IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 31, 2020, accepted September 10, 2020, date of publication September 22, 2020, date of current version October 9, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3025788

An Efficient and Privacy Protection Communication Scheme for Smart Grid

ALBERT GUAN[1] **AND D. J. GUAN**[2,3]

[1]Department of Applied Mathematics, National Sun Yat-sen University, Kaohsiung 80424, Taiwan, R.O.C.
[2]Department of Computer Science, National Sun Yat-sen University, Kaohsiung 80424, Taiwan, R.O.C.
[3]Department of Applied Mathematics, National Chiao Tung University, Hsinchu 30010, Taiwan, R.O.C.

Corresponding author: Albert Guan (albertguan@math.nsysu.edu.tw)

This work was supported in part by the Ministry of Science and Technology under Grant 107-2218-E-110-017-MY3.

ABSTRACT Smart grid is a modern electric power grid designed to improve efficiency and reliability of the production and distribution of electric power. In a smart grid, smart meters continuously generate electric power consumption data and send it to the server. These data have two important purposes: (1) The sum of the data generated by each meter in a certain period of time will be used for billing; (2) The sum of the data generated by all meters in a specific area at a given time will be used to predict the electric power required in that area for electric power distribution. These data are considered to be sensitive and should be properly protected. There have been many studies on the confidentiality and privacy protection of these data. Some schemes require trusted servers, some schemes require heavy computation, and some schemes need to send two sets of data, one for billing and the other for electric power distribution. In this article, we propose an efficient and privacy-preserving communication scheme for the transmission of meter data in a smart grid. No trusted authorities are required in the scheme. By sending only one set of data, the new communication scheme can ensure that both sums for billing and sums for electric power distribution can be computed accurately. The scheme uses only simple operations, such as addition and hashes. It is computationally lightweight and suitable for devices with limited computing resources.

INDEX TERMS Smart grid, privacy protection, secret sharing, differential privacy.

I. INTRODUCTION

A smart grid is a modernized electric power grid designed to improve efficiency and reliability of the production and distribution of electricity. Smart grids usually include smart meters, smart appliances, renewable energy and other resources. These components are integrated into an advanced metering infrastructure for remote meter configuration, dynamic tariffs, electric power quality monitoring and load control.

Smart meters that measure electric power consumption of customers are essential devices in a smart grid. A smart meter can be considered as an electronic meter with a communication link. Smart meters must constantly send their measuring data to nearby server. The time to send these data is usually every 15 minutes, an hour, or a day [1]. The customer's electricity bill is based on these data. This information can also be used to predict the electrical energy required in a particular area to better distribute electricity to that area. Therefore, protecting the data generated and sent by each meter is essential to the security of a smart grid.

For confidentiality, encryption and authentication can be used to prevent an adversary from learning or changing the information sent through public networks. In a smart grid, in which the electric power usage of customers are constantly being sent, confidentiality may not be sufficient to protect the privacy of the customers. There is a crucial concern of the privacy related to the collection and use of customers' energy consumption data. Smart meters can be used by others either maliciously or inadvertently in an unauthorized fashion to infer types of activities or occupancy of a home for specific periods of time. It is also possible that such information can be sought for legal proceedings as evidence to prove or disprove certain propositions. To protect the privacy of the customers, NIST recommends using anonymous techniques to avoid traces of meter readings [2]. Unless the servers to which the data are sent and stored are fully trusted, additional steps must be taken to protect the privacy of the customers.

The associate editor coordinating the review of this manuscript and approving it for publication was Firooz B. Saghezchi.

Both confidentiality and privacy protection are very important in the information security of a smart grid. They are different problems and require different technologies to solve these two problems. In confidentiality, the sender and the receiver are usually trusted. Encryption can be used to avoid disclosure of sensitive or confidential information to third-party adversaries. For privacy protection, in addition to the third-party adversaries, some information sent by the sender should also be kept confidential to the receiver.

There have been many studies on the encryption and authentication of data sent by smart meters. For example, Liu *et al.* proposed a lightweight authentication scheme [3]; Wu *et al.* presented an improved version of Liu's scheme [4]; Mahmood *et al.* proposed another lightweight authentication scheme for smart meters [5]. However, privacy protection has not been integrated in these schemes.

In this article, we propose an efficient and lightweight privacy-preserving communication scheme for the transmission of data in a smart grid. The main techniques used in the design of the communication scheme are secret sharing and differential privacy.

For security, we must ensure that the data sent by smart meters are encrypted. Our privacy protection communication scheme adds carefully calibrated noises to the data measured by smart meters before sending them to the server. The addition of the noises to the data plays the role of one-time pad encryption. Note that the noise in our scheme is a random number generated from a distribution with mean 0 and carefully chosen variance. It is not the noise from the environment in the communication channel.

For the authentication of the data sent by the meters to the server, a lightweight authentication method, such as Liu *et al.*'s method [3], can be used. In our scheme, the data to be authenticated is the one-time encrypted data, not the original meter readings, which need to be kept secret to the server. The authentication part of the scheme will not be described in detail in this article.

We assume that the servers are semi-honest. They follow the protocol, but may want to know the customers' timely electric power consumption information. We also assume that nearby smart meters can communicate with each other. By sending only one set of data, our communication scheme ensures that both sums for billing and sums for electric power distribution can be computed accurately. No trusted authorities are required in our scheme, and the scheme uses only simple operations, such as addition and hashes. It is a lightweight scheme suitable for devices with limited computing resources.

II. DESCRIPTION OF THE PROBLEM

A smart grid may contain many components. Figure 1 shows a simplified diagram of the smart grid, which contains only the components related to our communication scheme. Each smart meter constantly generates electric power consumption data and sends to the nearby server. The electricity company uses these data for two purposes: billing and electricity distribution.

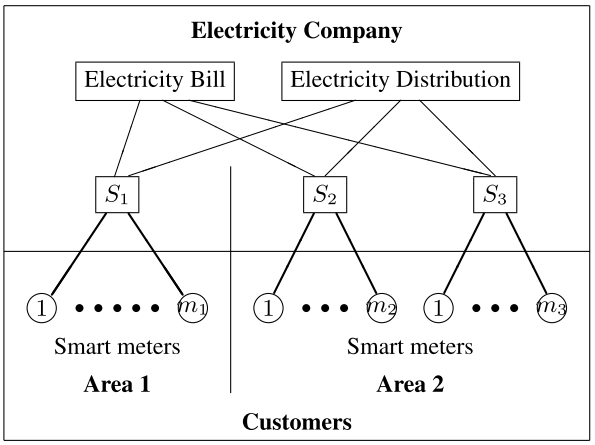

FIGURE 1. Simplified diagram of smart grid, which contains only the components related to our communication scheme.

Our goal is to protect the data sent by smart meters, from the malicious third party for secrecy, and from the electrical company for customer privacy. For simplicity, assume that each customer has a smart meter at home. Assumed that each smart meter transmits electric consumption data to the server every τ time units. According to the guidelines of European Regulators Group for Electricity and Gas, the value of τ is usually 15 minutes, an hour, or a day [1]. This data represents the amount of electric power consumption during time interval $[t-\tau, t]$.

Table 1 shows the data generated by $m = 6$ meters in $n = 8$ time intervals. In Table 1, $d_{i,j}$ is the electric power consumption measured by smart meter i during the time interval $[t_{j-1}, t_j]$. Our privacy protection communication scheme will add carefully calibrated noise to the data before sending them to the servers.

TABLE 1. An example of data generated by smart meters for $m = 6$ meters and $n = 8$ time intervals.

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
1	$d_{1,1}$	$d_{1,2}$	$d_{1,3}$	$d_{1,4}$	$d_{1,5}$	$d_{1,6}$	$d_{1,7}$	$d_{1,8}$
2	$d_{2,1}$	$d_{2,2}$	$d_{2,3}$	$d_{2,4}$	$d_{2,5}$	$d_{2,6}$	$d_{2,7}$	$d_{2,8}$
3	$d_{3,1}$	$d_{3,2}$	$d_{3,3}$	$d_{3,4}$	$d_{3,5}$	$d_{3,6}$	$d_{3,7}$	$d_{3,8}$
4	$d_{4,1}$	$d_{4,2}$	$d_{4,3}$	$d_{4,4}$	$d_{4,5}$	$d_{4,6}$	$d_{4,7}$	$d_{4,8}$
5	$d_{5,1}$	$d_{5,2}$	$d_{5,3}$	$d_{5,4}$	$d_{5,5}$	$d_{5,6}$	$d_{5,7}$	$d_{5,8}$
6	$d_{6,1}$	$d_{6,2}$	$d_{6,3}$	$d_{6,4}$	$d_{6,5}$	$d_{6,6}$	$d_{6,7}$	$d_{6,8}$

In general, the data generated by smart meters are used for three purposes. (1) billing, (2) electric power distribution, and (3) value-added services. These three types of uses of the data differs significantly in terms of their requirement on metering frequency and accuracy.

The primary use of the data generated by the meters is consumer billing. Since billing typically happens on monthly basis, the electric power consumption data needs not be processed in real-time, but the correctness of billing requires accurate measurement data.

Another important use of data generated by smart meters is to increase the efficiency and reliability of electric power distribution of the smart grid. The inclusions of renewable energy to smart grid makes electric power distribution even more important. The data generated by smart meters at specific time are mainly used for the prediction of electric power required in each area for the next time period. This type of data usage requires real-time or near real-time processing and fine-grained time intervals, but may accept lower accuracy.

The measuring data can also be used by customers, operators, and third-party service providers for providing various value-added services, such as managing and arranging the use of household appliances to reduce electricity bills.

In this article, we focus on the first two uses of metering data, namely billing and electric power distribution. Assume that the power company summarize the data in every n time intervals. For the description of the privacy protection communication scheme, the table will be shown in m rows and n columns. The sum of the i-th row is the electric power used by customer i from time t_0 to t_n. This sum is also called temporal aggregation. Each of the sum will be used by the power company to calculate the electricity bill for the customer.

The sum of the j-th column is the electric power used by all customers in this area at time t_j. This sum is also called spatial aggregation. The power company will use this sum to predict the amount of electricity needed in the area and allocate enough electricity for the area to meet customer needs.

Suppose the electricity company calculates the electricity usage for each customer every n time intervals, and customers receive their electricity bill in every $N = kn$ time intervals for some integer k. For example, assume that each meter sends out its electric power consumption data every 15 minutes, the electricity company calculates the electricity usage for each customer every day, and the electricity bill will be computed monthly, then $\tau = 15, n = 96, k = 30$, the number of data sent to the server will be $N = 30 \times 96 = 2880$ for each meter in 30 days.

We assume that these data are stored in a server operated by electricity company. Our goal is to design a robust and efficient communication scheme to transfer these data from smart meters to the server to ensure the security of data, and most importantly, to protect the privacy of customers. There may be different values of τ, k, or N for different countries, our communication scheme works correctly, as long as N is a multiple of n.

III. RELATED WORKS

In this article, we focus on the privacy protection of customer power consumption data. There are other situations where the privacy of customers needs to be protected. For example, Zhang *et al.* studied the privacy-preserving communication and power injection scheme over vehicle networks and 5G smart grid slice [6].

Many techniques have been studied for the protection of the privacy of customers in smart grids. Rechargeable batteries can be installed to hide the energy consumption of customers [7]. This type of technique require the installation of hardware. In this article, we focus on software technologies that can be applied to protect customer privacy.

Anonymous technology can also be used to protect customers' privacy in smart grid. Petrlic proposed a privacy protection scheme using a trusted intermediate gateway as a pseudonym server for billing applications [8]. This technique can hide the real identity of customers, but it requires a trusted pseudonym server.

Homomorphic encryption can do arithmetic operations directly on ciphertext without decryption. This technique can be used to protect customer privacy. Jawurek *et al.* proposed a secure computation of billing using homomorphic commitment [9]. Metering data are committed and aggregated first. Only the final sum will be opened to electricity company, and the correctness of the data can be proved by using zero-knowledge proof. Kong *et al.* proposed a group blind signature scheme in smart grid to accomplish conditional anonymity [10]. The integrity of electricity consumption data can be verified by homomorphic encryption. The problem with this type of scheme is that homomorphic encryption is usually computationally inefficient, especially for homomorphic encryption with addition and multiplication operations.

Lin *et al.* proposed a smart metering system supporting both privacy preserving billing and load monitoring with one set of data [11]. In their system, meter readings are stored in a semi-trusted storage system. The electricity company can only query for the sum of meter readings over a time period. The load monitoring unit can only query the sum of meter readings from meters in the area at a specific time. In this scheme, the storage system stores all the original data generated by smart meters. The correct operation of this method depends on the trusted storage system. Due to the intentional or unintentional behavior of the storage system, or the intrusion of attackers, sensitive data may be leaked to a malicious party.

Differential privacy was originally designed for statistical data set to limit the disclosure of private information. It is also useful in protecting customers' privacy in smart grid. Hale *et al.* applied differential privacy to the metering data both for billings and electric power distribution [12]. They showed that, with proper selection of parameters, both the billing and the electric power distribution aggregations may have some errors, but these data are still useful. Eibl and Engel studied the effect of differential privacy on real smart metering data, and showed that as long as the number of smart meters is large enough, the data are useful [13].

In summary, many technologies have been used to solve the confidentiality and privacy protection of customers in the smart grid. Some schemes require heavy computation, such as homomorphic encryption. Some schemes require the use of trusted servers. Some schemes require two sets of data, one set for spatial sums and the other for temporal sums. Some schemes can only provide a good approximation of the spatial sums and the temporal sums.

In this article, we propose a novel communication scheme for smart grids to achieve secrecy and, at the same time, to preserve the privacy of customers. In our scheme, carefully calibrated noise are added to the data before sending the data to the server. No measuring data from any smart meter are directly transmitted and stored in a storage system that the electricity company can access. Thus, the privacy protection of customers can be achieve perfectly. Our communication scheme can always ensure that the sum of each meter readings during a given period of time will be exact. Therefore, billings for the customers will always be accurate. Furthermore, in our scheme, the same set of data can be used for both billing and electric power distribution.

IV. PRELIMINARIES

In this section, we briefly introduce secret sharing scheme and differential privacy. The modified versions of the two schemes will be used in our privacy protection communication scheme.

A. SECRET SHARING

Let t and n be two positive integers, $t \leq n$. A (t, n)-threshold secret sharing scheme is a method for the n users to share a secret K. Each user i has a share s_i about the secret K. The goal of a secret sharing scheme is that the secret K can be computed correctly by using the shares of any subset of t users, while any $t-1$ or fewer users cannot compute any information about the secret K.

Shamir showed that a (t, n)-threshold secret sharing scheme can be implemented by polynomial interpolation [14]. The secret K, as well as each share, is represented by a point in a polynomial of degree $t-1$. Any t shares can uniquely determine the polynomial, but any subset of $t-1$ or fewer shares cannot.

Our privacy protection communication scheme uses a special case of (t, n)-threshold secret sharing scheme, namely the (n, n)-threshold secret sharing scheme. It can be implemented much more efficiently without polynomial interpolation.

Let p be an integer greater than the secret K. The first $n-1$ shares s_i, $i = 1, 2, \ldots, n-1$, can be randomly and uniformly selected from the set $\{0, 1, \ldots, p-1\}$. Then the last share s_n is computed by

$$s_n = [K - (s_1 + s_2 + \cdots + s_{n-1})] \bmod p.$$

It can be verified that

1) The sum of all shares $\left(\sum_{i=1}^{n} s_i\right) \bmod p$ is equal to K.
2) Any sum of the proper subset of the shares $\{s_1, s_2, \ldots s_n\}$, is a random number.

This implementation of the (n, n)-threshold secret sharing scheme is *perfect*, which means that no subset of $n-1$ or fewer users can compute any information about the secret K even if they have infinite computing power. This implementation of the (n, n)-threshold secret sharing scheme is also *ideal*, because the size of each share s_i is no more than the size of the secret K, that is, $|s_i| = |K|$. We will modify the above perfect and ideal secret sharing scheme to provide secrecy and protect the privacy of the customers.

B. DIFFERENTIAL PRIVACY

Differential privacy was originally design for statistical data set. It has been shown that an attacker can understand the confidential content of a statistical data set by creating a series of target queries. In 2003, Nissim and Dinur demonstrated that ''it is impossible to publish arbitrary queries on a private statistical data set without revealing some amount of private information.'' This is also called fundamental law of information retrieval.

Noise can be added to each query to limit the leakage of privacy in the data set. In 2006, Dwork *et al.* presented a method called ϵ-differential privacy, to formalize the amount of noise that needed to be added and proposed a generalized mechanism for adding the noise [15].

The intuition of ϵ-differential privacy is that a person's privacy cannot be compromised by releasing statistical information if their data are not in the data set. Therefore, with differential privacy, the goal is to give each individual roughly the same privacy that would result from having their data removed.

Dwork and Roth formally defined ϵ-differential privacy as follows [15]. Let ϵ be a positive real number. Let $\mathcal{A}$ be a randomized algorithm that takes a data set as input and compute an output representing the actions of the trusted party holding the data in response to a query. The algorithm $\mathcal{A}$ is said to provide ϵ-differential privacy if, for all data sets D_1 and D_2 that differ on a single element (i. e., the data of one person), and all subsets S of all possible responses of $\mathcal{A}$:

$$\Pr[\mathcal{A}(D_1) \in S] \leq e^{\epsilon} \Pr[\mathcal{A}(D_2) \in S].$$

In other words, a data set query response algorithm $\mathcal{A}$ is ϵ-differential if for all data set D_1 and D_2 differing on a single record, the probability of obtaining response $t \in S$ when the data set is D_1 is within $(1+\epsilon)$ times the probability of obtaining response $t' \in S$ when the data set is D_2. This also implies that the ratio of the two probabilities lies in $[e^{-\epsilon}, e^{\epsilon}] \approx [1-\epsilon, 1+\epsilon]$ when ϵ is small.

In our communication scheme for privacy protection, the data to be protected are the electric power consumption data for each smart meter in the grid. This is a sequence of $d_{i,j}$, $j = 1, 2, \ldots, n$. To ensure that the eavesdropper cannot learn any useful information about $d_{i,j}$, certain amount of noise must be added to it before it is transmitted to the server. The proper amount of noise can be determined by the principle of differential privacy.

In differential privacy, the amount of noise to be added to the real data is a trade-off between privacy protection and data usefulness. The smaller ϵ, the better protection of customer privacy. On the other hand, the larger the ϵ, the more accurate the data.

According to the principle of differential privacy, the amount of noise to be added to the data should be proportional to the sensitivity of the query function. To estimate the

sensitivity of our query function, let d be a positive integer, $\mathcal{D}$ be a collection of data sets, and $f : \mathcal{D} \to \mathbf{R}^d$ be a query function. The sensitivity of the function f, denoted Δf, is defined by

$$\Delta f = \max ||f(D_1) - f(D_2)||_1,$$

where the maximum is over all pairs of data sets D_1 and D_2 in $\mathcal{D}$ differing in at most one element, and $||\cdot||_1$ denotes the ℓ_1-norm. In our scheme, the query function is the row sum or the column sum of the data $d_{i,j}$. Therefore, the sensitivity of our query function is $\max\{d_{i,j}\}$.

There are many mechanisms which can provide ϵ-differential privacy, such as Laplace mechanism, exponential mechanism, and posterior sampling. The Laplace mechanism adds noise from the Laplace distribution, which can be expressed by the probability density function

$$\text{noise}(y) \propto \exp(-|y|/\lambda)$$

which has mean 0 and standard deviation $\sqrt{2}\lambda$. It can be shown that this method satisfies the definition of ϵ-differential privacy. We use this mechanism in our privacy protection communication scheme.

Note that our privacy protection communication scheme is not a direct application of differential privacy. We also integrate (n, n)-threshold secret sharing scheme to ensure that both the temporal and the spatial aggregation of the data can be computed accurately. Only in the case that certain meter cannot communicate with any other meters, the method of differential privacy is used to protect the privacy of the data generated by that meter. Even if all meters cannot communicate with other meters, the temporal sum for billing can still be accurately computed in our scheme.

V. THE PRIVACY PROTECTION SCHEME

In this section, we propose a communication scheme for the transmission of data generated by each smart meter that meets our goal: both temporal aggregation for billing and spatial aggregation for electric power distribution can be computed accurately by using only one set of data. Furthermore, the confidentiality and the privacy of customers are properly protected. We assume that smart meters in a specific area can communicate with some other smart meters.

A. MODIFICATION OF THE SECRET SHARING SCHEME

The main technique used in the design of the communication scheme is a modification of the (n, n)-threshold secret sharing scheme. For each meter i, the data that should be protected are $d_{i,j}, j = 1, 2, \ldots, n$. Our communication scheme adds certain amount of noise to $d_{i,j}$ before sending it to the server.

We first show that direct application of the (n, n)-threshold scheme to our communication scheme may have difficulty. Suppose that the (n, n)-threshold secret sharing scheme is modified to share a sequence of numbers instead of only one key. Let p be an integer greater than $\sum_{j=1}^{n} d_{i,j}$. Each meter i first selects $n-1$ random numbers $s_{i,j}, j = 1, 2, \ldots, n-1$ from $\{0, 1, \ldots, p-1\}$. The last random number s_n can then be computed by

$$s_n = -\left(\sum_{j=1}^{n-1} s_{i,j}\right) \bmod p.$$

Then, the j-th data to be sent to the server for meter i is

$$w_{i,j} = d_{i,j} + s_{i,j}, \quad j = 1, 2, \ldots, n.$$

It is easy to verify that

$$\left(\sum_{j=1}^{n} w_{i,j}\right) \bmod p = \left(\sum_{j=1}^{n} d_{i,j} + \sum_{t=1}^{n} s_{i,j}\right) \bmod p = \sum_{j=1}^{n} d_{i,j}.$$

The above method allows the electricity company to compute the correct sum of the data sent from each meter in a fixed time period from t_0 to t_n without knowing each $d_{i,j}$. However, in order to correctly compute the sum, the value of p must be greater than $\sum_{j=1}^{n} d_{i,j}$. If the value of p is too small, i. e. $p < \sum_{j=1}^{n} d_{i,j}$, then the sum $\sum_{j=1}^{n} d_{i,j}$ would not be correctly computed. To avoid having to estimate the correct value of p, the (n, n)-threshold secret sharing scheme must be further adapted so that it can be applied to the communication of smart meters.

We observed that the random value $s_{i,j}$ added to the data in secret sharing plays the same role as the random noise added to the data in differential privacy. Therefore, we can randomly select $n-1$ noises $s_{i,j}$ to be added to $d_{i,j}$. To ensure that the sum of each row is correct, the last noise $s_{i,n}$ must be computed from the first $n-1$ noises:

$$s_{i,n} = -\sum_{j=1}^{n-1} s_{i,j}.$$

According to the principle of differential privacy, the noise $s_{i,j}$ to be added to the data $d_{i,j}$ should be randomly selected from a probability distribution with mean 0, and variance proportional to the sensitivity of the query function. In our privacy protection communication scheme the query function is the sum $\sum_{j=1}^{n} d_{i,j}$. Thus, the sensitivity is

$$D_i = \max_{j=1}^{n}\{d_{i,j}\}.$$

The value of D_i for each meter i also require to be determined ahead of the time. However, a good estimation of D_i is sufficient to make the scheme works correctly. It can be verified that, even if some of the values of $d_{i,j}$ exceeds D_i, the desired sum $\sum_{j=1}^{n} d_{i,j}$ can still be computed correctly.

$$\sum_{j=1}^{n} w_{i,j} = \sum_{j=1}^{n}(d_{i,j} + s_{i,j}) = \sum_{j=1}^{n} d_{i,j} + \sum_{j=1}^{n} s_{i,j} = \sum_{j=1}^{n} d_{i,j}.$$

In other words, the sum $\sum_{j=1}^{n} d_{i,j}$ can be computed correctly as long as $\sum_{j=1}^{n} s_{i,j} = 0$. This equation holds due to the application of the (n, n)-threshold scheme.

Our ultimate goal is to compute both temporal sum and spatial sum correctly. Unlike differential privacy, usability is no longer a problem in our scheme because the sum can always be computed accurately in our scheme. The value of each $s_{i,j}$, except $s_{i,n}$, can be chosen randomly and uniformly in the interval $[-D_i/2, D_i/2]$. In fact, choosing $s_{i,j}$ randomly and uniformly in $[-D_i/2, D_i/2]$ increase the uncertainty (information entropy) of the data sent to the server. For example, the entropy of $d_{i,j}$ given $w_{i,j} = d_{i,j} + s_{i,j}$ is $\log(D_i)$, if $s_{i,j}$ is randomly and uniformly chosen from $[-D_i/2, D_i/2]$. If $s_{i,j}$ is chosen from Laplace distribution with mean 0 and variance D_i, the entropy of $d_{i,j}$ is only $\log(2be)$, where the variance $2b^2 = D_i$. The entropy $\log(2be) = \log\left(\sqrt{2D_i}\, e\right) < \log(D_i)$ whenever $D_i > 2e^2$.

The security of the above method is the same as one-time pad. Since the noises $s_{i,j}$, $j = 1, 2, \ldots, n$, are randomly chosen, it is impossible to compute the value of individual $d_{i,j}$ from the cipher text $w_{i,j}$, unless the values of these random noises $s_{i,j}$ are known.

By using the above method, it is straightforward to encrypt the data $d_{i,j}$ for temporal aggregation for billing. Note that, for each meter, only the first $n-1$ random noises can be randomly chosen. The last noise must be computed from the first $n-1$ noises. Therefore, if the above method is applied to the column sum or spatial sum, smart meter needs to communicate with each other to synchronize their random noises.

B. DESCRIPTION OF THE PRIVACY PRESERVING SCHEME

In our communication scheme, every smart meter in a specific area should be able to communicate with some other smart meters in that area. Smart meters can communicate by power-line network or any other network. If power-line network is used, then data are sent and received on a conductor that is also used simultaneously for electric power transmission. The privacy protection communication scheme is summarized in Figure 2.

The amount of noise $s_{i,j}$ to be added to the data $d_{i,j}$ is synchronized in the communication between meters. In the time interval $[t_0, t_n]$, n random noises $s_{i,j}$ are required for each meter i. They can set up their first $n-1$ random noises by the method described in step 1 of scheme, and compute the last one based on the first $n-1$ random noises.

To reduce the number of communications, they can also set up their first random noise, and compute the other $n-2$ noises by the following method. Suppose that meter i sends a request to meter k for setting up a random noise $s^k_{i,1} = \sigma_i r$ in step 1(c) of the scheme as shown in Figure 2. The random noises $s^k_{i,j}, j = 2, 3, \ldots, n-1$, can be computed by

$$s^k_{i,j} = (-1)^{j-1}\sigma_i h(r, j),$$

1) Each meter i tries to set up a random number with some other meters in the same area for generating noises to be added to their data $d_{i,j}$. Assume that meter i send a request to meter k to set up $s^k_{i,j}$ and $s^i_{k,j}$ to be added to $d_{i,j}$ and $d_{k,j}$, respectively. They perform the following steps.
   a) Meter i randomly and uniformly chooses a number r_i from $[-D_i/2, D_i/2]$, and sends r_i to meter k, where D_i is the estimation of $\max\{d_{i,j} \mid j = 1, 2, \ldots, n\}$.
   b) Meter k randomly and uniformly chooses a number r_k from $[-D_k/2, D_k/2]$, and sends r_k to meter i, where D_k is the estimation of $\max\{d_{k,j} \mid j = 1, 2, \ldots, n\}$.
   c) Let $r = \max\{r_i, r_k\}$, and let $\sigma_i = 1$ if $r_i > r_k$, otherwise let $\sigma_i = -1$. Let $\sigma_k = -\sigma_i$. Meter i sets $s^k_{i,j} = \sigma_i r$. Meter k sets $s^i_{k,j} = \sigma_k r$.
2) For each meter i let $s^i_{i,j} = 0$ and $s^k_{i,j} = 0$ for every k if meter i did not set up $s^k_{i,j}$ with meter k. At time $j = 1, 2, \ldots, n-1$, each meter i computes the random noises,
$$s_{i,j} = \sum_{k=1}^{m} s^k_{i,j}.$$
For $j = n$, each meter i computes $s_{i,n}$ by
$$s_{i,n} = -\sum_{j=1}^{n-1}\sum_{k=1}^{m} s^k_{i,j}.$$
3) At time $j = 1, 2 \ldots, n$, each meter i sends $w_{i,j} = d_{i,j} + s_{i,j}$ to the server.

FIGURE 2. Description of the privacy protection communication scheme.

for meter i. Summarily, meter k can compute

$$s^i_{k,j} = (-1)^{j-1}\sigma_j h(r, j).$$

In the above equations, h is a secure one-way hash function. The output of h is the range $[-D, D]$, where $D = \max\{D_i, D_j\}$.

The actual electricity consumption data of the customers are not directly stored in the server. Therefore, our communication scheme is secure and privacy-preserving, even if sensitive data may be leaked to a malicious party due to the intentional or unintentional behavior of the storage system or the intrusion of an attacker.

VI. ANALYSIS OF PRIVACY PROTECTION SCHEME

In this section, we show that our communication scheme is secure and privacy-preserving. We first give a formal security model of the security and privacy protection communication scheme.

Let $d_{i,j}$ be the electric power consumption data measured by smart meter i during time period $[t_{i-1}, t_i]$. For the security

of these $d_{i,j}$'s, we adopt the normal definition that the unauthorized parties cannot learn any information about the value of $d_{i,j}$. No trusted servers are required in our scheme. Therefore, we assume that the servers are semi-honest and define privacy-preserving as follows.

definition 1: The communication between smart meters and servers in a smart grid is privacy-preserving *if the semi-honest servers can only compute the temporal sums* $\sum_{j=1}^{m} d_{i,j}$ *and the spatial sums* $\sum_{i=1}^{n} d_{i,j}$*, but the servers do not know the value of each* $d_{i,j}$.

We now show that the privacy protection communication scheme shown in Figure 2 can always accurately compute the temporal sum for billing.

Theorem 1: The temporal aggregation for billing of each meter i can be computed accurately by computing $\sum_{j=1}^{n} w_{i,j}$.

Proof:

$$\sum_{j=1}^{n} w_{i,j} = \sum_{j=1}^{n} d_{i,j} + \sum_{j=1}^{n} s_{i,j} = \sum_{j=1}^{n} d_{i,j}.$$

This is because the term, $\sum_{j=1}^{n} s_{i,j} = 0$ for the (n, n)-threshold scheme. □

Next, we show that the spatial aggregation for electric power distribution at time t can also be computed accurately.

Theorem 2: Assume that each meter is communicated with some other meters in the same area for setting up random noises to be added to its data. The spatial aggregation for electric power distribution can be computed accurately by $\sum_{i=1}^{m} w_{i,j}$.

Proof:

$$\sum_{i=1}^{m} w_{i,j} = \sum_{i=1}^{m} d_{i,j} + \sum_{i=1}^{m} s_{i,j} = \sum_{i=1}^{m} d_{i,j}.$$

In the above equation, the term $\sum_{i=1}^{m} s_{i,j} = 0$, because it includes both $s_{i,j}^{k}$ and $s_{k,j}^{i}$, one is positive and the one is negative, for every pair of meters i and k in that area. □

To show that the privacy of all customers can be protected, we model the communication pattern of smart meters by a graph $G = (V, E)$. The vertex set V is the set of the m meters in that area, Let $V = \{1, 2, \ldots, m\}$. There is an edge between i and k if, and only if, meter i and meter k communicate with each other to establish random noises to be added to their data. The graph $G = (V, E)$ is called the *connection graph* for smart meters in the area.

Let G be the connection graph for some area in a smart grid, and S be a subset of vertices in G. Let $[S, \bar{S}]$ denote the set of edges with one endpoint in S and the other endpoint not in S. Define $D(S, j)$ to be the sum of measuring data sent from all meters in area S at time $j = 1, 2, \ldots, n$, that is,

$$D(S, j) = \sum_{i \in S} w_{i,j}.$$

For the proof of the privacy protection property of our smart grid communication scheme, we first prove the following theorem.

Theorem 3: Let $G = (V, E)$ *be the connection graph for some area in a smart grid, and* S *be a subset of* V*. The value of* $D(S, j)$ *can be computed accurately by the server, if and only if* $[S, \bar{S}] = \emptyset$.

Proof: Consider the spatial sum with respect to S at time j.

$$\sum_{i \in S} w_{i,j} = \sum_{i \in S} d_{i,j} + \sum_{i \in S} s_{i,j}.$$

The last term, $\sum_{i \in S} s_{i,j} = 0$ if and only if $[S, \bar{S}] = \emptyset$. In other words, if $[S, \bar{S}] \neq \emptyset$, then its value is the sum of random numbers which are totally unknown to the server. This implies that the server cannot compute the value of $D(S, t)$. □

Based on the above theorem, we have the following corollary, which gives another proof that, as long as each meter is connected with some other meters in this area, the spatial sum for electric power distribution can be computed accurately,

Corollary 3.1: The spatial sum for an area can be computed accurately by the server if and and only if no meters in this area is connected with meters in another areas.

Finally, we show that our privacy protection communication scheme preserves the privacy of all customers if every meter can communicate with some other meters in the same area.

Corollary 3.2: The value of $d_{i,j}$, $j = 1, 2, \ldots, n$, *cannot be computed by the server, if and only if, meter i is connected to some other meters.*

According to Corollary 3.2, to protect the privacy of a customer, every meter should be connected to some other meters in the same area to set up random noise to be added to its data before sending it to the server. In theory, this is sufficient to protect the measuring data for every meter. For example, if the connected component contains only two meters, then the sum of the electric power usages of the two meters can be computed, but the meter reading for each meter remains secret. In practice, we may want to avoid small connected component in the connection graph.

There are many ways to make sure that every meter is connected in the connection graph. For example, the connection graph can be an l-circulent graph, where l is a small integer. In this graph, a pair of meters i and k are connected if $i - k \equiv \beta \pmod{m}$ for some $\beta \in \{\alpha_1, \alpha_2, \ldots, \alpha_l\}$, where $\alpha_1, \alpha_2, \ldots, \alpha_l$ are l positive integers, with $\gcd(m, \alpha_1, \alpha_2, \ldots, \alpha_l) = 1$. In particular, when $l = 2$, the graph is also called a double-loop network. In this network, each meter connects to 4 other vertices $i \pm \alpha_1$ and $i \pm \alpha_2$.

The connection graph can also be a random graph. In this case, every meter i first sets a probability p. Then it tries to send a request to set up random noise with other meter k with probability p. The following theorem shows that, with proper value of p, the graph G will almost sure be a connected graph [16].

Theorem 4 Alon and Spencer: Let ϵ be a positive number, and p be the probability that meters i and k establish random noises to be added to their data. If $p > ((1+\varepsilon)\ln n)/n$, then the connection graph $G(V, E)$ will almost surely be connected.

In the case that some meter i cannot set up any random number with other meters to generate noise to protect its data, each such meter i can choose a number form the Laplace distribution with mean 0 and variance D_i. Even in this case, our scheme can still ensure that the temporal sum for billing is accurate, only the spatial sum may induce some errors. In the extreme case, every meter cannot communicate with other meters. This is equivalent to the case $p = 0$. When this happened, for only the temporal sum, our communication scheme degraded to Hale *et al.*'s scheme [12]. They showed that the spatial sum is still useful for electric power distribution, as long as the number of meters n in that area is large. Note that, in our communication scheme, even in this extreme case, the temporal aggregation for billing can still be accurately computed.

VII. CONCLUSION AND DISCUSSION

We have presented a communication scheme for smart meters in a smart grid to send their measuring data to the server in a secure and privacy-preserving way. In our scheme, only one set of measuring data is required to be sent to and stored in the server. The same set of data can be used for computing the temporal sum which is used for billing, and the spatial sum which is used for electric power distribution. Smart meters need to communicate with other meters to generate proper amounts of noise to be added to the measuring data. When a meter cannot communicate with other meters, it needs to generate a random noise from a probability distribution with mean 0. We have shown that, even if all meters cannot communicate with each other, the temporal sum for billing can still be computed accurately. The spatial sum used for electric power distribution may have some errors, but still useful for electric power distribution.

The main techniques used in our communication scheme are secret sharing and differential privacy. Both these techniques require only simple computations. The (n, n)-secret sharing scheme used in our communication scheme is similar to one-time pad encryption. Only addition is required to do encryption, no modular exponentiation or other heavy computations. Hashes may be required to compute some random noises to reduce communications between smart meters. Thus, addition and hashes are the only computations required by our scheme. Therefore, our scheme is a lightweight scheme. It is more suitable for devices with low computing resources.

REFERENCES

[1] *Final Guidelines of Good Practice on Regulatory Aspect of Smart Metering for Electricity and Gas*, document dRAFT-NISTIR-7628, European Regulators Group for Electricity and Gas, Feb. 2011.

[2] *Smart Grid Cyber Security Strategy and Requirements*, document dRAFT-NISTIR-7628, USA Interagency Rep., National Institude of Standards and Technology, 2009.

[3] Y. Liu, C. Cheng, T. Gu, T. Jiang, and X. Li, "A lightweight authenticated communication scheme for smart grid," *IEEE Sensors J.*, vol. 16, no. 3, pp. 836–842, Feb. 2016.

[4] L. Wu, J. Wang, S. Zeadally, and D. He, "Anonymous and efficient message authentication scheme for smart grid," *Secur. Commun. Netw.*, vol. 2019, p. 12, May 2019, doi: 10.1155/2019/4836016.

[5] K. Mahmood, S. Ashraf Chaudhry, H. Naqvi, T. Shon, and H. Farooq Ahmad, "A lightweight message authentication scheme for smart grid communications in power sector," *Comput. Electr. Eng.*, vol. 52, pp. 114–124, May 2016.

[6] Y. Zhang, J. Li, D. Zheng, P. Li, and Y. Tian, "Privacy-preserving communication and power injection over vehicle networks and 5G smart grid slice," *J. Netw. Comput. Appl.*, vol. 122, pp. 50–60, Nov. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1084804518302480

[7] D. Varodayan and A. Khisti, "Smart meter privacy using a rechargeable battery: Minimizing the rate of information leakage," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 1932–1935.

[8] R. Petrlic, "A privacy preserving concept for smart grids," in *Proc. Sicherheit Vernetzten Syst. 18. DFN*. Books on Demand GmbH, 2010, pp. B1–B14.

[9] M. Jawurek, M. Johns, and F. Kerschbaum, "Plug-in privacy for smart metering billing," in *Privacy Enhancing Technologies*, S. Fischer-Hübner and N. Hopper, Eds. Berlin, Germany: Springer, 2011, pp. 192–210.

[10] W. Kong, J. Shen, P. Vijayakumar, Y. Cho, and V. Chang, "A practical group blind signature scheme for privacy protection in smart grid," *J. Parallel Distrib. Comput.*, vol. 136, pp. 29–39, Feb. 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0743731519301285

[11] 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 *Applied Cryptography and Network Security*, F. Bao, P. Samarati, and J. Zhou, Eds. Berlin, Germany: Springer, 2012, pp. 544–560.

[12] M. Hale, P. Barooah, K. Parker, and K. Yazdani, "Differentially private smart metering: Implementation, analytics, and billing," 2019, *arXiv:1902.06310*. [Online]. Available: http://arxiv.org/abs/1902.06310

[13] G. Eibl and D. Engel, "Differential privacy for real smart metering data," *Comput. Sci. Res. Develop.*, vol. 32, nos. 1–2, pp. 173–182, Mar. 2017.

[14] A. Shamir, "How to share a secret," *Commun. ACM*, vol. 22, no. 11, pp. 612–613, Nov. 1979, doi: 10.1145/359168.359176.

[15] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, nos. 3–4, pp. 211–407, 2014.

[16] N. Alon and J. H. Spencer, *The Probabilistic Method*, 4th ed. Hoboken, NJ, USA: Wiley, Jan. 2016.

ALBERT GUAN received the bachelor's degree in applied mathematics from National Sun Yat-sen University, in 2008, and the Ph.D. degree in computer science from National Chiao Tung University, in 2017. From 2017 to 2018, he was a Postdoctoral Research Fellow with the Department of Electrical Engineering, National Taiwan University. In 2018, he joined the Department of Applied Mathematics, National Sun Yat-sen University, as an Assistant Professor. His research interests include discrete mathematics, cryptography, and its applications.

D. J. GUAN received the Ph.D. degree in computer science from Purdue University, in 1989. He was with the Department of Applied Mathematics, National Sun Yat-sen University, from August 1989 to January 1999. He moved to the Department of Computer Science, in February 1999. He has retired from Sun Yat-sen University. He is currently working as an Adjunct Professor with National Sun Yat-sen University and National Chiao Tung University. His research interests include algorithms, cryptography, combinatorics, and information security.