PPMA: Privacy-Preserving Multisubset Data Aggregation in Smart Grid

Shaohua Li, Kaiping Xue, *Senior Member, IEEE*, Qingyou Yang, and Peilin Hong

***Abstract*—Privacy-preserving data aggregation has been extensively studied in smart grid. However, almost all existing schemes aggregate the total electricity consumption data of the whole user set, which sometimes cannot meet the fine-grained demands from control center in smart grid. In this paper, we propose a privacy-preserving multisubset data aggregation scheme, named PPMA, in smart grid. PPMA can aggregate users' electricity consumption data of different ranges, while guaranteeing the privacy of individual users. Detailed security analysis shows that PPMA can protect individual user's electricity consumption privacy against a strong adversary. In addition, extensive experiments results demonstrate that PPMA has less computation overhead and no more extra communication and storage costs.**

***Index Terms*—Multisubset aggregation, privacy-preserving, security, smart grid.**

I. Introduction

SMART GRID integrates the traditional power grid with information and communication technologies, to achieve efficient and reliable electricity generation, transmission, distribution, and control [1], [2]. In smart grid, both electricity and information are exchanged between utilities and users. This two-way mechanism allows smart grid to collect and analyze the situation of power generation, transmission, and consumption, etc., in real time; thus, ensuring a reasonable allocation of power, but also to ensure timely response to potential safety and security threats to the grid.

Collecting users' electricity consumption data is one of the most important processes of smart grid. Based on the collected data, control center (CC) in smart grid can use the means of data mining to obtain useful information, such as dynamic demand response [3], [4] and energy use improve [5]. To collect nearly real-time electricity consumption data, smart meters (SMs) are deployed on the consumers side, they usually record and report these data periodically, e.g., every 15 min. All these data will be transmitted to the CC for further processing and analysis. Generally, the CC has strong storage and computing capacity, and it can dynamically adjust the power grid, such as electricity supply, pricing, etc., to meet real-time demands.

Communications in smart grid must be secured against malicious attacks [6]–[8]. From consumers' perspective, privacy-preserving is quite necessary, as data reported by SMs will expose consumers detailed real-time using, which can reveal owner's personal information. For example, the low/high power consumption may indicate that householders are out/in the house [9]; periods of high activity correlate with more people being inside the home, e.g., for a get-together or party [10]. To preserve the privacy of consumers, one straightforward approach is to encrypt the data content before report, but this will increase the computing and communication cost.

To protect users' privacy against malicious attackers, in recent years, some data aggregation schemes [11]–[14] have been proposed to report overall consumption data in a geographical region, and protect users' fine-gained data from disclosure at the same time. Lu *et al.* [15] proposed a privacy-preserving data aggregation scheme, named EPPA, to aggregate multidimensional data with a superincreasing sequence. To enable the gateway (GW) to perform data aggregation, these schemes use homomorphic encryption techniques to encrypt consumption data. Homomorphic encryption can allow some computations in the plaintext space to be carried out on the corresponding ciphertext space, e.g., the multiplication on ciphertext corresponds to the addition on plaintext. All these schemes and many other related works choose Paillier cryptosystem [16] as their homomorphic algorithm. Meanwhile some others utilize different techniques, such as lattice-based [17] or ElGamal-based [18].

As mentioned above, the existing data aggregation schemes consider that CC is only concerned with the total electricity consumption of the whole user set, which sometimes cannot meet the fine-grained demands from the CC. For example, in order to make better power generation prediction [19], [20] and develop real-time pricing schemes [21], [22], CC needs to know not only the aggregation result of the whole set of users, but also the number of users whose electricity consumption is in a specific range, and the total consumption of these users. Lu *et al.* [23] first proposed a practical scheme to realize privacy-preserving set aggregation, and this scheme achieves the aggregation of the two subsets separately, but this scheme cannot handle the sit-

Manuscript received March 22, 2017; revised June 2, 2017; accepted June 26, 2017. Date of publication June 29, 2017; date of current version February 1, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61379129 and Grant 61671420, in part by the National Key Research and Development Plan of China under Grant 2016YFB0800301, in part by Youth Innovation Promotion Association CAS under Grant 2016394, and in part by the Fundamental Research Funds for the Central Universities. Paper no. TII-17-0517. *(Corresponding author: Kaiping Xue.)*

The authors are with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China (e-mail: lshhsl@mail.ustc.edu.cn; kpxue@ustc.edu.cn; qingyou@mail.ustc.edu.cn; plhong@ustc.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2017.2721542

uation with more than two subsets. EPPA, we have introduced previously, can deal with this kind of set aggregation to some extent in smart grid. However, it cannot hold the assumption that the CC is honest-but-curious [24], [25], which can collude with the residential GW.

In this paper, we propose a privacy-preserving multi-subset aggregation (PPMA) scheme in smart grid, in which, the users set in a residential area are divided into multisubset according to their electricity consumption in each period, and CC can obtain the sum of electricity consumption and the number of users for each subset. Specifically, the main contributions of this paper can be summarized as follows:

1) We propose a privacy-preserving multisubset aggregation scheme for smart grid. To our knowledge, it is the first attempt to achieve multisubset data aggregation in smart grid. Given several consecutive electricity consumption ranges, users can be divided into several subsets, then the proposed scheme can aggregate the number of users and the sum of electricity consumption for each subset, with only one single aggregated data.
2) To protect the users' privacy in smart grid communication network, we introduce a trust third party (TTP) to the smart grid and utilize Paillier homomorphic cryptosystem [16] to aggregate users' electricity consumption data in a privacy-preserving way.
3) Furthermore, we give two extensions to support billing and the scenario with dynamic users.

The rest of this paper is organized as follows: Section II discusses the related work. In Section III, we introduce our system model including system architecture, adversary model, and design goal. Then, in Section IV, we review Paillier cryptosystem and give the definition of subsets as our preliminaries. We present our proposed scheme PPMA in Section V, followed by security and performance analysis in Sections VI and VII, respectively. Section VIII gives two extensions about billing and dynamic users. Finally, Section IX makes a conclusion.

II. Related Work

The cyber security issue is a big obstacle in the way of development of smart grid [6]–[8], [26]. Generic security protection methods in traditional networks, such as key management, source authentication, and integrity protection mechanism have thus been introduced to construct a more secure communication among components in smart grid [27]–[30]. However, owing to the distinctive features of smart grid, users' privacy may be leaked out unconsciously, and result in a more serious loss [9], [10].

To protect users' private information from disclosure, privacy-preserving techniques have been proposed in recent studies. One direct way of privacy protection is hiding users' real identity, Chim *et al.* [31] and Diao *et al.* [32] respectively utilized anonymous credentials to conceal user's real ID during verifying the legality of messages. This can also be achieved by using zero-knowledge proof, which has been presented in the literatures [33], [34]. Except hiding users' real ID to achieve privacy preservation, many schemes focus on data obfuscation, which aims to hide individual electricity consumption data. A widely used method is adding the noise with a given distribution (like Laplace Distribution) into each user's metering data, so as to prevent user's privacy being leaked out by well-designed attacks, such as differential attack [13], [35], [36].

Another way of privacy protection, called data aggregation, is to aggregate the electricity consumption for a specific region without revealing the individual user's consumption data. The existing data aggregation schemes usually utilize homomorphic encryption to aggregate the electricity readings and guarantee the privacy for the users. In the literature [37], Ruj and Nayak proposed a secure framework to aggregate the electricity consumptions of all users in a specific region and achieve privacy preserving with additive homomorphic encryption. Meanwhile, it supports access control based on attribute-based encryption. Li *et al.* [38] proposed an efficient privacy-preserving demand response (EPPDR) scheme which achieves privacy-preserving electricity demand aggregation and efficient response. Chen *et al.* [13] proposed a multifunctional data aggregation scheme (MuDA) to compute multiple statistical functions of users' data, such as summation, variance, etc., in a privacy-preserving way. Since homomorphic encryption usually results in a larger computational overhead, Dong *et al.* [18] considers an efficient ElGamal-based homomorphic encryption, and Abdallah *et al.* [17] proposes a lightweight lattice-based homomorphic data aggregation scheme.

While all these schemes are proposed to aggregate the total electricity consumptions in a residential area, and solve other problems at the same time, e.g., MuDA can compute multifunction, and EPPDR achieves efficient demand response. However, as we pointed out before, when the CC needs to find out the distribution of electricity consumption, none of the existing privacy-preserving schemes can meet this demand. The scheme proposed in [23] can achieve two-subset aggregation, in which the CC is able to obtain the sum of electricity consumption and the corresponding amount of users in each subset, but it cannot be extended to aggregate multisubset (more than two subsets). Our proposed scheme utilizes two superincreasing sequences (similar to EPPA) and Paillier cryptosystem [16], to aggregate multisubset in a privacy-preserving way. At the same time, this scheme can also protect privacy against the malicious GW and CC.

III. System Model

A. System Architecture

In our proposed scheme, there are four entities including SMs, a GW, a CC, and an offline TTP. As our scheme focuses on data aggregation in smart grid, we consider a residential area with n users $\mathbb{U} = \{U_1, U_2, \ldots, U_n\}$. As shown in Fig. 1, there exists one CC and one TTP in our proposed smart grid architecture, and each residential area contains a GW to aggregate users reports.

1) *SMs:* Each user $U_i \in \mathbb{U}$ is equipped with a SM, which is used to collect the corresponding user's electricity consumption data and reporting encrypted data to GW periodically, e.g., every 15 min. We take n SMs (or n users) in a residential area into account in our proposed scheme.
2) *GW:* As a communication relay and aggregator, GW aggregates all data reported by n users, and forwards the

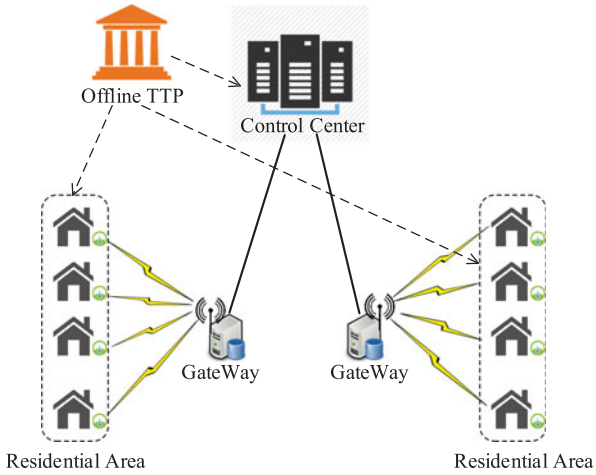


Fig. 1. System architecture of our proposed PPMA.

final result to CC. The existence of GW reduces not only the communication overhead between users and CC, but also the computational cost of CC.

3) *CC:* CC has strong computing capacity, and it is responsible for collecting, processing, and analyzing the data reported by GW. With these information, CC can adjust the electricity generation, transmission, and distribution to meet the dynamic demands.
4) *TTP:* TTP is a fully trusted entity in smart grid, whose duty is to run the *key generation* procedure in Paillier cryptosystem, and further generate and distribute secure parameters for CC and all users in $\mathbb{U}$. In our proposed scheme, TTP only participates in the system setup, and will not involve in the multisubset data aggregation procedure.

B. Adversary Model

In this paper, we assume that there exists an adversary $\mathscr{A}$ residing in the residential area to violate individual user's electricity consumption privacy as possible as he/she can. The adversary $\mathscr{A}$ is capable of launching multiple attacks:

1) $\mathscr{A}$ can eavesdrop the communication between residential users and GW, to get the report data directly. Moreover, the adversary $\mathscr{A}$ can compromise some users in a residential area, which means $\mathscr{A}$ will obtain all the consumption data and the corresponding security parameters of these users. Then, $\mathscr{A}$ will try his/her best to disclose uncompromised users' privacy.
2) $\mathscr{A}$ can intrude in the database of GW and CC to steal the individual user reports and security parameters, and even compromise them. Although CC in reality is very powerful, since the adversary $\mathscr{A}$ is also strong enough under our assumption, the data stored in CC cannot completely avoid the risk of disclosure. Considering CC is curious about the users' privacy, it can be considered as a powerful adversary too.

Surely, the adversary $\mathscr{A}$ can also perform other attacks to obtain information he/she is interested in. However, in this paper, we only focus on protecting individual user's electricity consumption privacy, other attacks are beyond the scope of our work.

C. Design Goal

Our goal is to design an efficient and privacy-preserving multisubset aggregation scheme in smart grid. Specifically, the following three objectives should be achieved.

1) *The functionality of multisubset aggregation needs to be implemented correctly:* The main contribution of our work is to meet such a novel requirement. Different from the traditional aggregation schemes, which aggregate all the report data in a residential area, our proposed PPMA allows CC to know the distribution of electricity consumption and the corresponding number of users in different ranges.
2) *The individual user's electricity consumption privacy should be guaranteed under our adversary model:* The adversary model considers possible security threats to individual user's electricity consumption privacy, and these private data related to users' behavior habits and even home security. For example, low electricity consumption may indicate householders are out of the house. Therefore, our proposed scheme should achieve the privacy-preserving of individual user's electricity consumption.
3) *The high efficiency of computation and communication in the aggregation should be achieved:* Due to the weaker computing power and communication capabilities of SMs, our proposed solution cannot increase the computational complexity and communication overhead while implementing the required functionality. Meanwhile, although CC has strong computing power, as it needs to handle many other tasks, our scheme should not add much more extra computational cost, either.

IV. Preliminaries

A. Paillier Cryptosystem

Paillier cryptosystem [16] is one of the most popular techniques to achieve homomorphic additive encryption. In Paillier cryptosystem, if two integers a and b are encrypted as $E_k(a)$ and $E_k(b)$ with a same key k, there exists a relationship between plaintext operation and ciphertext operation, such that

$$E_k(a) \cdot E_k(b) = E_k(a+b).$$

Generally, Paillier cryptosystem is composed of the following phases: *key generation*, *encryption*, and *decryption*.

1) *Key Generation:* Select two large and independent prime numbers p and q randomly, and we compute $\lambda = lcm(p-1, q-1)$ and $N = p \cdot q$, where λ is the least common multiple of $p-1$ and $q-1$. Then, define a function $L(x) = \dfrac{x-1}{N}$, choose a generator $g = (1+N)$, and compute $\mu = (L(g^{\lambda} \mod N^2))^{-1} \mod N$. The public key is (N, g), and the private key is (λ, μ).
2) *Encryption:* Given message $m \in \mathbb{Z}_N^*$, first, we select a random number $r \in Z_{N^2}^*$, and then the ciphertext can be

computed as follows:

$$c = E(m) = g^m \cdot r^N \mod N^2.$$

3) *Decryption:* Let c be the ciphertext to decrypt, where $c \in \mathbb{Z}^*_{N^2}$, and we can compute the plaintext message as

$$m = L(c^{\lambda} \mod N^2) \cdot \mu \mod N.$$

The homomorphic encryption we utilized in our solution is still Paillier cryptosystem, but we use another form of it. In our proposed scheme, we consider such a ciphertext with the generator $(1 + N)$:

$$c = (1 + N)^m \cdot r^{N \cdot \lambda} \mod N^2$$

where m is the plaintext to be encrypted, and λ is the private key. As $\phi(N^2) = N \cdot \phi(N) = N \cdot \lambda$, we can conclude the following equations according to Euler theorem:

$$\begin{aligned} c &= (1 + N)^m \cdot r^{N \cdot \lambda} \mod N^2 \\ &= (1 + N)^m \cdot r^{\phi(N^2)} \mod N^2 \\ &= (1 + N)^m \mod N^2. \end{aligned} \tag{1}$$

To further decrypt c, we expend the power $(1 + N)^m$ with Binomial theorem $(1 + N)^m = \sum_{i=1}^{m} \binom{m}{i} N^i$, and then simplify the formula with $\mod N^2$, as all items with $i \geq 2$ turn to zero, we get

$$c = (1 + N)^m = 1 + mN \mod N^2. \tag{2}$$

To learn the detailed derivation and proof of Formulas (1) and (2), readers can refer to the literature [16].

B. Definition of Subsets

PPMA realizes the data aggregation of multisubset. Here, we define the subset in detail. We assume that k subsets is equivalent to k consecutive electricity consumption ranges, such that

$$[R_1, R_2), [R_2, R_3), \cdots, [R_k, +\infty).$$

For every user in a residential area, his/her electricity consumption in each period belongs to one of k subsets, for example, if the electricity consumption e satisfies $R_1 \leq e < R_2$, then e is in the first subset. Our scheme aims to compute the total electricity consumptions of each subset in a period, and also the corresponding number of users in each subset.

V. Our Proposed PPMA Scheme

Here, we will present our privacy-preserving multisubset aggregation scheme. In the proposed scheme, the CC can obtain the number of users and the sum of electricity consumptions in each subset, without knowing the individual electricity consumption of each user. We will first give an overview of our proposed PPMA. Then, we will divide the scheme into four steps to elaborate separately.

A. Overview

The purpose of our proposed PPMA is to aggregate the number of users and the total electricity consumptions in each subset, while preventing the individual user's privacy from revealing against malicious adversary. In the system initialization phase, TTP will generate the parameters of Paillier cryptosystem as well as the private parameters of CC and each user in a residential area. During the initialization, the CC and users in a residential area should receive their parameters securely, and leave no information to someone else. Once the system setup is finished, all SMs can report the electricity consumptions to GW periodically. The report needs to be encrypted using the public key of Paillier cryptosystem and private parameter of the individual user. After receiving reports from n SMs, GW aggregates these ciphertexts into a single one, and sends the result to CC. Then, CC decrypts the report with its private parameter, and obtains the number of users as well as the total electricity consumptions in each subset with our designed algorithm.

Our proposed PPMA scheme can preserve the privacy of individual users against malicious adversary. Specifically, GW only has the ciphertext of each report, and CC just knows the number of users in each subset and the corresponding summation of electricity consumptions. Therefore, both GW and CC cannot obtain individual user's electricity consumption data. Moreover, if there has a strong adversary that compromise both GW and CC, as only TTP has the private key of Paillier cryptosystem, the adversary still cannot decrypt individual user's encrypted report. The detailed security analysis will be presented in Section VI.

B. System Setup

There are four entities that include users, a GW, a CC, and an offline trusted third party (TTP) in the system model of our proposed scheme.

1) *TTP:* TTP chooses two distinct large primes p and q randomly, where $|p| = |q| = l$, and then computes the Paillier cryptosystem's public key $(N = pq, g)$ and private key (λ, μ). Then, TTP takes a pseudo-random number generator to generate n random numbers $\{x_1, \ldots, x_n\}$ in $\mathbb{Z}^*_N$, and computes $x_0 \in \mathbb{Z}^*_N$ to satisfy

$$x_0 + \sum_{i=1}^{n} x_i = 0 \mod \lambda. \tag{3}$$

In practice, for providing security, the size of each random numbers should not be less than 1024 bits. Finally, TTP sends x_0 to CC and respectively sends x_i to U_i $(i = 1, 2, \ldots, n)$ via secure channels, then publishes $\{N, g\}$.

2) *CC:* Assume that CC needs to know the aggregation of k consecutive electricity consumption ranges $[R_1, R_2), [R_2, R_3), \cdots, [R_k, E]$, where R_1 is always equals to 0 and E is the maximum electricity consumption, to be noted, $m_i \in [R_i, R_{i+1})$ means that m_i is less than R_{i+1} and greater than or equal to R_i. Then, CC generates two groups of coefficients, the first one is $\{a_1, a_2, \cdots, a_k\}$ satisfied $a_i > \sum_{j=1}^{i-1} a_j \cdot (R_{j+1} - R_j) \cdot n$, where $a_1 = 1$ and $i = 2, 3, \ldots, k$. The second one is $\{b_1, b_2, \ldots, b_k\}$ satisfied $b_i > b_0 + b_0 \cdot n^{i-1}$, where $b_0 > a_k \cdot E \cdot n + \sum_{j=1}^{k-1} a_j \cdot R_{j+1} \cdot n$ and $i = 1, 2, \ldots, n$. Finally, CC chooses a hash function $H : \{0, 1\}^* \rightarrow \mathbb{Z}^*_N$, and publishes $\{(R_1, g^{a_1}, g^{b_1}), (R_2, g^{a_2}, g^{b_2}), \cdots, (R_k, g^{a_k}, g^{b_k}), H\}$.

When the initialization process ends, CC has private x_0 and each user U_i holds his/her private x_i. The parameters published

by CC means that, if the electricity consumption of the user U_i belongs to the jth subset in some period, U_i should use private x_i and the jth parameters (g^{a_j}, g^{b_j}) as the encryption parameters.

C. Encrypt Electricity Consumption

As user's consumption data are reported to GW periodically, e.g., every 15 min, to protect user's privacy from disclosure, a SM needs to encrypt these private data. From the information that CC has published, SM can know what ranges CC wants to collect. Assume that all residential users are $\mathbb{U} = \{U_1, U_2, \ldots, U_n\}$. If U_i's electricity consumption data $m_i \in [R_j, R_{j+1})$, U_i is denoted to lie in the subset $\mathbb{U}_j \subset \mathbb{U}$. Obviously, $\mathbb{U} = \mathbb{U}_1 \cup \mathbb{U}_2 \cup \cdots \cup \mathbb{U}_k$, and $\mathbb{U}_i \cap \mathbb{U}_j = \varnothing$, for $i, j = 1, 2, \ldots, k, i \neq j$.

For each $U_i \in \mathbb{U}$, as every period is small, it is reasonable to assume that m_i lies in $\{0, 1, 2, \ldots, E\}$. At each time interval t, without loss of generality, we suppose $m_i \in [R_j, R_{j+1})$, and $\Delta m_i = m_i - R_j$. U_i computes the ciphertext with the secret key x_i and (g^{a_j}, g^{b_j})

$$c_i = g^{a_j \cdot \Delta m_i} \cdot g^{b_j} \cdot H(t)^{N \cdot x_i} \mod N^2. \tag{4}$$

Then, U_i sends the ciphertext c_i to GW.

D. Aggregate Data

After receiving all the c_i from the users, GW performs the following aggregation process:

$$\begin{aligned} C &= \prod_{i=1}^{n} c_i \\ &= g^{a_1 \cdot \sum\limits_{i \in \mathbb{U}_1} \Delta m_i + a_2 \cdot \sum\limits_{i \in \mathbb{U}_2} \Delta m_i + \cdots + a_k \cdot \sum\limits_{i \in \mathbb{U}_k} \Delta m_i} \\ &\quad \cdot g^{b_1 \cdot |\mathbb{U}_1| + b_2 \cdot |\mathbb{U}_2| + \cdots + b_k \cdot |\mathbb{U}_k|} \cdot H(t)^{N \cdot \sum\limits_{i=1}^{n} x_i} \mod N^2. \end{aligned} \tag{5}$$

Then, GW sends the aggregated ciphertext C to CC.

E. Decrypt Aggregated Data

Once receiving the aggregated ciphertext C, CC performs the following steps to recover the aggregated data, and compute the number of users and the sum of electricity consumptions in each subset from it.

1) *Step 1:* CC uses its secret key x_0 to compute:

$$\begin{aligned} V &= C \cdot H(t)^{N \cdot x_0} \\ &= g^{a_1 \cdot \sum\limits_{i \in \mathbb{U}_1} \Delta m_i + a_2 \cdot \sum\limits_{i \in \mathbb{U}_2} \Delta m_i + \cdots + a_k \cdot \sum\limits_{i \in \mathbb{U}_k} \Delta m_i} \\ &\quad \cdot g^{b_1 \cdot |\mathbb{U}_1| + b_2 \cdot |\mathbb{U}_2| + \cdots + b_k \cdot |\mathbb{U}_k|} \\ &\quad \cdot H(t)^{N \cdot \left(x_0 + \sum\limits_{i=1}^{n} x_i \right)} \mod N^2 \\ &\xrightarrow{x_0 + \sum\limits_{i=1}^{n} x_i = 0 \mod \lambda} \\ &= g^{a_1 \cdot \sum\limits_{i \in \mathbb{U}_1} \Delta m_i + a_2 \cdot \sum\limits_{i \in \mathbb{U}_2} \Delta m_i + \cdots + a_k \cdot \sum\limits_{i \in \mathbb{U}_k} \Delta m_i} \\ &\quad \cdot g^{b_1 \cdot |\mathbb{U}_1| + b_2 \cdot |\mathbb{U}_2| + \cdots + b_k \cdot |\mathbb{U}_k|} \mod N^2. \end{aligned} \tag{6}$$

Algorithm 1: Recover(D).

for $i = k$ to 1 **do**
  $|\mathbb{U}_i| = (D - D \mod b_i)/b_i$
  $D = D - (b_i \cdot |\mathbb{U}_i|)$
end for
for $i = k$ to 1 **do**
  $M = (D - D \mod a_i)/a_i$
  $D = D - (a_i \cdot M)$
  $M_i = M + R_i \cdot |\mathbb{U}_i|$
end for
return $\{|\mathbb{U}_1|, |\mathbb{U}_2|, \cdots, |\mathbb{U}_k|\}$,$\{M_1, M_2, \cdots, M_k\}$

2) *Step 2:* Then, we let

$$\begin{aligned} D = &\left(a_1 \cdot \sum_{i \in \mathbb{U}_1} \Delta m_i + a_2 \cdot \sum_{i \in \mathbb{U}_2} \Delta m_i + \cdots + a_k \cdot \sum_{i \in \mathbb{U}_k} \Delta m_i \right) \\ &+ (b_1 \cdot |\mathbb{U}_1| + b_2 \cdot |\mathbb{U}_2| + \cdots + b_k \cdot |\mathbb{U}_k|). \end{aligned} \tag{7}$$

CC can recover the aggregated data as follows:

$$D = \frac{V - 1}{N} \mod N^2. \tag{8}$$

3) *Step 3:* To recover the number of users and the sum of electricity consumptions in each subset from D, CC executes Algorithm 1, and obtains $\{|\mathbb{U}_1|, |\mathbb{U}_2|, \ldots, |\mathbb{U}_k|\}$, $\{M_1, M_2, \ldots, M_k\}$, where $|\mathbb{U}_i|$ is the number of users lies in $\mathbb{U}_i$, and M_i is the total electricity consumptions of users lies in $\mathbb{U}_i$. Obviously, $|\mathbb{U}| = \sum_{i=1}^{k} |\mathbb{U}_i|$ is the number of all users, and $M = \sum_{i=1}^{k} M_i$ is the total electricity consumptions of all users in a residential area.

In each time period, these four procedures will be executed repeatedly. Based on the result obtained from $Step3$, CC can dig out the power usage trends, the distribution of electricity consumption or use machine learning and other methods to further analyze the data. The ultimate goal is to improve electricity utilization, develop reasonable dynamic price, etc., in which the most important one is to prevent blackouts.

VI. Correctness and Security Analysis

In this section, we discuss the security issues of our proposed scheme, specifically, to protect users privacy against a strong adversary $\mathscr{A}$. As mentioned in our adversary model, the strong adversary $\mathscr{A}$ can be able to launch multiple attacks to violate individual user's privacy. In the following, we will first prove the correctness of our scheme, and then analyze data privacy protection from the aspects of attacks that may be initiated by the adversary.

A. Proof of Correctness

The correctness of our scheme depends on whether CC can decrypt aggregated data correctly. To proof that, we here demonstrate Formulas (6), (8) and Algorithm 1. Note that, the other processes are clearly correct, so we do not need to give additional proofs of them.

1) *Formula (6):* With CC's private key x_0, this formula eliminates the term containing $H(t)$ in C. We now bring D in Formula (7) into Formula (5), and get

$$C = g^D \cdot H(t)^{N \cdot \sum_{i=1}^{n} x_i} \mod N^2.$$

Based on Formula (3), we can assume that $x_0 + \sum_{i=1}^{n} x_i = k \cdot \lambda$, where $k \in \mathbb{Z}_N^*$. We reintroduce Formula (6) as follows:

$$\begin{aligned} V &= C \cdot H(t)^{N \cdot x_0} \\ &= g^D \cdot H(t)^{N \cdot (x_0 + \sum_{i=1}^{n} x_i)} \\ &= g^D \cdot H(t)^{N \cdot k \cdot \lambda} \\ &= g^D \cdot (H(t)^k)^{N \cdot \lambda} \\ &\xrightarrow{r = H(t)^k} \\ &= g^D \cdot r^{N \cdot \lambda} \mod N^2. \end{aligned}$$

According to the description in our preliminaries, we can get $V = g^D \mod N^2$.

2) *Formula (8):* From the description in our preliminaries, we can get

$$\begin{aligned} V &= g^D \mod N^2 \\ &= (1+N)^D \mod N^2 \\ &= 1 + N \cdot D \mod N^2 \\ \Rightarrow D &= \frac{V-1}{N} \mod N^2. \end{aligned}$$

3) *Algorithm 1:* This algorithm has two *for* loops, the first one recovers $\mathbb{U}_i, i = 1, 2, \ldots, k$, and the second one recovers $M_i, i = 1, 2, \ldots, k$. Here, we will indicate how these two loops work.

1) To explain the first *for* loop, as an example, we first consider $\mathbb{U}_k$. Since

$$\begin{aligned} & a_1 \cdot \sum_{i \in \mathbb{U}_1} \Delta m_i + a_2 \cdot \sum_{i \in \mathbb{U}_2} \Delta m_i + \cdots + a_k \cdot \sum_{i \in \mathbb{U}_k} \Delta m_i \\ & \le a_1 \cdot \sum_{i \in \mathbb{U}_1} (R_2 - R_1) + a_2 \cdot \sum_{i \in \mathbb{U}_2} (R_3 - R_2) + \cdots \\ & \quad + a_k \cdot \sum_{i \in \mathbb{U}_k} (E - R_k) \\ & < a_k \cdot (E - R_k) \cdot n + \sum_{j=1}^{k-1} a_j \cdot (R_{j+1} - R_j) \cdot n \\ & < b_0 \end{aligned}$$

and because

$$\begin{aligned} & b_1 \cdot |\mathbb{U}_1| + b_2 \cdot |\mathbb{U}_2| + \cdots + b_{k-1} \cdot |\mathbb{U}_{k-1}| \\ & < b_{k-1} \cdot |\mathbb{U}_1| + b_{k-1} \cdot |\mathbb{U}_2| + \cdots + b_{k-1} \cdot |\mathbb{U}_{k-1}| \\ & \le b_{k-1} \cdot n \end{aligned}$$

we can get

$$\begin{aligned} & a_1 \cdot \sum_{i \in \mathbb{U}_1} \Delta m_i + a_2 \cdot \sum_{i \in \mathbb{U}_2} \Delta m_i + \cdots + a_k \cdot \sum_{i \in \mathbb{U}_k} \Delta m_i \\ & \quad + b_1 \cdot |\mathbb{U}_1| + b_2 \cdot |\mathbb{U}_2| + \cdots + b_{k-1} \cdot |\mathbb{U}_{k-1}| \\ & < b_0 + b_{k-1} \cdot n \\ & < b_k. \end{aligned}$$

Therefore, we can get $(D - D \mod b_k)/b_k = (b_k \cdot |\mathbb{U}_k|)/b_k = |\mathbb{U}_k|$, and by using the same method, we can recover all $|\mathbb{U}_i|, i = 1, 2, \ldots, k$.

2) After recovering the number of users in each subset, D will be equal to $a_1 \cdot \sum_{i \in \mathbb{U}_1} \Delta m_i + a_2 \cdot \sum_{i \in \mathbb{U}_2} \Delta m_i + \cdots + a_k \cdot \sum_{i \in \mathbb{U}_k} \Delta m_i$. For recovering M_k, since any $\Delta m_i, i \in \mathbb{U}_j$ is less than $(R_{j+1} - R_j)$, we have

$$\begin{aligned} & a_1 \cdot \sum_{i \in \mathbb{U}_1} \Delta m_i + a_2 \cdot \sum_{i \in \mathbb{U}_2} \Delta m_i + \cdots + a_{k-1} \cdot \sum_{i \in \mathbb{U}_{k-1}} \Delta m_i \\ & \le a_1 \cdot \sum_{i \in \mathbb{U}_1} (R_2 - R_1) + a_2 \cdot \sum_{i \in \mathbb{U}_2} (R_3 - R_2) + \cdots \\ & \quad + a_{k-1} \cdot \sum_{i \in \mathbb{U}_{k-1}} (R_k - R_{k-1}) \\ & < \sum_{j=1}^{k-1} a_j \cdot (R_{j+1} - R_j) \cdot n \\ & < a_k. \end{aligned}$$

Therefore, we can further get

$$(D - D \mod a_k)/a_k = \left(a_k \cdot \sum_{i \in \mathbb{U}_k} \Delta m_i \right) / a_k = \sum_{i \in \mathbb{U}_k} \Delta m_i$$

and

$$M_k = \sum_{i \in \mathbb{U}_k} (R_k + \Delta m_i) = R_k \cdot |\mathbb{U}_k| + \sum_{i \in \mathbb{U}_k} \Delta m_i.$$

With the same processes, we can obtain $\{M_1, M_2, \cdots, M_k\}$ finally.

B. Eavesdropping Cannot Reveal Individual User's Electricity Consumption Privacy

The adversary $\mathscr{A}$ can lurk in a residential area to eavesdrop on communication data between the users and GW. Assume that the adversary $\mathscr{A}$ has acquired a report data of the user U_i, i.e., c_i. In our proposed scheme, user U_i's consumption data sensed by the corresponding SM can be formed as $c_i = g^{a_j \cdot \Delta m_i} \cdot g^{b_j} \cdot H(t)^{N \cdot x_i} \mod N^2$. Let $m_i = a_j \cdot \Delta m_i + b_j$ and $r_i = H(t)^{x_i}$, then the ciphertext $c_i = g^{m_i} \cdot r_i^N \mod N^2$ is still a legal ciphertext of Paillier cryptosystem. Since Paillier cryptosystem is semantic secure against the chosen plaintext attack, the adversary $\mathscr{A}$ is not able to recover U_i's private data, i.e., a_j, Δm_i, and b_j. Thus, the individual user's electricity consumption privacy is protected from eavesdropping.

C. Compromising Some Users Cannot Reveal Other User's Electricity Consumption Privacy

A strong adversary $\mathscr{A}$ can compromise some users in a residential area. In our system, considering n users in one residential area, the adversary $\mathscr{A}$ is able to compromise k ($k \leq n-1$) users, where "compromise" means adversary $\mathscr{A}$ can disclose a user's private data as well as the secret parameter x. Since all secret parameters are randomly generated by TTP, they do not have any correlation with each other. Knowing one or some users' secret parameters reveals nothing about the others'. Consider such an extreme situation, that the adversary $\mathscr{A}$ compromise $n-1$ users, i.e., $U_1, U_2, \ldots, U_{n-1}$. Then, the adversary obtains $x_1, x_2, \cdots, x_{n-1}$. Since we have $x_0 + x_n + \sum_{i=1}^{n-1} x_i = 0 \mod \lambda$, without any knowledge of x_0 and λ, the adversary $\mathscr{A}$ still cannot reveal x_n. Thus, no matter how many users the adversary has compromised, he/she cannot disclose the electricity consumption privacy of uncompromised users.

D. Compromising GW Cannot Reveal Individual User's Electricity Consumption Privacy and the Aggregation Result

After receiving encrypted data from residential users periodically, GW can aggregate these ciphertexts by computing $C = \prod_{i=1}^{n} c_i \mod N^2$. Thus, if the adversary $\mathscr{A}$ compromises GW, or hacks into the database, he/she can only get the encrypted consumption data and the aggregated ciphertext. As analyzed in Section VI-B, due to the semantic security of Paillier cryptosystem, the ciphertext reveals nothing about the plaintext, so the individual user's electricity consumption privacy is preserved. Moreover, the aggregated ciphertext C, as shown in Formula (5), has the same form with user's report data, which means aggregated ciphertext C is also a valid ciphertext of Paillier cryptosystem. Since the adversary $\mathscr{A}$ cannot obtain any security parameters from GW, he/she is not able to disclose the aggregation result. Thus, even though GW is compromised, the individual user's electricity consumption privacy can be ensured.

E. Compromising CC Cannot Reveal Individual User's Electricity Consumption Privacy

After receiving an aggregated ciphertext from GW, CC decrypts it and performs Algorithm 1 to obtain $\{|\mathbb{U}_1|, |\mathbb{U}_2|, \ldots, |\mathbb{U}_k|\}$, $\{M_1, M_2, \ldots, M_k\}$, where $|\mathbb{U}_i|$ is the number of users lying in $\mathbb{U}_i$, and M_i is the total electricity consumptions of users lying in $\mathbb{U}_i$. If a strong adversary $\mathscr{A}$ compromises CC, he/she can get all these information. However, in our system, CC contains just one security parameter x_0, and this parameter can only be used to decrypt the aggregated ciphertext, different from other aggregation schemes, CC has no private key, i.e., λ, μ, of Paillier cryptosystem. Instead, we let TTP to retain them. Therefore, the adversary $\mathscr{A}$ cannot decrypt ciphertexts obtained from attacks mentioned in Sections VI-B or VI-D. Moreover, as $\{|\mathbb{U}_1|, |\mathbb{U}_2|, \ldots, |\mathbb{U}_k|\}$ and $\{M_1, M_2, \ldots, M_k\}$ are the aggregation results, even if the adversary $\mathscr{A}$ obtains them, he/she cannot associate these data with individual user's electricity consumption. Thus, individual user's electricity consumption privacy will not be revealed even though CC is compromised.

TABLE I
COMPUTATIONAL COMPLEXITY

	SM	GW	CC
PPMA	$1 \times O_e + 1 \times O_m$	$(n-1) \times O_m$	$1 \times O_e + 1 \times O_m$
Paillier	$1 \times O_e + 1 \times O_m$	$(n-1) \times O_m$	$1 \times O_e$
PPSDA	$1 \times O_e + 1 \times O_m$	$(n-1) \times O_m$	$1 \times O_m + 1 \times O_e + DEC$

Note: $1 \times O_e$ means one exponentiation operation in $\mathbb{Z}_{N^2}$.$1 \times O_m$ means one multiplication operation in $\mathbb{Z}_{N^2}$.DEC means the special operation for decryption in PPSDA.

VII. PERFORMANCE EVALUATION

In this section, we first measure the performance of our proposed scheme in terms of the computation overhead of SM, GW, and CC. In addition, since the message recover algorithm (refer to Algorithm 1) is carried out on plaintext, the number of subsets has almost no effects on computing overhead, so we set the number of subsets to two, and compare the computation cost with privacy-preserving set data aggregation (PPSDA) scheme in [23], which only supports two subsets aggregation. We also consider a traditional approach with the standard Paillier cryptosystem (simply denoted as the Paillier in this paper), which only aggregates overall electricity consumptions. Moreover, in this section, we also give a brief performance analysis of communication and storage overhead comparison.

Specifically, we implement our scheme by Java and run our experiments on a computer with 3.1 GHz processor, 12 GB memory, and Ubuntu 16.04 platform. We use 512-bit length primes p and q for Paillier homomorphic cryptosystem, and limit the maximum electricity consumption data in a period to 100. To be noted, for security reasons, we suggest that the length of primes p and q should be at least 1024-bit in practical use.

A. Computational Complexity and Efficiency

When a residential user U_i needs to encrypt the data, it requires one exponentiation and one multiplication operation in $\mathbb{Z}_{N^2}$ to generate c_i. To be noted, the complexity of two multiplication operations in $\mathbb{Z}_{N^2}$ can be negligible compared with the exponentiation operations. After receiving the ciphertext from n users, GW should perform $n-1$ multiplication operations in $\mathbb{Z}_{N^2}$ and forward the result to CC. However, the overhead of these operations cannot be ignored, as n is usually hundreds or even thousands. To decrypt the aggregated ciphertext, different from the standard decryption of Paillier cryptosystem, which requires $L(x)$ function, in our scheme, CC needs only one exponentiation operation. To further recover the number of users and the sum of consumption data in each subset, CC just performs several or (up to) dozens of (depend on the size of k) subtraction or division operations in $\mathbb{Z}_{N^2}$; thus, these operations can be considered to have less impact on the overall performance overhead.

We compare the computational complexity of PPMA with the Paillier and PPSDA, as shown in Table I. We can see that, the

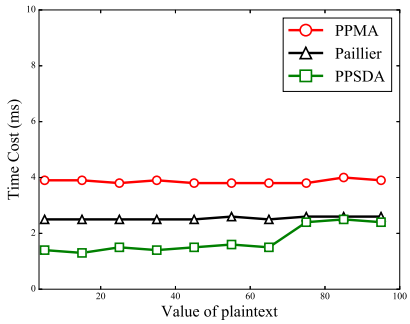


Fig. 2. Encryption cost at SM.

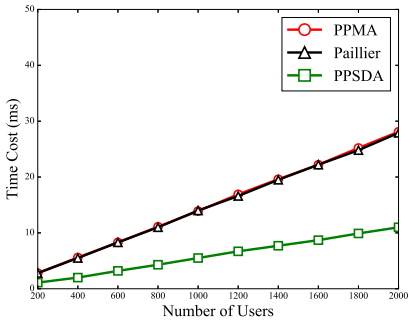


Fig. 3. Aggregation cost at GW.

computing overheads of SM and GW in these three schemes are consistent. As for CC, our proposed PPMA achieves the lowest computing cost compared to the other two schemes.

We also compare the computing efficiency of our PPMA scheme in SM, GW, and CC with the Paillier and PPSDA. In our proposed scheme, as we fix the maximum electricity consumption data, there are actually two factors that may affect the performance obviously, i.e., the number of users and the number of subsets.

In these three schemes, the number of users will not affect the encryption cost obviously, and the number of subsets may only have impact on our scheme, as the other two schemes have fixed number of subsets (the Paillier has only one set and PPSDA has two). We set the number of users to 500 and the number of subsets to two, and execute the experiments, the results are showed in Fig. 2. From the figure, we can see that the encryption cost of our scheme changes little with the increase of electricity consumption value, which means the number of subsets also has little impact on the overhead of encryption, as the number of subsets only affects the size of report data of different electricity consumption. The encryption cost of the Paillier is stable at about 2.6 ms, and PPSDA is overall optimal. In general, the Paillier and PPSDA are slightly more efficient than our proposed scheme, but the difference is small (about 2 ms), which can be accepted.

Fig. 3 shows the comparison result of the aggregation computation cost of our proposed scheme and the other two schemes. As we have analyzed above, the number of subsets has no effect on the aggregation computation cost of our scheme, so we set it to two and conduct the experiment. In Fig. 3, the aggregation function in PPSDA is more efficient than that in our scheme and the Paillier, and our scheme almost has the same aggregation

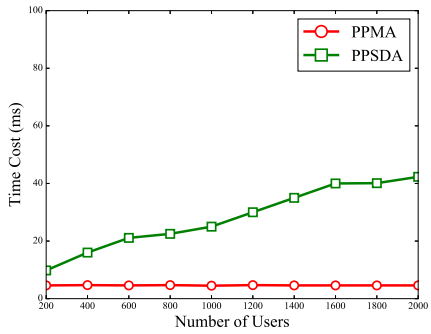


Fig. 4. Decryption cost at CC.

cost with the Paillier. Note that, when the total number of users increases to 2000, the aggregation in our scheme costs 27 ms and PPSDA costs 12 ms, although the latter is intuitively half the cost of the former, in practice, since the aggregation is only required to be performed once in each period, both of them are efficient enough to aggregate data in time.

In order to explore the impact of the number of users on the decryption overhead at CC, we tested our proposed scheme with the number of users changes from 200 to 2000 and compared it with PPSDA. Fig. 4 shows the comparison results in terms of the decryption computation cost. From the figure, we found that the decryption cost of our proposed scheme remains at around 4.60 ms regardless of the number of users. This is extremely efficient and practical, as CC usually needs to process a large amount of user data reported by different GWs in every period. From the figure, we can see that the decryption overhead of our proposed scheme is smaller relative to PPSDA, and the computation cost of PPSDA goes up rapidly as the number of users increases.

B. Communication and Storage Overhead

We analyze the communication overhead from two aspects: SM-to-GW and GW-to-CC. In every period, each SM reports one ciphertext of consumption information to GW, the overhead is only associated with the size of ciphertext. The ciphertexts in different schemes are only a little different in size, especially for PPMA and PPSDA. For the GW-to-CC communication, as the same as being analyzed above, there is no much more extra cost.

We ignore the storage overhead of GW and CC, as they always have a lot of storage space compared to SMs. Due to the limited resources of SMs, it becomes a major consideration of storage overhead in smart grid. Compared to other schemes with only data aggregation, each SM in PPMA requires additional storage of one secure key x and the parameters of k subsets. As both the secure key and each parameter size is 1024 bits, the extra storage size in each SM is $(k+1) \times 1024$ bits approximately. As the number of subsets is generally from several to dozens, the required additional storage size will be just a few KB. Therefore, PPMA will not greatly increase the storage overhead.

VIII. Extensions

Our PPMA scheme can aggregate multisubset data of all users in one time point. However, sometimes we need to aggregate

one user's data of a number of continuous time periods, e.g., to charge users [39]. Our work can be easily modified to support such aggregation in a privacy-preserving way. In addition, dynamic users, i.e., user addition or user removal, is quite normal in reality, and our scheme can also be modified to support that.

A. Extension to Support Billing

No matter in traditional grid or current smart grid, how to charge users is always a realistic need. Here, we give an extension scheme to support billing. Assume a billing cycle contains τ time periods, i.e., from t_1 to t_τ, from $t_{\tau+1}$ to $t_{2\tau}$, etc. To clearly give our extension, we consider charging U_i of the first billing cycle.

During this billing cycle, the ciphertext U_i report in t_j is c_{ij} $(1 \le j \le \tau)$, and all these ciphertexts will be stored on the GW. While computing c_{ij}, a temporary value $H(t_j)^{x_i}$ will also be computed, which U_i will store locally.

When billing is required, U_i first computes

$$h = \prod_{j=1}^{\tau} H(t_j)^{x_i} = \left(\prod_{j=1}^{\tau} H(t_j)\right)^{x_i}$$

sends h to TTP. Then, GW receives h^{-1} from TTP and computes

$$\begin{aligned} V' &= \left(\prod_{j=1}^{\tau} c_{ij}\right)/h \mod N^2 \\ &= g^{D'} \mod N^2 \end{aligned}$$

where

$$\begin{aligned} D' = &\left(a_1 \cdot \sum_{i \in \mathbb{U}_1} \Delta m_{ij} + a_2 \cdot \sum_{i \in \mathbb{U}_2} \Delta m_{ij} + \cdots + a_k \cdot \sum_{i \in \mathbb{U}_k} \Delta m_{ij}\right) \\ &+ (b_1 \cdot |\mathbb{U}_1| + b_2 \cdot |\mathbb{U}_2| + \cdots + b_k \cdot |\mathbb{U}_k|). \end{aligned}$$

GW can decrypt V' to get D' with Formula (8). By utilizing Algorithm 1 with the input D', GW or CC can recover the sum usage $\sum_{j=1}^{\tau} m_{ij}$ for U_i. Finally, CC can charge U_i with a fee of $p_1 \cdot \sum_{j=1}^{\tau} m_{ij}$, where p_1 is the electricity price of the first billing cycle.

Note that, the key of billing is to recover the sum usage of U_i correctly, and this is guaranteed by the coefficients a_i and b_i $(i = 1, 2, \ldots, k)$. In our basic PPMA scheme, these coefficients are associated with the number of users n and ranges. However, in this extension, they are the number of periods τ and ranges. Therefore, CC should derive these coefficients with $\max\{n, \tau\}$ instead of n, to make sure both the basic scheme and the extension could work.

B. Extension to Support Dynamic Users

User increasing and decreasing will change the number of users in an aggregation region, here we give an extension to support this dynamic scenario. When a set of new users $\mathbb{U}_a$ are added into the system and a set of old users $\mathbb{U}_r$ are removed from the system, TTP first chooses a subset of users $\mathbb{U}_c$ that still remain in the system, and then generates each $U_a \in \mathbb{U}_a$ with a new private key x_{ua} and computes

$$\sum_{U_c \in \mathbb{U}_c} x'_{uc} = \sum_{U_c \in \mathbb{U}_c} x_{uc} + \sum_{U_r \in \mathbb{U}_r} x_{ur} - \sum_{U_a \in \mathbb{U}_a} x_{ua}$$

where x_{ur} is the private key of each $U_r \in \mathbb{U}_r$ that will be revoked, x_{uc} is the old private key of $U_c \in \mathbb{U}_c$, and x'_{uc} is the new one generated by the TTP based on the above equation. All the x'_{uc} will be sent to each $U_c \in \mathbb{U}_c$ in a secure channel to replace the old x_{uc}.

It is important to note that the subset of users $\mathbb{U}_c$ should be chosen careful. Considering the security issue in Section VI-C, the size should be large enough to prevent possible attacks. This is actually a tradeoff between security and performance.

IX. Conclusion and Future Work

In this paper, for the first time we have defined the subsets as many consecutive electricity consumption ranges, and underlined the necessity of aggregating these subsets to CC. As an initial attempt to achieve multisubset data aggregation, we have proposed PPMA to aggregate the count of users and the sum of electricity consumption in each subset. In particular, detailed performance evaluation and security analysis show that our PPMA scheme preserves the privacy of users' electricity consumption and achieves lower computation overhead at the same time.

Currently, we have not considered the possible user failures, which means some SMs may be damaged and they cannot ensure to report data correctly. As a result, CC cannot further decrypt ciphertext correctly. Furthermore, how to support dynamic pricing is still a challenge issue, and we will continue to study to address these problems in our future work.

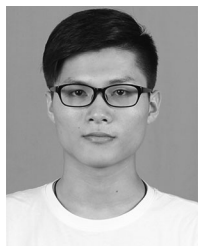
Acknowledgment

The authors sincerely thank the anonymous referees for their invaluable suggestions that have led to the present improved version of the original manuscript.

References

[1] V. C. Gungor, B. Lu, and G. P. Hancke, "Opportunities and challenges of wireless sensor networks in smart grid," *IEEE Trans. Ind. Electron.*, vol. 57, no. 10, pp. 3557–3564, Oct. 2010.
[2] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid—The new and improved power grid: A survey," *IEEE Commun. Surv. Tuts.*, vol. 14, no. 4, pp. 944–980, Fourth Quart. 2012.
[3] Y. Simmhan *et al.*, "Cloud-based software platform for big data analytics in smart grids," *Comput. Sci. Eng.*, vol. 15, no. 4, pp. 38–47, 2013.
[4] K. Sakurama and M. Miura, "Communication-based decentralized demand response for smart microgrids," *IEEE Trans. Ind. Electron.*, vol. 64, no. 6, pp. 5192–5202, Jun. 2017.
[5] S. Aman, Y. Simmhan, and V. K. Prasanna, "Improving energy use forecast for campus micro-grids using indirect indicators," in *Proc. 11th IEEE Int. Conf. Data Mining Workshops*, 2011, pp. 389–397.
[6] H. Khurana, M. Hadley, N. Lu, and D. A. Frincke, "Smart-grid security issues," *IEEE Security Privacy*, vol. 8, no. 1, pp. 81–85, 2010.
[7] D. B. Rawat and C. Bajracharya, "Cyber security for smart grid systems: Status, challenges and perspectives," in *Proc. IEEE Region 3 SoutheastCon Conf.*, 2015, pp. 1–6.
[8] X. Li *et al.*, "Securing smart grid: Cyber attacks, countermeasures, and challenges," *IEEE Commun. Mag.*, vol. 50, no. 8, pp. 38–45, Aug. 2012.

[9] Y. Yan, Y. Qian, H. Sharif, and D. Tipper, "A survey on cyber security for smart grid communications." *IEEE Commun. Surveys Tuts.*, vol. 14, no. 4, pp. 998–1010, Fourth Quart. 2012.
[10] A. Molina-Markham *et al.*, "Private memoirs of a smart meter," in *Proc. 2nd ACM Workshop Embedded Sens. Syst. Energy-Efficiency Build.*, 2010, pp. 61–66.
[11] F. D. Garcia and B. Jacobs, "Privacy-friendly energy-metering via homomorphic encryption," in *Proc. Int. Workshop Security Trust Manage.*, 2010, pp. 226–238.
[12] L. Chen, R. Lu, and Z. Cao, "PDAFT: A privacy-preserving data aggregation scheme with fault tolerance for smart grid communications," *Peer-to-Peer Netw. Appl.*, vol. 8, no. 6, pp. 1122–1132, 2015.
[13] L. Chen *et al.*, "MuDA: Multifunctional data aggregation in privacy-preserving smart grid communications," *Peer-to-Peer Netw. Appl.*, vol. 8, no. 5, pp. 777–792, 2015.
[14] Q. Yang *et al.*, "A privacy-preserving and real-time traceable power request scheme for smart grid," in *Proc. IEEE Int. Conf. Commun.*, 2017, pp. 1–6.
[15] R. Lu *et al.*, "EPPA: An efficient and privacy-preserving aggregation scheme for secure smart grid communications," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 9, pp. 1621–1631, Sep. 2012.
[16] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Proc. Int. Conf. Theory Appl. Cryptographic Techn.*, 1999, pp. 223–238.
[17] A. Abdallah and X. Shen, "A lightweight lattice-based homomorphic privacy-preserving data aggregation scheme for smart grid," *IEEE Trans. Smart Grid*, to be published, doi: 10.1109/TSG.2016.2553647.
[18] X. Dong *et al.*, "An elgamal-based efficient and privacy-preserving data aggregation scheme for smart grid," in *Proc. IEEE Global Commun. Conf.*, 2014, pp. 4720–4725.
[19] Z. Zhu *et al.*, "An integer linear programming based optimization for home demand-side management in smart grid," in *Proc. IEEE PES Innovative Smart Grid Technol.*, 2012, pp. 1–5.
[20] S. Moon and J.-W. Lee, "Multi-residential demand response scheduling with multi-class appliances in smart grid," *IEEE Trans. Smart Grid*, to be published, doi: 10.1109/TSG.2016.2614546.
[21] P. Luh, Y. Ho, and R. Muralidharan, "Load adaptive pricing: An emerging tool for electric utilities," *IEEE Trans. Autom. Control*, vol. 27, no. 2, pp. 320–329, Apr. 1982.
[22] C. Chen, S. Kishore, and L. V. Snyder, "An innovative RTP-based residential power scheduling scheme for smart grids," in *Proc. IEEE Int,. Conf. Acoust., Speech,Signal Process.*, 2011, pp. 5956–5959.
[23] R. Lu, K. Alharbi, X. Lin, and C. Huang, "A novel privacy-preserving set aggregation scheme for smart grid communications," in *Proc. IEEE Global Commun. Conf.*, 2015, pp. 1–6.
[24] Z. Yang, S. Yu, W. Lou, and C. Liu, "P^2: Privacy-preserving communication and precise reward architecture for V2G networks in smart grid," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 697–706, Dec. 2011.
[25] H. J. Jo, I. S. Kim, and D. H. Lee, "Efficient and privacy-preserving metering protocols for smart grid systems," *IEEE Trans. Smart Grid*, vol. 7, no. 3, pp. 1732–1742, May 2016.
[26] N. Saxena and B. J. Choi, "State of the art authentication, access control, and secure integration in smart grid," *Energies*, vol. 8, no. 10, pp. 11883–11915, 2015.
[27] T. W. Chim, S. M. Yiu, L. C. K. Hui, and V. O. K. Li, "PASS: Privacy-preserving authentication scheme for smart grid network," in *Proc. IEEE Int. Conf. Smart Grid Commun.*, 2011, pp. 196–201.
[28] N. Liu *et al.*, "A key management scheme for secure communications of advanced metering infrastructure in smart grid," *IEEE Trans. Ind. Electron.*, vol. 60, no. 10, pp. 4746–4756, Oct. 2013.
[29] Z. Wan, G. Wang, Y. Yang, and S. Shi, "SKM: Scalable key management for advanced metering infrastructure in smart grids," *IEEE Trans. Ind. Electron.*, vol. 61, no. 12, pp. 7055–7066, Dec. 2014.
[30] T. W. Chim, S.-M. Yiu, V. O. Li, L. C. Hui, and J. Zhong, "PRGA: Privacy-preserving recording & gateway-assisted authentication of power usage information for smart grid," *IEEE Trans. Dependable Secure Comput.*, vol. 12, no. 1, pp. 85–97, Jan./Feb. 2015.
[31] T. W. Chim, S.-M. Yiu, L. C. Hui, and V. O. Li, "Privacy-preserving advance power reservation," *IEEE Commun. Mag.*, vol. 50, no. 8, pp. 18–23, Aug. 2012.
[32] F. Diao, F. Zhang, and X. Cheng, "A privacy-preserving smart metering scheme using linkable anonymous credential," *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 461–467, Jan. 2015.
[33] D. Mashima and A. Roy, "Privacy preserving disclosure of authenticated energy usage data," in *Proc. IEEE Int. Conf. Smart Grid Commun.*, 2014, pp. 866–871.
[34] Y. Gong, Y. Cai, Y. Guo, and Y. Fang, "A privacy-preserving scheme for incentive-based demand response in the smart grid," *IEEE Trans. Smart Grid*, vol. 7, no. 3, pp. 1304–1313, May 2016.
[35] Z. Shi *et al.*, "Diverse grouping-based aggregation protocol with error detection for smart grid communications," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 2856–2868, Nov. 2015.
[36] H. Bao and R. Lu, "A new differentially private data aggregation with fault tolerance for smart grid communications," *IEEE Internet Things J.*, vol. 2, no. 3, pp. 248–258, 2015.
[37] S. Ruj and A. Nayak, "A decentralized security framework for data aggregation and access control in smart grids," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 196–205, Mar. 2013.
[38] H. Li *et al.*, "EPPDR: An efficient privacy-preserving demand response scheme with adaptive key evolution in smart grid," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 8, pp. 2053–2064, Aug. 2014.
[39] X. Liang *et al.*, "UDP: Usage-based dynamic pricing with privacy preservation for smart grid," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 141–150, Mar. 2013.



Shaohua Li received the B.S. degree in information security from the Department of Information Security, University of Science and Technology of China (USTC), Hefei, China, in 2016. He is currently working toward the Graduate degree in communication and information system in the Department of Electronic Engineering and Information Science (EEIS), USTC.

His research interests include network security protocol design and analysis.

Kaiping Xue (M'09–SM'15) received the B.S. degree in information security from the Department of Information Security, University of Science and Technology of China (USTC), Hefei, China, in 2003 and the Ph.D. degree in information and communication engineering from the Department of Electronic Engineering and Information Science (EEIS), USTC, in 2007.

He is currently an Associate Professor in the Department of Information Security and Department of EEIS, USTC. His research interests include next-generation Internet, distributed networks and network security.

Qingyou Yang received the B.S. degree in information security from the School of the Gifted Young, University of Science and Technology of China (USTC), Hefei, China, in July 2016. He is currently a graduate student of Communication and Information System in the Department of Electronic Engineering and Information Science (EEIS), USTC.

His research interests include network security and cryptography.

Peilin Hong received the B.S. and M.S. degrees in electronic and information engineering from the Department of Electronic Engineering and Information Science (EEIS), University of Science and Technology of China (USTC), Hefei, China, in 1983 and 1986, respectively.

She is currently a Professor and an Advisor for Ph.D. candidates in the Department of EEIS, USTC. Her research interests include next-generation Internet, policy control, IP QoS, and information security. She has published 2 books and more than 150 academic papers in several journals and conference proceedings.