ATLANTIS PRESS

Advances in Computer Science Research, volume 87

3rd International Conference on Mechatronics Engineering and Information Technology (ICMEIT 2019)

Grouping-Based Aggregation Protocol with Error Tolerance for Privacy-Preservation in Smart Grid

Guanlin Si, Yue Sun, Wei Chen, Jian Li

State Grid Jibei Electric Power Co. Ltd. Research Institute, Beijing 100045, China

Abstract. Smart grid, as the next generation of power grid characterized by "two-way" communication, has been paid great attention to realizing reliable and flexible electricity delivery for our future lives. In order to support the two-way communication in smart grid, a large number of smart meters (SMs) should be deployed for customers to report their real-time data to control center. However, this kind of real-time report would disclose users' privacy, bring down the users' willingness to participate in smart grid. In order to address this problem, in this paper, we propose a grouping-based aggregation protocol with error tolerance for privacy-preservation in smart grid (GBAP-ETPP), which employs the counting bloom filter to save storage space and improve the query efficiency. We also solve the false positive of bloom filter by choosing appropriate IDs in the system setup. In addition, we group users and all users in the same group share the same key. Thus, we can normally run our algorithm by substitution even there are some malfunctioning SMs. Detailed security analysis and theoretical proof show that our scheme can guarantee the security and privacy requirements of all the users in the smart grid.

Keywords: privacy-preserving, homomorphic encryption, error tolerance, authentication, bloom filter.

1. Introduction

As a new generation of energy network, smart grid is considered as a useful way to resolve the severe environmental and resource issues. it is the product of the combination of energy network and information technology. Differed from the unidirectional centralized grid, the control mode of the smart grid is more flexible and reliable. The user in the smart grid is not only a consumer but also a producer. The smart grid can supply users with electricity; on the contrary, the users can also provide the smart grid with their superfluous electricity which comes from their household energy. What's more, to realize the optimal scheduling, the smart grid installs a SM at each home to collect the real-time electricity consumption data, draw the real-time load curve, and make the plan for electricity generation. Not only that, the smart grid also adopts many new service modes. For example, the control center can make the dynamic price of the electricity to encourage users to adjust their power consuming behavior, collect the electricity consuming plan and make the electricity generation plan in advance.

Smart grid ensures the security, reliability and economy of the electric power system, while, it also threatens the privacy of the users. When the SM collects the real-time electricity data from the user to the control center, user's activity may be disclosed by observing his real-time data. what's more, the adversary may acquire the user's habit, for instance, when you get up, open the computer and when you leave home and come back and so on.

Privacy-preserving and validity-authentication are two important concerns in the security field of the smart grid. There are many solutions to protect user's privacy and authenticate use's identity. For example, we can use blind signature to protect user's privacy, but it needs to sign the data before sending to the control center, which isn't satisfied with real-time property of the smart grid. In addition, some scholars proposed a scheme based on data-obfuscation, which adds a random number to the real electricity consumption data to hide the user's activity, while this solution may cause large errors if the random numbers are not reasonable. There also exists a novel scheme based on the virtual ring to solve the problem. The main idea is that each member is distributed with the same private key for encryption so that the control center can't know the sender's identity even it can obtain the data. The flaw of this scheme is lack of effect authentication because of the uncertainty of identity. Recently, the homomorphic encryption is more and more popular used in the security of smart grid such as the paillier encryption. This scheme is a good way to realize the privacy-preserving, but there still exist

some problems in this kind of encryption algorithm. The obstacle is a lack of error-tolerance. In our paper, we proposed a scheme called Grouping-Based Aggregation Protocol with Error Tolerance for Privacy-Preserving in Smart Grid (GBAP-ETPP) based on the paillier encryption and realize the error –tolerance effectively.

The rest of the paper is organized as follows: Related work is reviewed in section2. The system model and design goals are described in section3. Basic cryptographic concepts are summarized in section4. Our schemes are presented in section5. Security analysis and theoretical proof are presented in sections6 and 7. At last, we conclude in section 8.

2. Related Work

Recently, various techniques have been proposed to address the problems of data security and data privacy in smart grid. A scheme using the battery to hide the real-time data is proposed in [1] and [14]. In these schemes, smart grid and the household battery provide users with electricity at the same time. When the household consuming curve goes high, the battery discharges. Otherwise, it charges. In this way, we can hide the user's real-time data to protect user's privacy. The downside is that the effect depends on the battery capacity, besides, charging and discharging the battery frequently is detrimental to the battery life. Some scholars propose a scheme based on blind signature to solve the privacy-preserving and validity-authentication in [2]. The main idea of this scheme is as follows: the user times his data with a random number which calls blind factor and sends the blind data to the third party. later, the third party authenticates the user's identity, signs the blind data with its private key and returns the signed data to the user. Thus, the user can obtain the right signature by multiplying the signed data with his inverse of the blind factor, while the third party doesn't know the content of the user's data. The downside of this scheme is that users should send their electricity data to the third party for authentication before communicating with the control center, which is against the real-time property of the power grid. An effective scheme based on virtual ring is presented in [3]. It groups the users by their geographical position and distributes each member in the same group with the same key. In this way, control center can obtain all of the users' data without knowing the sender's ID. Obviously, it's a good way to protect user's privacy, but the validity-authentication can't be guaranteed because of the anonymity.

In [4], Marmol Implements a novel scheme using homomorphic encryption to protect the real-time data from being revealed to the adversary. Control center can only obtain the total electricity of all the users in the smart grid, but this scheme is short of effect measures for error-tolerance. If there is a SM malfunctioning, the whole system will not run in the right way. Based on [4], Bao, Haiyong presents a scheme to realize the differential privacy in [5] and Zhiguo introduces some measures for error-tolerance in [6], while both of them are not ideal in the computational complexity.in [7], the author shows a scheme based on data obfuscation, which adds a random number to each electricity data to protect the real data from being disclosed by the adversary and control center. But it will cause some large errors if the random numbers are not reasonable. [8]- [13] are related to data aggregation algorithm, but there exist some flaws in complexity or accuracy.

3. System Model and Design Goals

3.1. System Modal

As shown in figure 1, smart grid is divided into four parts, which is comprised of control center (CC), Key initialization center (KIC), Cipher text aggregation device (CTD), and residential users.

1) Residential users

We divide all the users into different groups in accordance with their geographical location and the number of users in each group is the same. Each residential user is installed with a smart meter (SM) to collect the real-time electricity of all the home applications and send these encrypted values to the CTD.

2) Cipher text aggregation device

Cipher text aggregation device is a special device that is responsible for verifying the validity of the encrypted values which come from the SMs, taking measures for possible mistakes such as the absence of SM's values, and operating these values in order to acquire the summary of the real time electricity.

3) Key initialization center

Key initialization center is responsible for initializing all the keys for SMs and CC. its products public and private keys by pallier encryption, send the private key to CC for decryption. In addition, its products an ID for each legal user where the ID of each member in the same group are the same for the first five digits in order to judge which group the ID belongs to. At last, KIC inputs these IDs into k hash functions, set the values in the positions of the array according to the calculation result, send this bit array to CTD for authentication.

4) Control center

Control center can acquire the summary of real –time electricity in the smart grid using the private key distributed by the KIC. With these data, CC can learn of the trend of the power consumption and make the power generation plan immediately.

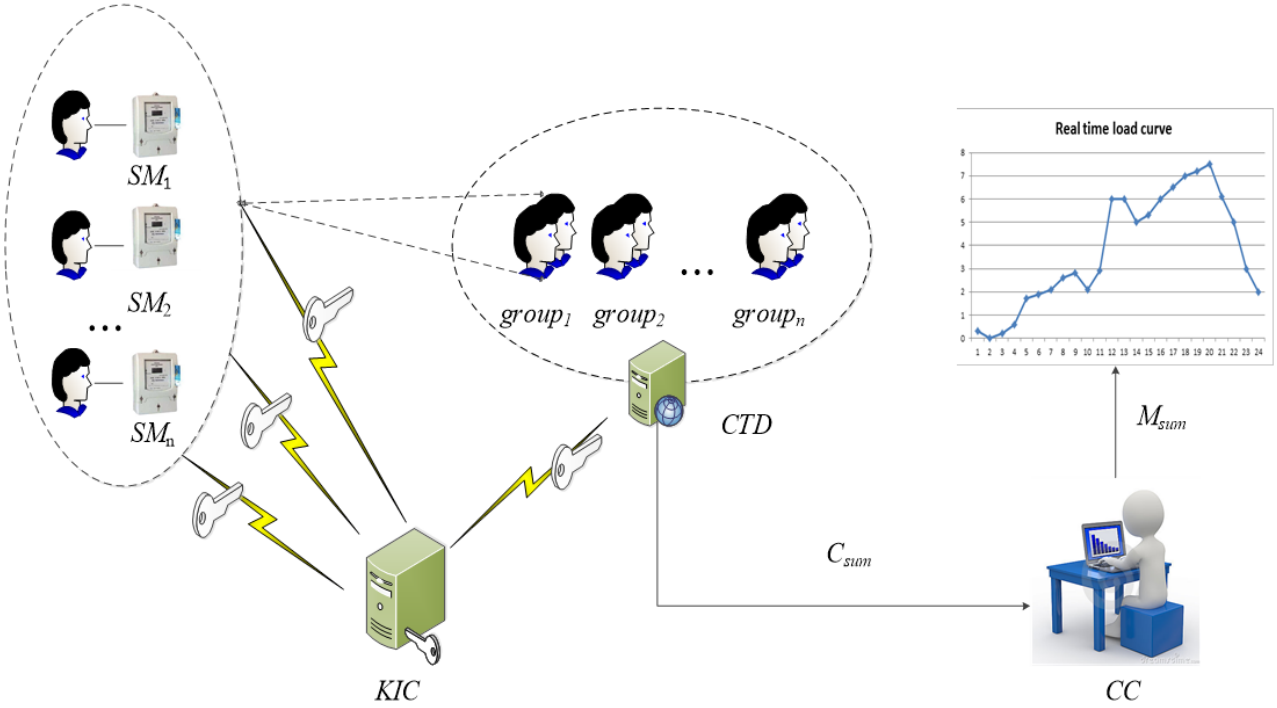

Figure.1 System model

3.2. Design Goals

Considering the above scenarios, our design goals are as follows:

1)Privacy preservation: A residential user's data is inaccessible to any other users or the CC, KIC, what's more, the outside adversary can't acquire the real electricity of users even he knows the cipher text and the public key.

2) Validity authentication: CTD can authenticate the validity of the user. if a user doesn't register at the KIC, it will be regarded as an illegal user. Besides, the CC will authenticate the validity of the message come from CTD to prevent adversary from sending forged data.

3) Error-detection and fault-tolerance:

The malfunctioning SMs in the smart grid can be detected by the KIC without privacy disclosure. Besides, the KIC can still aggregate the data of functioning meters even when there are malfunctioning ones.

4. Preliminaries

4.1. Notations

Table.1. Notations in GBAP-ETPP

Acronym	Descriptions
SM	Smart meter
CC	Control center
KIC	Key initialization center
CTD	Cipher text aggregation device
S_u	The private key of the key initialization center
P_u	The public key of the key initialization center
T	Time stamp
M_i	Plaintext
C_i	Ciphertext
H_i	The hash function
SK_i	The private key of smart meter

4.2. The Counting Bloom Filter

A bloom filter provides an efficient representation of a set $A = a_1, a_2, \ldots a_n$ of n elements to support membership queries. Firstly, we choose k hash functions and an array which has m bits. Secondly, we input each element into those hash functions, and set the corroding positions of the array into 1, while, the other positions are set in 0;

Given an element b, we can input it into the hash functions compiling the values with the former array to judge whether it belongs to the set. If all the values in the positions of the array are not 1, this indicates that the element b doesn't belong to set A. otherwise, the element b may belong to set A.

In this paper, we use the counting bloom filter to realize the authentication. The only difference between these two filters is that the value in the array of the counting bloom filter is the total times of all the hash mappings.

4.3. Homomorphic Encryption

Homomorphic encryption is a special encryption which can operate the cipher text to achieve ideal effect without knowing the plaintext. For example, there are three entities in our system: sender, intermediate, and receiver. All of the senders encrypt themselves values and send them to the intermediate. The receiver wants to achieve the summary of the values from all the senders. In order to achieve this purpose, the intermediate can multiply these cipher texts and send them to the receiver, thus, receiver can achieve the summary of the values coming from the senders while the intermediate provides the middle service without knowing the real values. We call this algorithm additive homomorphic encryption, and paillier encryption is a good example.

4.4. Log Normal Distribution

Log normal distribution is a distribution in mathematics field and many phenomena follow the Log normal distribution. Research shows that the real-time electricity follows the log normal distribution. The probability density function of the distribution is $f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)}{2\sigma^2}}$; The expectation can be demonstrated by $E(x) = e^{\frac{\mu+\sigma^2}{2}}$ and the variance can be demonstrated by $D(x) = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$.

5. Our Solution

This section presents our grouping-based aggregation protocol with error tolerance for privacy-preserving in smart grid in details. We call it GBAP-ETPP, which contains Setup, Encrypt, Aggregation, and Decrypt. The flow diagram of CTD is showed in figure 2:

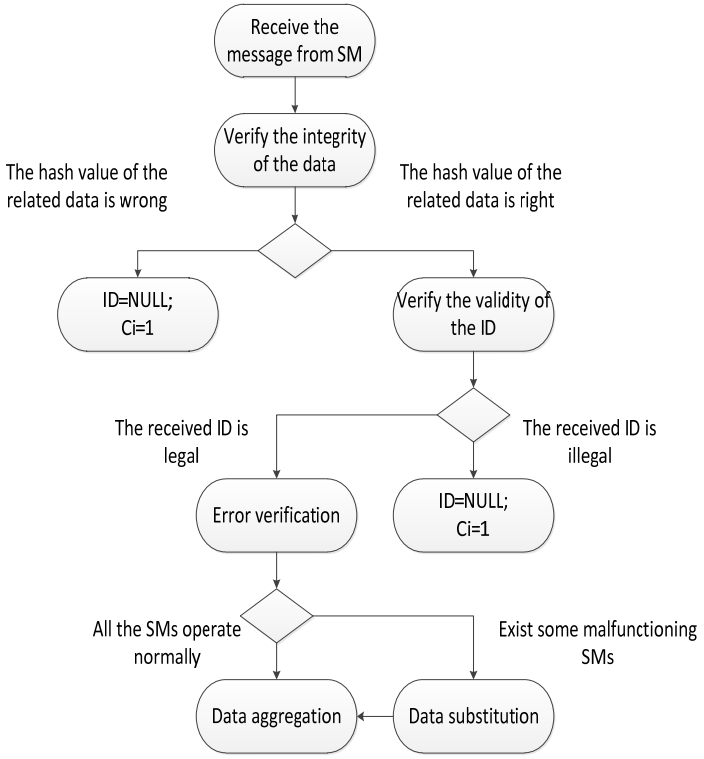

Figure. 2 The flow diagram of CTD

5.1 System Setup

The single KIC can bootstrap the whole system in the beginning. Concretely, KIC executes the following procedures for the system initialization.

- Given a security parameter k, KIC randomly selects two large primes p, q, where $|p| = |q| = k/2$. We calculate $n = pq$, choose a generator g in $Z^*_{n^2}$ and $y = lcm(p-1, q-1)$, $u = (L(g^y \bmod n^2))^{-1} \bmod n$, where $L(u) = (u-1) \div u$. KIC keeps (y, u) secretly and publishes (g, n)
- Divides all of the SMs into different groups which have the same member. If there is a group that doesn't have enough members, we can add some virtual SMs into this group. Then, KIC products a security parameter denoted by sk_i for each SM, where these parameters satisfy $sk_1 + sk_2 + sk_3 + \ldots sk_n = 0$. Next, KIC sends (g, n, h, sk_i) to each SM and sends (y, u) to the Power grid control center, where h is a random parameter.
- Creates an array which has m bits initialized to 0 and selects k hash functions. Then, it products ID for each user. inputs these IDs into the hash functions and set corroding value in the counting array according to the number of mappings; It is remarkable that the first five digits of user's ID are the same for users in the same group and each ID should not map to the same position with other IDs over k times. By running RSA algorithm, we can earn the public and private key for the CTD denoted by (p_u, s_u). At last, KIC sends k hash functions, the counting array and (p_u, s_u) to the CTD.

5.2 Data Encryption

SM installed at user's home is responsible for collecting the electricity and sending the data to CC in encrypted form. The specific steps are as follows:

- Computes $C_i = g^{m_i} r^n \bmod n\, h^{sk_i}$, where mi denotes the real electricity and r is a random number selected by the SM.
- Computes the hash function of the C_i together with its ID and current time stamp T as:
- $H_i = h(C_i \| ID \| T)$. At last, SM sends $[C_i, ID, T \| H_i]$ to the CTD.

5.3 Data Aggregation

CTD is located at the community to receive the users' message, inspect the validity of users by bloom filter, and aggregate the message before sending to the CC. In addition, it is also in charge of checking whether there exists absence of some SMs and taking corresponding measures. The specific steps are as follows:

- For each $[C_i, ID, T \| H_i]$ received, re-compute the message authentication code Hi based on the received C_i, ID, T to see whether it is the same as the one attached.
- Inputs the ID into the k hash functions and observes the values in the corresponding positions of the bit array. If the values in the corresponding positions are not all-natural numbers, or the ID maps to the same location with other IDs up to k times, this means the ID is illegal and we can drop the power message. Otherwise, it must be the legal one.
- Counts the number of the IDs in the same group by comparing their first five digits. If there exist some SMs whose values are absent, then CTD can use the other member's values in the same group to substitute the absent.
- Computes $C_{sum} = C_1 \times C_2 \times C_3 \times \ldots C_n$ and use s_u to acquire the signature as $sig_{s_u}(p_u)$. Then, sends $[C_{sum}, p_u, sig_{s_u}(p_u)]$ to the CC.

5.4 Data Decryption

After receiving the message of the CTD, the control center goes following steps:

- Uses the public key of the CTD to decrypt the $sig_{s_u}(p_u)$ to see whether it is equal to p_u in order to judge the validity of the message.
- Uses (y, u) to acquire the M_{sum} (denotes the summary of the electricity consumption) by running paillier algorithm.

6. Security Analysis

In this section, the security of the proposed protocol will be discussed.

1) Privacy preservation:

By encrypting the user's data with the key (g, n, h, sk_i) distributed by the KIC, we can ensure the privacy preservation of user's data. Because each SM has a special key sk_i and keep it privately, so the other users, the outside adversary and the KIC can't acquire the real electricity of the user even they know the cipher text. Besides, the CC can't acquire the real electricity of a special user because it only knows the aggregation of users' data.

2) Validity authentication:

CTD can authenticate the validity of the data which come from all of the SMs through bloom filter. If an adversary forges an ID and send a wrong data to CTD, we can distinguish it by inputting the ID into k hash functions and compiling the results with the pre-set value in the corroding positions of the counting array.

3) Error-detection and fault-tolerance:

Because all of the SMs are grouped by the electric type and each group has the same number of members, besides, we can distinguish which group a SM's data belongs to by the first five digits of the ID. Therefore, if there is a malfunctioning SM in some group, we can detect this error by compiling the total number of the received IDs with the normal number and using the data of other SM which has the same first five digits of the absentee's ID to substitute for normal aggregation. Thus, we can realize the error-detection and fault-tolerance.

7. Performance Evaluation

The most important idea of our paper is to use substitution to realize the error-tolerance, so it is necessary to prove the substitution between two numbers in the same group to be right. Because the real-time electricity consumption data follows the log normal distribution, and we can calculate its expectation and standard deviation. Based on these two values, we can easily get the expression of the error in our scheme. Input the number of malfunctioning SMs and the total number of the SMs in the smart grid, and we can get the error rates 'curve as follows:

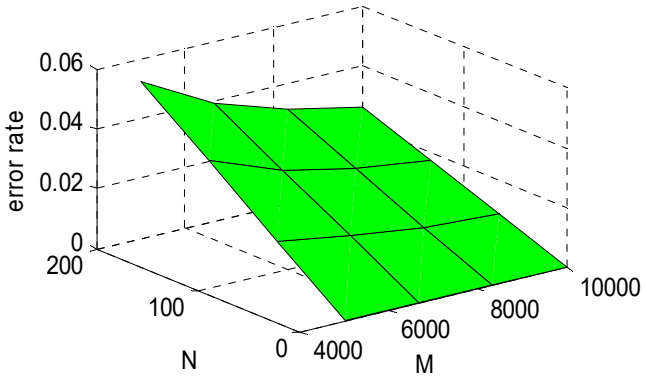

Figure. 3 The error rate in GBAP-ETPP

Figure.3. N denotes the number of the malfunctioning SMs. M denotes the total number of the SMs in the smart grid.

Through this picture, we can see that the rate of error in the worst case is0.05, while the rate of error is close to 0.02 in most cases, which is an acceptable value in the smart grid. Therefore, we can prove our solution is reasonable in the accuracy.

Next, we will try to test our solution in the time complexity. We compare our solution with a solution without adopting bloom filter in the time complexity in figure.3. Through analyzing this figure, we can easily find that our solution based on the bloom filter is more competitive than the other one.

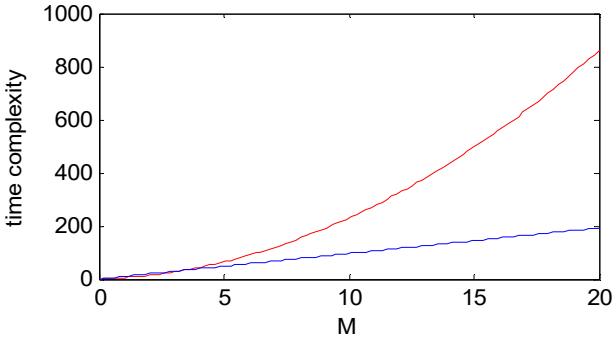

Figure. 4 The time complexity

The red curve denotes a solution without adopting bloom filter and the blue curve denotes our solution based on bloom filter M denote the number of SMs in smart grid.

8. Conclusion

This paper proposes a Grouping-Based Aggregation Protocol with Error Tolerance for Privacy-Preserving, which is based on the homomorphic encryption. Our solution can ensure the privacy-preservation, validity-authentication, and error-tolerance. Besides, we analyze our solution and give corroding proof of our substitution theory. The analysis shows that our solution satisfies the security requirements and has a good feasibility.

References

[1]. McLaughlin, Stephen, Patrick McDaniel, and William Aiello. "Protecting consumer privacy from electric load monitoring." Proceedings of the 18th ACM conference on Computer and communications security. ACM, 2011.

[2]. Cheung, Jeanno CL, et al. "Credential-based privacy-preserving power request scheme for smart grid network." Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE. IEEE, 2011.

[3]. Badra, Mohamad, and Sherali Zeadally. "Design and performance analysis of a virtual ring architecture for smart grid privacy." Information Forensics and Security, IEEE Transactions on 9.2 (2014): 321-329.

[4]. Marmol, Felix Gomez, et al. "Do not snoop my habits: preserving privacy in the smart grid." Communications Magazine, IEEE 50.5 (2012): 166-172.

[5]. Bao, Haiyong, and Rongxing Lu. "A new differentially private data aggregation with fault tolerance for smart grid communications." Internet of Things Journal, IEEE 2.3 (2015): 248-258.

[6]. Shi, Zhiguo, et al. "Diverse Grouping-Based Aggregation Protocol with Error Detection for Smart Grid Communications." Smart Grid, IEEE Transactions on 6.6 (2015): 2856-2868.

[7]. Beussink, Andrew, et al. "Preserving consumer privacy on ieee 802.11 s-based smart grid ami networks using data obfuscation." Computer Communications Workshops (INFOCOM WKSHPS), 2014 IEEE Conference on. IEEE, 2014.

[8]. Chim, Tat Wing, et al. "PRGA: Privacy-preserving recording & gateway-assisted authentication of power usage information for smart grid." Dependable and Secure Computing, IEEE Transactions on 12.1 (2015): 85-97.

[9]. Chen, Jie, Junping Shi, and Yueyu Zhang. "EPPDC: an efficient privacy-preserving scheme for data collection in smart grid." International Journal of Distributed Sensor Networks 2015 (2015): 60.

[10]. Dong, Xiaolei, Jun Zhou, and Zhenfu Cao. "Efficient privacy‐preserving temporal and spacial data aggregation for smart grid communications." Concurrency and Computation: Practice and Experience (2015).

[11]. Bao, Haiyong, and Le Chen. "A lightweight privacy‐preserving scheme with data integrity for smart grid communications." Concurrency and Computation: Practice and Experience (2015).

[12]. Jo, Hyo Jin, In Seok Kim, and Dong Hoon Lee. "Efficient and Privacy-Preserving Metering Protocols for Smart Grid Systems."

[13]. Efthymiou, Costas, and Georgios Kalogridis. "Smart grid privacy via anonymization of smart metering data." Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on. IEEE, 2010.

[14]. Liang, Xiaohui, et al. "UDP: Usage-based dynamic pricing with privacy preservation for smart grid." Smart Grid, IEEE Transactions on 4.1 (2013): 141-150.