sent from smart meters to utility companies or to utility components like gateways. On the other hand, downlink transmissions control messages are sent from utility companies to smart meters, groups of smart meters or to compatible electrical devices on smart grids. These messages are intended to remotely control smart meters and enable a number of features which include device shut down, monitoring statuses and aggregation of power usage. Aggregated data and control messages are subject to several security and privacy concerns. Aggregated data may be manipulated and malicious entities can abuse broadcast messages to cause power outages for households and neighborhoods. Currently, a number of smart grids struggle against attackers for the control of key infrastructure. In 2015 Ukraine experienced an attack resulting in complete power outages for 103 cities and partial outages in 186 [81].

Security and privacy are of paramount importance in smart grid for both suppliers and consumers. Consumers privacy is a primary concern for customers, power consumption records may disclose household activities, occupancy and the variety of appliances in a house. For example, low power consumption is an indicator that home owners are possibly away. Additionally, power signatures can be analyzed to detect appliance types. Leaked utility information can be quite profitable for marketing. A number of consumers are engaged in industrial or manufacturing work. It is in the interest of both public and private parties to ensure privacy. This includes safeguarding against leakages and cyberattacks.

There has been extensive research concerning privacy and security on the smart grid. Much literature have proposed schemes to address confidentiality, data integrity and authenticity in uplink transmissions [8, 10, 22, 52, 53, 82, 83]. However, only few security schemes have been proposed to secure downlink communications in smart grid. Moreover, these schemes have been suggested to tackle security goals separately. More precisely, their focuses are only on either data integrity and authenticity or confidentiality. Of course, in smart grids a scheme that provides authenticity or/and data integrity does not necessarily provide confidentiality as well. For example, if the utility company wants to send a command to shut down, it can implement a digital signature scheme to prevent malicious users from issuing such a commands against consumers on the smart grid [84, 85].

On the other hand, consumers on the smart grid would like their privacy safeguarded not only in uplink transmission, but in the downlink communications as well. This would prevent unauthorized consumers from guessing to whom a command was issued to. In this case, an adversary or even a legitimate user may discern the destination in a residential area network to which a shutdown command was issued. As a result of this, consumers will lose sensitive information during the downlink communication. Therefore, there is an urgent need for authenticity-confidentiality schemes in the smart grid to efficiently achieve both authenticity and confidentiality.

Signcryption is a cryptographic primitive that simultaneously performs both encryption and digital signing on a message. Signcryption schemes are constructed to be more efficient than combining two separate schemes for encryption and signing. Additionally, Signcryption is intended to provide essential security services as well as reduce computational cost and communication overhead.

In this chapter, we propose an efficient privacy-preserving scheme for smart gird downlink communication using identity based signcryption. The proposed scheme, called EIBSC, provides consumer privacy and authenticity as well as data integrity in downlink communication while saving computational cost and communication overhead. The proposed scheme is characterized by employing concealing technique to provide consumer privacy. Concretely, each residential consumer has a smart meter connected to smart appliances forming House Area Networks. Smart meters in residential areas can be connected via a gateway, which acts as an intermediate point between smart meters, and a control center in utility companies. With our proposed scheme, control and command messages can be securely and efficiently transmitted from the control center to smart meter(s). Specifically, the contributions of this chapter are:

- We propose an identity-based signcryption scheme implemented concealing technique for downlink communication in smart gird in order to protect consumer privacy, authenticity and data integrity.
- The proposed scheme is based on a tree-based network in which downlink communication can be more efficient using minimum spanning trees and privacy preservation is provided using the concealing destination technique.
- The proposed scheme is much more efficient in terms of computational costs and ciphertext size compared to other signcryption schemes and outperforms existing competing schemes.

The remainder of this chapter is organized as follows. In Section 6.2, we introduce the system model, residential tree-based network construction, security requirements and design goal. In

Section 6.3, we recall the bilinear pairing. Then, we present our EIBSC scheme in Section 6.4, followed by its security analysis and performance evaluation in Section 6.5 and Section 6.6, respectively. Finally, we draw our conclusion in Section 6.7.

6.2 System Model And Design Goals

In this section, we formalize the system requirements, residential tree-based network construction, security requirements, and identify our design goals.

6.2.1 System Model

In our system model, we consider a typical Residential Tree-based Network (RTN) consists of smart meter devices in a smart grid system. Smart grid systems consist of a control center, a RTN gateway (R-Gateway), and a number of smart meters (SMs) $SM_1, SM_2, \cdots$, as shown in Fig. 6.1.

- **Control Center (CC)**: Control center is a trusted entity whose responsibilities include initializing a system, collecting uplink transmissions from smart meters, real-time monitoring and issuing downlink messages and commands to smart meters in smart grid systems.
- **RTN Gateway (R-Gateway)**:R-Gateway is a network entity which serves as a relay between the control center and smart meters for transmissions. RTN Gateways direct message to specific smart meter destinations.
- **Smart Meters Network** (SMs) $\{SM_1, SM_2, \cdots\}$: Smart meters are an important component that can electrically record the nearly real-time data concerning electrical consumption. Smart meters only report data to R-Gateway. They also receive requests and commands from the latter.

Communication Model. In the residential tree-based networks (RTN), we first construct a tree-based network to allow downlink transmission coverage for smart meters in a local area; further details are described network construction subsection below. Communication between each SM_i and the R-Gateway is through relatively cheap WiFi technology. However, when a smart meters network range is large, it is impossible for some SM_i to directly communicate with R-Gateway. In this case, multi-hop communication will be formed in the RTN. On the other hand, the communication between R-Gateway and the control center is dependent upon the

high-bandwidth, low delay, reliable and secure channels. These can guarantee smart grid two-way communication, facilitating for demand response, dynamic pricing, and system monitoring in smart grid systems.

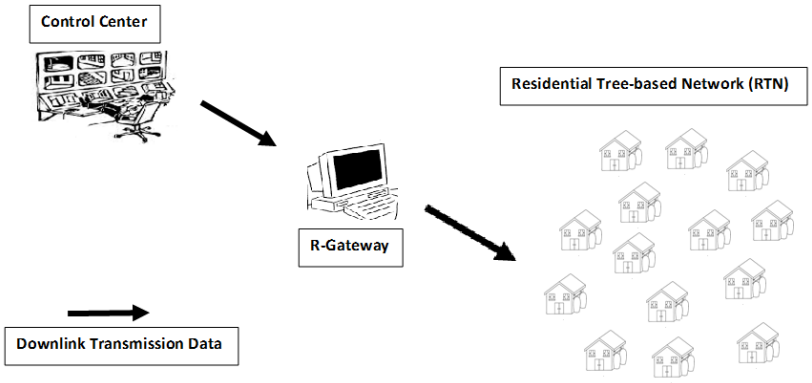

Figure 6.1: System model under consideration: a residential tree-based network (RTN) with a number of smart meters (SMs) and RTN gateway.

6.2.2 Construction of Tree-based Network

We consider the residential smart meter networks as graphs $G(V, E)$, where V denotes the set of smart meters (i.e., vertices) and E denotes the set of wireless links (i.e., edges) between two smart meters. In addition, residential tree-based networks are a minimum spanning tree, meaning they contain smart meters with the smallest cost communication paths (i.e., shortest available wireless links). More precisely, the graph should be connected and each smart meter should have the shortest communication path to the gateway. A routing table contains all pertinent details for network topology. Address information remain static throughout the network as shown in Fig. 6.2. The control center takes advantage of this feature and can send commands to any smart meter by leveraging the routing table. A copy of the routing table is deployed to all participating smart meters at the installation phase in a top-down manner.

6.2.3 Security Requirements

Security is crucial for the success of secure smart meters communications as well as the protection of user's privacy. In our security model, we consider the control center and the gateway are trustable, and the consumers represented by smart meters are honest. However, there exists

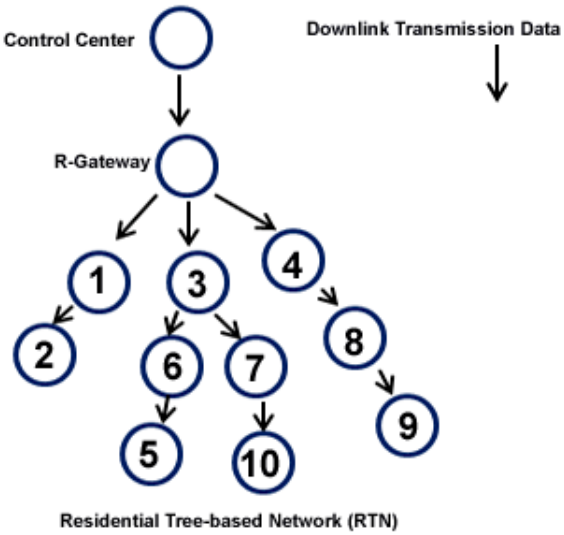

Figure 6.2: Routing Table:a residential tree-based network (RTN) with a number of smart meters (SMs) and RTN gateway.

an adversary $\mathcal{A}$ which resides in the RTN, eavesdropping the control center's messages. The adversary $\mathcal{A}$ could also launch some active attacks to threaten the data integrity or could remotely disconnect smart meters. The adversary $\mathcal{A}$ can launch a passive attack against the RTN and only monitors communications channel, threatening the privacy. Therefore, to prevent the adversary $\mathcal{A}$ from violating the consumer privacy and detect a adversary $\mathcal{A}$'s malicious actions in the downlink transmission, the following security requirements should be satisfied.

- *Downlink controlling messages should be secure*. Only the authorized commands from the control center can be accepted by smart meters. In other words, if a command is not from the control center, the requested command will not be executed by the smart meter.
- *Downlink data integrity should be provided*. Data manipulation by unauthorized parties (i.e., the adversary $\mathcal{A}$) or even legitimate consumers should be detected. Only the control center is in charge of issuing controlling messages to smart meters. Therefore, detection of bogus data is expected to meet smart grid's application requirements in uplink and downlink transmission.
- *User's Privacy should be protected*. Protection of consumer data is of paramount importance in uplink and downlink transmission. Although the adversary $\mathcal{A}$ can eavesdrop the

WiFi communication in the RTN, it cannot identify the content of downlink messages and cannot determine to whom these messages are destined to in smart meter network.

6.2.4 Design Goals

Under the aforementioned system model and security requirements, our intent is to develop an efficient identity-based signcryption scheme (EIBSC) for privacy-preserving downlink data communication to satisfy the aforementioned security requirements. Specifically, the following two desirable goals should be achieved:

- *Security and privacy preservation*: As stated above, the proposed EIBSC scheme should deliver real-time data security, data integrity, authentication as well as the residential consumer privacy to satisfy the above security requirements. If the scheme does not satisfy security requirements, then consumer privacy will be infringed. As a result, this will hinder the proliferation of smart grids.
- *Efficiency*: The proposed scheme should also be efficient in regards to computational cost and ciphertext size compared to existing schemes.

6.3 Bilinear Pairing

In this section, we recall the bilinear pairing technique [55], which serves as the basis of the proposed scheme. Let $\mathbb{G}_1$ and $\mathbb{G}_2$ be two cyclic groups of the same large prime order p, and $P_1 \in \mathbb{G}_1$ be the generator of $\mathbb{G}_1$. An *admissible* bilinear pairing $\hat{e} : \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$ is a map with the following properties: i) *Bilinearity:* For all $P, Q \in \mathbb{G}_1$ and any $a, b \in \mathbb{Z}_p^*$, we have $\hat{e}(aP, bQ) = \hat{e}(P, Q)^{ab}$; ii) *Non-degeneracy:* $\hat{e}(P_1, P_1) \neq 1_{\mathbb{G}_2}$; and iii) *Computability:* There is an efficient algorithm to compute $\hat{e}(P, Q)$ for all $P, Q \in \mathbb{G}_1$. Such an *admissible* bilinear pairing $\hat{e} : \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$ can be implemented by the modified Weil/Tate pairings over elliptic curves [55].

Definition 2 (Bilinear Parameter Generator)**.** *A bilinear parameter generator $\mathcal{G}en$ is a probabilistic algorithm that takes a security parameter k as input, and outputs a 5-tuple $(p, P_1, \mathbb{G}_1, \mathbb{G}_2, \hat{e})$ where p is a k-bit prime number, $\mathbb{G}_1, \mathbb{G}_2$ are two groups with order p, $P_1 \in \mathbb{G}_1$, is a generator, and $\hat{e} : \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$ is a non-degenerated and efficiently computable bilinear map.*

Next, we state the following three underlying problems, which serve as a basis of our proposed scheme.

Definition 3 (Computational Diffie-Hellman (CDH) Problem)**.** *The CDH problem is stated as follows: Given the elements* $(P_1, aP_1, bP_1) \in \mathbb{G}_1$ *for unknown* $a, b \in \mathbb{Z}_p^*$*, to compute* $abP_1 \in \mathbb{G}_1$.

Definition 4 (Bilinear Diffie-Hellman (BDH) Problem)**.** *The BDH problem is stated as follows: Given the elements* $(P_1, aP_1, bP_1, cP_1) \in \mathbb{G}_1$ *for unknown* $a, b, c \in \mathbb{Z}_p^*$*, to compute* $\hat{e}(P_1, P_1)^{abc} \in \mathbb{G}_1$.

6.4 Proposed EIBSC Scheme

In this section, we will present our identity based signcryption scheme (EIBSC). Before proceeding to the scheme's details, an overview of EIBSC is introduced.

6.4.1 Overview of EIBSC scheme

From time to time, utility companies represented by control centers need to send information to remote smart meters in residential areas. While the topology of the residential area networks is fixed and routing information is stored in the control center's database, the control center can send downlink data to smart meters using a hop-by-hop transport protocol; exploiting the fixed tree-based network. Fig. 6.3 outlines such a multi-hop topology, where the gateway (R-Gateway) serves as the root node of the smart meters network, and smart meters act as nodes in the residential area. In order to achieve reliable multi-hop communication in a tree-based network topology between the control center and smart meters, the control center first sends the downlink messages to the gateway based on the network routing table. This in turn forwards the messages to subsequent nodes. Moreover, in order to achieve efficient transmission, data integrity and privacy in the downlink communication, the proposed scheme employs the signcryption scheme and concealing technique. While the signcryption scheme performs digital signature and encryption simultaneously, the concealing technique conceals the destination of the downlink message. This makes it difficult for adversaries or even a legitimate consumer to discern to whom a message is destined.

6.4.2 Description of EIBSC scheme

In this section, we will present our efficient Identity Based Signcryption (EIBSC) scheme. The EIBSC scheme is composed mainly of four algorithms: system initialization algorithm, registration and private key extraction algorithm, signcryption algorithm, and unsigncryption algorithm.

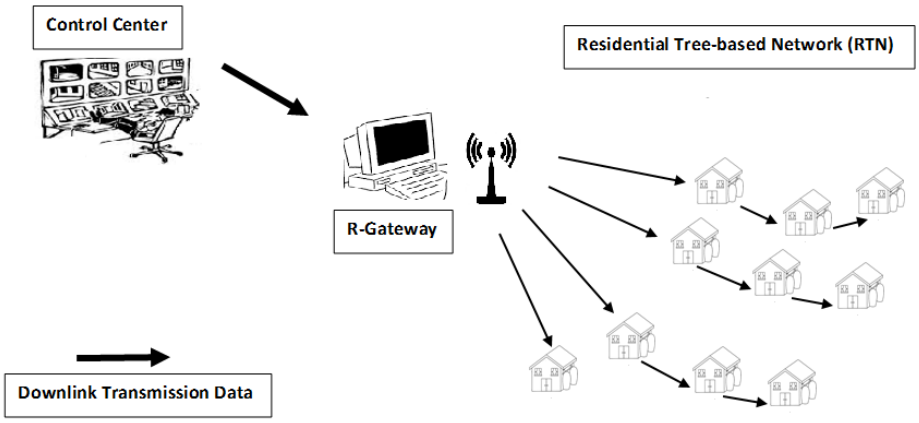

Figure 6.3: Fixed and reliable multi-hop downlink transmission for tree-based smart meters network.

1) System Initialization

In the system initialization algorithm, the control center is responsible for system parameters configuration. In particular, given the security parameter k, the control center first generates the bilinear parameters $(p, P_1, \mathbb{G}_1, \mathbb{G}_2, \hat{e})$ by running $\mathcal{G}en(k)$, and chooses three secure cryptographic hash functions H_1, H_2 and H_3, where $H_1 : \{0,1\}^* \rightarrow \mathbb{G}_1$; $H_2 : \mathbb{G}_2 \rightarrow \{0,1\}^n$ where n is the length of plaintext and $H_3 : \{0,1\}^n \rightarrow \mathbb{Z}_p^*$. Then, the control center picks a random number $s \in \mathbb{Z}_p^*$, and calculates $P_{pub} = sP_1$. With these settings, the control center keeps s as the master key secretly, and publishes the public parameters $Params = (p, P_1, \mathbb{G}_1, \mathbb{G}_2, n, \hat{e}, H_1, H_2, H_3, P_{pub})$.

2) Registration and Private Key Extraction

In this algorithm, the gateway gets initial authentication and submits its identity ID_c to the control center. The control center in turn uses the master secret key s to calculate the gateway's private key as $S_c = sH_1(ID_c)$ and sends S_c back to the gateway over a secure channel. Similarly, each smart meter $SM_1, SM_2, \cdots$ submits its identity (ID_s) to the control center for registration and obtains a private key as $S_s = sH_1(ID_s)$ from the control center via a secure channel. Moreover, the control center's identity and private key are ID_g and $S_g = sH_1(ID_g)$, respectively.

3) Signcryption(S_g, ID_s, m)

To signcrypt message $m \in \{0,1\}^n$, the signcryption performs encryption and signing simultaneously using the public parameters $Params$ and (S_g, ID_s, m) as follows:

- The control center generates a random integer $r \in \mathbb{Z}_p^*$, and
- Calculates $C_1 = rP_1$.
- Calculates $Q_s = H_1(ID_s)$.
- Calculates $K = H_2(\hat{e}(rQ_s, P_{pub}))$.
- Calculates the ciphertext $C_{enc} = m \oplus K$.
- Calculates $h = H_3(m)$.
- Calculates $C_{sign} = hS_g + rP_{pub}$, where S_g is the control center's private key.
- Output $\sigma = (C_1, C_{enc}, C_{sign})$ is the signcryption of the control center on message m.

After determining the destination, i.e., the smart meter (SM_t with its ID_t) that will receive the signcrypted message, the control center runs `Algorithm` 3. This algorithm takes two inputs: the destination and message m. It first constructs the path to the intended smart meter based on the routing table of the residential tree-based network. Then, it performs signcryption (σ) and chooses a random number $t \in \mathbb{Z}_p^*$ to compute $C_{hide} = (t, H_1(tS_t))$ where H_1 is a cryptographic hash function defined in the public parameters and S_t is the smart meter's private key. Finally, it returns three outputs: a set of smart meters T_{SMs} containing the intended destination, concealing value (C_{hide}), and signcrypted message σ where σ is a 3-tuple (C_1, C_{enc}, C_{sign}).

4) Unsigncryption(S_s, ID_g, σ)

unsigncryption algorithm is in charge of decrypting and verifying sent by the control center. The intended smart meter needs to perform unsigncryption algorithm as follows:

- Calculates $Q_g = H_1(ID_g)$.
- Calculates $K' = H_2(\hat{e}(S_s, C_1))$.
- Calculates $m = C_{enc} \oplus K'$.

Algorithm 3 Concealing Technique and Multi-Hop Forwarding

INPUT: SM_t, and m.
OUTPUT: $(T_{SMs}, C_{hide}, \sigma)$.

procedure CONCEALING TECHNIQUE AND MULTI-HOP FORWARDING
  Set T_{SMs}= Select smart meters on the path to SM_t.
  if SM_t is not leaf node **then**
    if SM_{t+1} exists **then**
      Add SM_{t+1} **to** T_{SMs} ▷ SM_{t+1} is the intermediate leaf node in the path.
    else
      Add SM_{t+k} and **ALL** nodes SM_{t-n} **to** T_{SMs} ▷ SM_{t-n} are direct ancestors of the non-intermediate leaf SM_{t+k}
    end if
  end if
  compute σ = **Signcryption**(S_g, ID_s, m) ▷ where ID_t is the identity of SM_t.
  choose randomly r where $r \leftarrow \mathbb{Z}_p^*$.
  compute C_{hide} = $(t, H_1(tS_t))$.
  return $(T_{SMs}, C_{hide}, \sigma)$
end procedure

- Calculates $h = H_3(m)$.

If the following equation 6.1 holds, then the decrypted message (m) is authenticated. Otherwise, it is invalid and the downlink data will be rejected.

$$\hat{e}(P_1, C_{sign}) \stackrel{?}{=} \hat{e}(P_{pub}, hQ_g + C_1) \tag{6.1}$$

The correctness of recovering the message is as follows:

- The sender calculates

$$\begin{aligned} K &= H_2(\hat{e}(rQ_s, P_{pub}) \\ &= H_2(\hat{e}(rQ_s, sP_1)) \\ &= H_2(\hat{e}(Q_s, P_1)^{rs}). \end{aligned}$$

- The recipient calculates

$$\begin{aligned} K' &= H_2(\hat{e}(S_s, C_1) \\ &= H_2(\hat{e}(sQ_s, rP_1)) \\ &= H_2(\hat{e}(Q_s, P_1)^{rs}) = K. \end{aligned}$$

The correctness of equation 6.1 is as follows:

$$\begin{aligned}\hat{e}(P_1, C_{sign}) &= \hat{e}(P_1, hS_g + rP_{pub}) \\ &= \hat{e}(P_1, hsQ_g + rsP_1) \\ &= \hat{e}(P_1, s(hQ_g + rP_1)) \\ &= \hat{e}(P_1, (hQ_g + rP_1)^s) \\ &= \hat{e}(sP_1, hQ_g + rP_1) \\ &= \hat{e}(P_{pub}, hQ_g + C_1).\end{aligned}$$

In order to send a downlink message to a specific smart meter, the control center runs `Algorithm` 3 and forwards the outputs of the algorithm, namely $(T_{SMs}, C_{hide}, \sigma)$ to the gateway. The gateway, in turn further forwards the outputs to a subsequent smart meter in the tree-based network (as in T_{SMs}). Upon receiving outputs, each smart meter in T_{SMs}, including the intended smart meter will perform two tasks. The first one is to continue forwarding the message $(T_{SMs}, C_{hide}, \sigma)$ downlink according to the deployed routing table in the smart meter and smart meters in T_{SMs}. It is worth noting that the intended smart meter will continue to forward the received message to other smart meters. In the example shown in Fig 6.4, suppose smart meter 8 is the intended smart meter that will receive a message from its immediate parent, that is, smart meter 4 and then will further forward down the received message to smart meter 9. As a result, the real destination (or intended recipient) of a downlink smart grid communication is concealed. Such a design not only efficiently assists in forwarding downlink communication, but also prevents unauthorized entities from learning the destination of downlink messages, providing privacy preservation.

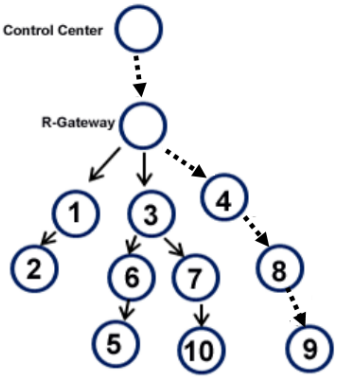

Figure 6.4: An illustration of concealing technique.

The second task is to verify whether downlink data is intended to itself or not. This is done by computing $C_{verify} = H_1(tS_t)$ where S_t is the smart meter's private key and t is extracted from Chide. C_{verify} is compared to $H_1(tS_t)$ extracted from C_{hide}. if they are equal then the smart meter recovers and verifies σ by running $Unsigncryption(S_s, ID_g, \sigma)$. Otherwise, C_{verify} does not match C_{hide} there is no need to perform $Unsigncryption$ on σ.

6.5 Security Analysis

In this section, we analyze the security properties of the proposed EIBSC scheme. In particular, following the security requirements discussed earlier, our analysis will focus on how the proposed EIBSC scheme can achieve downlink message source authentication and data integrity as well as consumer data privacy preservation.

- The proposed EIBSC scheme can provide the downlink messages source authentication. Since the encrypted message C_{enc}, digital signature C_{sign} and the value C_1 in σ are based on Definition 2 and 3 that are provably secure in the random oracle model [55] can be intercepted and forged by adversary $\mathcal{A}$, adversary $\mathcal{A}$ cannot encrypt any message without knowing r which is randomly chosen from and used in $C_1 = rP_1$ and $C_{enc} = m \oplus H_2(\hat{e}(rQ_s, P_{pub}))$, and cannot sign any message without having a control center's private key used in $C_{sign} = hS_g + rP_{pub}$. In other words, the intercepted and modified messages $(\bar{C}_1, \bar{C}_{enc}, \bar{C}_{sign})$ are not the same (C_1, C_{enc}, C_{sign}) encrypted and signed by the control center. In addition, since the proposed scheme employs a random integer in encryption, it can resist the possible replay attack Thus, σ is resilient against forgery. The proposed EIBSC scheme provides the downlink message source authentication.

- The proposed EIBSC scheme provides data integrity. Since the downlink messages are signed (e.g., $C_{sign} = hS_g + rP_{pub}$) by the control center's signature, adversary $\mathcal{A}$ cannot sign it without the control center's signature which is provably secure under the BDH problem in the random oracle model [9]. Similarly, when a downlink message $C_{enc} = m \oplus K$ where K is $H_2(\hat{e}(rQ_s, P_{pub})$ is encrypted by the control center, it can resist deliberate forgery attack. As a result, the adversary $\mathcal{A}$'s malicious behaviors in the downlink transmission can be detected in the proposed EIBSC scheme.

- The proposed EIBSC scheme can provide consumer data privacy preservation. The concealing technique algorithm provides a mechanism preventing adversary $\mathcal{A}$ or legitimate consumers on the smart grid from recognizing the destinations of downlink messages.

Moreover, only the intended consumer will signcrypt the downlink message from the control center and the other participating consumers know nothing about the message and the destination in communication which guarantees privacy preservation. Although the adversary $\mathcal{A}$ can manipulate the downlink data, it cannot learn anything about communication and such malicious behavior can be easily detected by the proposed scheme.

6.6 Performance Evaluation

The efficiency of the proposed scheme can be evaluated with respect to computational cost and ciphertext length. We first calculate the computation time for $Signcryption$ and $Unsigncryption$ using an MNT curve of embedding degree $k = 6$ and 160-bit p on an Intel Pentium IV 3.0 GHZ machine [86]. Since the point multiplication in $\mathbb{G}_1$ and pairing computations dominate each message's computational overhead, only these operations are counted in the calculation. Table 6.1 shows the measured time (in milliseconds) for the aforementioned operations. In this table, T_{PMUL} denotes the time of point multiplication (that is, mul) in $\mathbb{G}_1$ and $T_{\hat{e}}$ denotes the time pairing (that is, $\hat{e}$).

We have 4 point multiplication operations and 1 pairing operation in signcryption and 1 point multiplication operation and 3 pairing operations in unsigncryption. Accordingly, the execution time of signcryption would be as follows:

$$T_{sign} = 4 \times T_{PMUL} + 1 \times T_{\hat{e}} = 4 \times 0.6 + 1 \times 4.5 = 6.9\ ms.$$

While the execution time of unsigncryption is calculated as:

$$T_{unsign} = 1 \times T_{PMUL} + 4 \times T_{\hat{e}} = 1 \times 0.6 + 3 \times 4.5 = 13.5\ ms.$$

Table 6.1: Execution time of cryptographic operations

Operation	**Execution Time**
T_{PMUL}	0.6 ms
$T_{\hat{e}}$	4.5 ms

In order to show the efficiency of the proposed scheme, we compare the proposed scheme with two ID-based signcryption schemes Libert-Quisquater [49] and Lal *et al.* [50] implemented according to their original descriptions. Table 6.2 gives the detailed computations and total

execution time in these schemes compared to our proposed scheme. We can see from these results that the proposed scheme outperforms the competing schemes in terms of computational costs. Table 6.2 also shows the ciphertext-size of the comparative schemes. We can see the length of a signcrypted text of the proposed scheme is similar to the length in Libert-Quisquater.

Table 6.2: Computational overhead and ciphertext-size

Schemes	**Signcryption**			**Unsigncyption**			**Ciphertext-size**
	mul	$\hat{e}$	$time(ms)$	mul	$\hat{e}$	$time(ms$	
Libert-Quisquater [49]	2	2	10.2	1	4	18.6	$\|m\| + \|\mathbb{Z}_p^*\| + \|\mathbb{G}\|$
Lal *et al.* [50]	6	1	8.1	1	3	13.5	$\|m\| + \|\mathbb{G}\|$
EIBSC	4	1	6.9	1	3	13.5	$\|m\| + 2\|\mathbb{G}\|$

6.7 Concluding Remarks

In this chapter, we have proposed an efficient identity-based signcryption scheme for privacy-preserving data in downlink communication, called EIBSC, for smart grid. Compared to other identity-based signcryption schemes, the proposed scheme can efficiently achieve privacy-preserving data in downlink communication from the control center to smart meter networks in residential areas. The proposed scheme employs concealing techniques to provide privacy preservation. Additionally, the proposed scheme is much more efficient in regards to computational overhead and ciphertext size. Our security analysis has shown that the proposed scheme can resist various security threats in smart grids.

Chapter 7

Conclusions and Future Work

In this chapter, we summarize our contributions in this thesis and propose our future research work.

7.1 Contributions

The major contributions of this thesis can be summarized as follows:

- First, we propose a novel data sharing framework for the smart grid, where two popular infrastructure are combined: the smart grid and cloud computing together. The proposed framework allows the electricity consumption reports generated in smart grid to be stored in the cloud, and the distributed energy resources can obtain the statistics and analysis results from the cloud computing. Therefore, the proposed framework can take advantage of cloud computing for the smart grid. Additionally, the proposed framework makes use of the homomorphic encryption technique to facilitate the statistics and analysis on the encrypted electricity consumption reports, and the proxy re-encryption technique to keep the statistics and analysis results secret from the cloud.

- Second, by considering residential user privacy and efficiency issues in data aggregation in a Residential Area Network (RAN) of smart meter devices, we propose an efficient lightweight privacy-preserving data aggregation scheme (ELPDA)to address the security and privacy challenges. In ELPDA, based on one-time masking technique, each smart meter's data can be efficiently encrypted and aggregated. Compared with popular Paillier

Cryptosystem based aggregation (PCBA) algorithm applied in smart grid, the proposed ELPDA is much more efficient, reducing the aggregation delay in the whole RAN. Additionally, We carry out extensive simulations to examine the average aggregation delay in ELPDA. The simulation results will demonstrate the efficiency of the proposed ELPDA scheme. ELPDA outperforms the PCBA algorithms in terms of average aggregation delay in smart grid.

- Third, inspired by the facts that the gateways may be corrupted, we present a security-enhanced data aggregation scheme from trapdoor hash functions, Pailliar encryption and homomorphic authenticators. To the best of our knowledge our proposed scheme is the first one against malicious gateways and a successful attempt to construct authentication schemes from trapdoor hash functions with key exposure. We analyze the security strength and performance of the security-enhanced data aggregation scheme. In particular, we employ provable security technique to reduce the security of our scheme to the well-known mathematical hard problems and underlying cryptographic tools. Through the performance comparison, we demonstrated that our scheme is indeed significantly more efficient than the existing schemes in terms of both communication and computational overheads.

- Fourth, for downlink communication in smart gird we propose an identity-based signcryption scheme implemented concealing technique in order to protect consumer privacy, authenticity and data integrity. The proposed scheme (EIBSC) implements a tree-based network in which downlink communication can be much more efficient using minimum spanning trees and privacy preservation is provided using the concealing destination technique and identity-based signcryption. Additionally, the proposed scheme (EIBSC) is much more efficient in terms of computational costs and ciphertext size compared to other signcryption schemes and outperforms existing competing schemes.

7.2 Future Work

In the future work, we plan not only to study the methods to design data aggregation schemes that are able to detect and trace the misbehaviors of legitimate consumers in smart grid, but also to design a concrete scheme supporting our requirements for the proposed framework and full homomorphism.

Since supervisory control and data acquisition(SCADA) is a crucial system in modern industry such water and waste water, energy and smart grid, we also plan to further investigate SCADA and malware in order to satisfy significant security features for smart grid.

While there is no general solution or scheme for a wide variety of many privacy issues in smart grid and how we should properly address them using different techniques, we plan to investigate differential privacy as a promising topic. Unlike other types of privacy models, differential privacy might be applied to statistical databases and micro-data datasets for smart grid in order to preserve consumer data privacy. Differential privacy protects the privacy of aggregated datasets, but might not protect privacy in data aggregation. Therefore, differential privacy can be applied in a specific setting in order to learn as much as possible about a group in a database while learning as little as possible about a single individual(i.e., a householders or consumer).

References

[1] NYISO. (2016). *NYISO interim report on the august 14, 2003 blackout* [Online]. Available: http://www.hks.harvard.edu/hepg/Papers/NYISO.blackout.report.8.Jan.04.pdf

[2] M. Jacobs. (2016). *13 of the Largest Power Outages in History and What They Tell Us About the 2003 Northeast Blackout* [Online]. Available: http://blog.ucsusa.org/2003-northeast-blackout-and-13-of-the-largest-power-outages-in-history-199

[3] M. Davis. (2016). *Smartgrid device security adventures in a new medium* [Online]. Available: http://www.blackhat.com/presentations/bh-usa-09/MDAVIS/BHUSA09-Davis-AMI-SLIDES.pdf

[4] McAfee. (2016). *Global energy cyberattacks: Night dragon* [Online]. Available: http://www.mcafee.com/us/resources/white-papers/wp-global-energy-cyberattacks-night-dragon.pdf

[5] D. Bobkoff. (2016). *10 Years After The Blackout, How Has The Power Grid Changed?* [Online]. Available: http://www.npr.org/2013/08/14/210620446/10-years-after-the-blackout-how-has-the-power-grid-changed

[6] K. Alharbi, X. Lin, and J. Shao, "A privacy-preserving data sharing framework for smart grid," *IEEE Internet of Things Journal*, to appear.

[7] K. Alharbi and X. Lin, "LPDA: a lightweight privacy-preserving data aggregation scheme for smart grid," in *Proc. of IEEE Wireless Communications & Signal Processing (WCSP)*, pp. 1–6, 2012.

[8] F. Li, B. Luo, and P. Liu,"Secure information aggregation for smart grids using homomorphic encryption,"in *Proc. of IEEE SmartGridComm'10*, pp. 327–332, 2010.

[9] N. Saputro and K. Akkaya,"Performance evaluation of smart grid data aggregation via homomorphic encryption,"in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 2945-2950, 2012.

[10] R. Lu, X. Liang, X. Li, X. Lin, and X. Shen, "Eppa: An efficient and privacy-preserving aggregation scheme for secure smart grid communications," *IEEE Trans on Parallel and Distributed Systems*, vol. 23, no. 8, 2012.

[11] J. Ni, K. Alharbi, X. Lin, and X. Shen, "Security-Enhanced Data Aggregation against Malicious Gateways in Smart Grid," in *Proc. of IEEE GLOBECOM 2015*, 2015.

[12] K. Alharbi and X. Lin, "Efficient and Privacy-preserving Smart Grid Downlink Communication Using Identity Based Signcryption," in *Proc. of IEEE GLOBECOM 2016*.

[13] National Institute of Standards and Technology. (2016). *Nist framework and roadmap for smart grid interoperability standards, release 2.0.* [Online]. Available: http://www.nist.gov/smartgrid/upload/NIST Framework Release 2-0 corr.pdf

[14] SmartGrid.gov. (2016). *What is the Smart Grid?* [Online]. Available: https://www.smartgrid.gov/the_smart_grid/smart_grid.html

[15] K. Alfaheid,"Secure and compromise-resilient architecture for advanced metering infrastructure," MASc Thesis, University of Ontario Institute of Technology, 2011.

[16] D. Chen, D. Irwin, P. Shenoy, and J. Albrecht, "Combined heat and privacy: preventing occupancy detection from smart meters," in *Proc. of IEEE International Conference on Pervasive Computing and Communications*, 2014.

[17] P. McDaniel and S. McLaughlin, "Security and privacy challenges in the smart grid," *IEEE Security Privacy*, vol. 7, no. 3, pp. 75-77, 2009.

[18] Y. Chen, A. Oudalov, and A. Timbus, "The provision of frequency control reserves from multiple micro-grids," *IEEE Transactions on Industrial Electronics*, vol. 587, no. 1, pp. 173-183, 2011.

[19] M. Wen, R. Lu, J. Lei, H. Li, X. Liang, and X. Shen, "SESA: an efficient searchable encryption scheme for auction in emerging smart grid marketing," *Security and Communication Networks*, vol. 7, pp. 234-244, 2013.

[20] A. Metke and R. Ekl, "Smart grid security technology," in *Proc. of Innovative Smart Grid Technologies (ISGT)*, pp. 1-7, 2010.

[21] E. Pallotti and F. Mangiatordi, "Smart grid cyber security requirements," in *Proc. of EEEIC*, pp. 1-4, 2011.

[22] Z. M. Fadlullah, N. Kato, R. Lu, X. Shen, and Y. Nozaki, "Toward secure targeted broadcast in smart grid," *IEEE Communications Magazine*, vol. 50, no. 5, pp. 150-156, 2012.

[23] M. Fouda, Z. M. Fadlullah, N. K. abd R. Lu, and X. Shen, "A lightweight message authentication scheme for smart grid communications," *IEEE Trans on Smart Grid*, vol. 2, no. 4, pp. 675-685, 2011.

[24] J. Clemente, "The security vulnerabilities of smart grid," *Journal of Energy Security*, 2009.

[25] W. Stallings, "Classical Encryption Techniques,"in *Cryptography and network security,* principles and practices,4th ed. Pearson Practice Hall, 2006, pp. 32.

[26] A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone, "Overview of Cryptography,"in *Handbook of Applied Cryptography,*CRC Press, Boca Raton, Florida, USA, 1997, pp. 30–32.

[27] M. Blaze, G. Bleumer, and M. Strauss, "Divertible protocols and atomic proxy cryptography," in *Proc. of Advances in Cryptology(EUROCRYPT)*, pp. 1127–144, 1998.

[28] M. Luther, "What Is IBE,"in *Introduction to identity-based encryption*. Artech House, 2008, pp. 1–7.

[29] S. Al-Riyami and K. Paterson, 'Certificateless public key cryptography," in *Proc. of International Conference on the Theory and Application of Cryptology and Information Security*, pp. 452–473, 2003.

[30] A. Sahai and B. Waters, "Fuzzy identity-based encryption," in *Proc. of Advances in Cryptology (EUROCRYPT)*, pp. 457–473, 2005.

[31] P. Biswas. (2016). *Attribute Based Encryption* [Online]. Available: http://www.slideshare.net/prosunjit/attribute-based-encryption

[32] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data," in *Proc. of ACM conference on Computer and communications security (CCS '06)*, pp. 89–98, 2006.

[33] IEEE Standard. (2016). *1402-2000 - IEEE Guide for Electric Power Substation Physical and Electronic Security* [Online]. Available: https://standards.ieee.org/findstds/standard/1402-2000.html

[34] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Proc. of Advances in cryptology-EUROCRYPT*, pp. 223–238, 1999.

[35] Z. Erkin, J. R. Troncoso-Pastoriza, R. Lagendijk, and F. Prez-Gonzlez,"An Overview of Privacy-Preserving Data Aggregation in Smart Metering Systems," *IEEE Signal Processing Magazine*, 2013.

[36] F. Garcia, and B. Jacobs, "Privacy-friendly energy-metering via homomorphic encryption," *Security and Trust Management*, pp. 226–238, 2010.

[37] X. Li, X. Liang, R. Lu, X. Shen, X. Lin, and H. Zhu, "Securing smart grid: cyber attacks, countermeasures, and challenges," *IEEE Communications Magazine,* vol. 58, no. 8, pp. 38–45, 2012.

[38] C.-I. Fan, S.-Y. Huang, and Y.-L. Lai, "Privacy-enhanced data aggregation scheme against internal attackers in smart grid," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 1, pp. 666-675, 2014.

[39] K. Ohara, Y. Sakai, F. Yoshida, M. Iwamoto, and K. Ohta, "Privacy-preserving smart metering with verifiability for both billing and energy management," in *Proc of AsiaPKC'14*, Kyoto, Japan, June 3, 2014.

[40] T. Dimitriou, "Secure and scalable aggregation in the smart grid," in *Proc of IEEE NTMS'14*, Dubai, 2014.

[41] H.-Y. Lin, W.-G. Tzeng, S.-T. Shen, and B.-S. P. Lin, "A practical smart metering system supporting privacy preserving billing and load monitoring," in *Proc. of Applied cryptography and network security (ACNS'12)*, vol. 7341, pp. 544-560, 2012.

[42] K. Kursawe, G. Danezis, and M. Kohlweiss, "Privacy-friendly aggregation for the smart-grid," in *Proc. of Privacy Enhancing Technologies*, pp. 175–191, 2011.

[43] Y. Feng, Q. Yi, and H. R. Qingyang, "HIBaSS: hierarchical identity-based signature scheme for AMI downlink transmission," *Security and Communication Networks*, vol. 8, no. 16, pp. 2901–2908, 2015.

[44] S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving secure, scalable, and fine-grained data access control in cloud computing," in *Proc. of IEEE International Conference on Computer Communications (INFOCOM)*, pp. 1–9, 2010.

[45] C. Wang, K. Ren, S. Yu, and K. M. R. Urs, "Achieving usable and privacy-assured similarity search over outsourced cloud data," in *Proc. of IEEE International Conference on Computer Communications (INFOCOM)*, pp. 451–459, 2012.

[46] J. Shao, R. Lu, and X. Lin, "Fine: A fine-grained privacy-preserving location-based service framework for mobile devices," in *Proc. of IEEE International Conference on Computer Communications (INFOCOM)*, pp. 244–252, 2014.

[47] Y. Zheng, " Digital signcryption or how to achieve cost (signature & encryption) $\ll$ cost (signature) + cost (encryption)," in *Proc. of Annual International Cryptology Conference*, pp. 165–179, 1997.

[48] H.K.-H. So, S.H.M. Kwok, E.Y. Lam, and King-Shan Lui,"Zero-configuration identity-based signcryption scheme for smart grid," in *Proc. of IEEE Internation Conference on Smart Grid Communications,* Gaithersburg, MD, pp. 321–326, 2010.

[49] B. Libert and J.J. Quisquater, "New identity based signcryption schemes from pairings," in *Proc. of IEEE Information Theory Workshop,* Paris, France, pp. 155–158, 2003.

[50] S. Lal and P. Kushwah, "ID based generalized signcryption," *Cryptology ePrint Archive*, http://eprint.iacr.org/2008/84, 2008.

[51] K. Zhang, R. Lu, X. Liang, J. Qiao, and X. Shen, "PARK: A privacy-preserving aggregation scheme with adaptive key management for smart grid," in *Proc. of IEEE Int. Conf. Commun. China (ICCC)*, pp. 236–241, 2013.

[52] F. Li and B. Luo, "Preserving data integrity for smart grid data aggregation," in *Proc. of IEEE Third International Conference on Smart Grid Communications (SmartGridComm)*, pp. 366–371, 2012.

[53] D. Li, Z. Aung, J. R. Williams, and A. Sanchez, "Efficient authentication scheme for data aggregation in smart grid with fault tolerance and fault diagnosis," in *Proc. of IEEE Innovative Smart Grid Technologies*, pp. 1–8, 2012.

[54] Y. Lei and L. Fengjun, "Detecting False Data Injection in Smart Grid In-Network Aggregation," in *Proc. of IEEE Smart-GridComm*, pp. 40813, 2013.

[55] D. Boneh and M. Franklin, "Identity-based encryption from the Weil pairing," *SIAM J. Comput.*, vol. 32, no. 3, pp. 586–615, 2003.

[56] A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone, "Mathematical Background,"in *Handbook of Applied Cryptography,*CRC Press, Boca Raton, Florida, USA, 1997, pp. 72.

[57] H.-Y. Lin and W.-G. Tzeng. (2016). *An Efficient Solution to the Millionaires' Problem Based on Homomorphic Encryption* [Online]. Available: https://eprint.iacr.org/2005/043.pdf.

[58] MIRACL Library. (2016). *Multiprecision Integer and Rational Arithmetic c/c++ Library* [Online]. Available: https://certivox.org/display/EXT/MIRACL.

[59] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid - The new and improved power grid: A survey," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 4, pp. 944–980, 2012.

[60] V. C. Güngör, D. Sahin, T. Kocak, S. Ergüt, C. Buccella, C. Cecati, and G. P. Hancke, "Smart grid technologies: communication technologies and standards," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 4, pp. 529–539, 2011.

[61] J. C. Cha and J. H. Cheon, "An identity-based signature from gap diffie-hellman groups," in *Proc. of Public Key Cryptography*, pp. 18–30, 2003.

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

[63] W. Wang, and Z. Lu, "Cyber security in the smart grid: survey and challenges," *Computer Networks,* vol. 57, no. 5, pp. 1344–1371, 2013.

[64] F. Diao, F. Zhang, and X. Cheng, "A privacy-preserving smart metering scheme using linkable anonymous credential," *IEEE Transactions on Smart Grid,* to appear.

[65] K. Kursawe, and C. Peters, "Structural weaknesses in the open smart grid protocol," C*ryptology ePrint Archive*, https://eprint.iacr.org/2015/088.pdf, 2015.

[66] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin, "Private memoirs of a smart meter," in *Proc of BuildSys'10*, pp. 61–66, 2010.

[67] C. Rottondi, G. Verticale, and A. Capone, "Privacy-preserving smart metering with multiple data consumers," *Computer Networks*, vol. 57, no. 7, pp. 1699–1713, 2013.

[68] H. Li, X. Lin, H. Yang, X. Liang, R. Lu, and X. Shen, "EPPDR: an efficient privacy-preserving demand pesponse scheme with adaptive key evolution in smart grid," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 8, pp. 2053–2064, 2014.

[69] The National Academies. (2016). *Terrorism and the Electric Power Delivery System. The National Academies Press* [Online]. Available: http://www.nap.edu/catalog.php?record id=12050

[70] National Electric Sector Cybersecurity Organization Resource. (2016). *Analysis of Selected Electric Sector High Risk Failure Scenarios*. The National Academies Press [Online]. Available: http://www.nap.edu/catalog.php?record id=12050

[71] S. Hohenberger, V. Koppula, and B. Waters, "Universal signature aggregators," *Cryptology ePrint Archive,* https://eprint.iacr.org/2014/745.pdf, 2014.

[72] A. Shamir, and Y. Tauman, "Improved Online/Offine Signature Schemes," in *Proc of CRYPTO 2001,* LNCS 2139, pp. 355–367, 2001.

[73] H. Shacham, B. Waters, "Compact proofs of retrievability," in *Proc of Asiacrypt'08,* Sydney, Australia, Jan. 8-11, pp. 90-107. 2008.

[74] A. L. Ferrara, M. Green, S. Hohenberger, and M. Pedersen, "Practical short signature batch verification," in *Proc. of CT-RSA'09*, vol. 5473 of LNCS. Springer-Verlag, pp. 309–324, 2009.

[75] R. Deng, J. Chen, X. Cao, Y. Zhang, S. Maharjan, and S. Gjessing, "Sensing-performance tradeoff in cognitive radio enabled smart grid," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 302–310, 2013.

[76] H. Liang, B. Choi, W. Zhuang, and X. Shen, "Towards optimal energy store-carry-and-deliver for phevs via v2g system," in *Proc. of IEEE International Conference on Computer Communications (INFOCOM)*, pp. 1674–1682, 2012.

[77] K. Alharbi, X. Lin, and J. Shao, "A framework for privacy-preserving data sharing in the smart grid," in *Proc. of IEEE ICCC 2014*, 2014, pp. 214–219.

[78] C. Gentry, "A fully homomorphic encryption scheme," Ph.D. dissertation, STANFORD UNIVERSITY, 2009.

[79] J. Shao and Z. Cao, "Cca-secure proxy re-encryption without pairings," in *Proc. of International Workshop on Public Key Cryptography*, pp. 357–376, 2009.

[80] B. Lynn. (2016). *The pairing-based cryptography library* [Online]. Available: https://crypto.stanford.edu/pbc/

[81] J. Pagliery. (2016). *Scary questions in Ukraine energy grid hack* [Online]. Available: http://money.cnn.com/2016/01/18/technology/ukraine-hack-russia/.

[82] R. Berthier, W. Sanders, and H. Khurana, "Intrusion detection for advanced metering infrastructures: Requirements and architectural directions," in *Proc. of IEEE Smart Grid Communications (SmartGridComm)*, pp. 350–355, 2012.

[83] R. Lu, K. Alharbi, X. Lin, and C. Huang, "A Novel Privacy-Preserving Set Aggregation Scheme for Smart Grid Communications," in *Proc. of IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, 2015.

[84] R. Benzie. (2016). *Thousands of smart meters in Ontario to be removed over safety worries* [Online]. Available: http://www.thestar.com/news/queenspark/2015/01/22/thousands-of-smart-meters-in-ontario-to-be-removed-over-safety-worries.html.

[85] P. Paganini. (2016). *Smart meters in Spain can be hacked to hit the National power network* [Online]. Available: http://securityaffairs.co/wordpress/29353/security/smart-meters-hacking.html.

[86] M. Scott. (2016). *Efficient implementation of crytographic pairings* [Online]. Available: http://ecrypt-ss07.rhul.ac.uk/ Slides/Thursday/mscottsamos07.pdf.