A Privacy-preserving Multi-dimensional Data Aggregation Scheme with Forward Security in Smart Grid

Guojun Wang[1], Xueya Xia[2], Sai Ji[2], Chin-Feng Lai[3]

[1] School of Information & Security, Yancheng Polytechnic College, China
[2] School of Computer & Software, Nanjing University of Information Science & Technology, China
[3] Department of Engineering Science, National Cheng Kung University, Taiwan
113097096@qq.com, xiaxueya5258@163.com, jisai@nuist.edu.cn, cinfon@ieee.org

Abstract

Smart grid is a new generation of power system on the strength of rapid-speed bidirectional communication network. In smart grid, the electricity data from users needs to be collected to realize efficient energy management, which may reveal their privacy. In the past period of time, data aggregation protocols have been extensively studied to solve this problem. However, these protocols treat users' electricity data as single-dimensional data, which is inconsistent with reality. Moreover, key leakage is also a serious security threat in the smart grid. Therefore, we propose a multi-dimensional data aggregation protocol with forward security. In this protocol, a cube data structure is designed to represent the multi-dimensional electricity consumption data of users in a region over multiple time periods, realizing the dual aggregation of electricity data both in time and space. In addition, security analysis shows that the proposed protocol satisfies multiple security properties and performance analysis shows that our protocol is more superior in the smart grid environment.

Keywords: Smart grid, Multi-dimensional data aggregation, Cube data structure, Forward security

1 Introduction

With the rapid development of ubiquitous sensing and 5G mobile communication technologies, smart grid has been widely deployed in recent years. The mobile service application of power grid system based on wireless communication technology becomes the main application of information and communication technology in smart power grid due to its features of providing network access function anytime and anywhere without complex wiring. As a large number of smart meters and smart appliances are installed in residential areas, the network edge is further extended to the user end [1]. The framework of smart grid is shown in Figure 1. Power flow, information flow and service flow being highly integrated has become the principal feature in smart grid [2]. As bidirectional communications are deeply involved in smart grid, the security risks on the user side will become more and more prominent. For example, the user's real-time electricity consumption data may be stolen by the attacker during the communication process, the privacy information such as the user's living habits may be leaked, and the user's safety is endangered [3]. Therefore, Data confidentiality, especially privacy protection of users, has become an issue that must be considered [4].

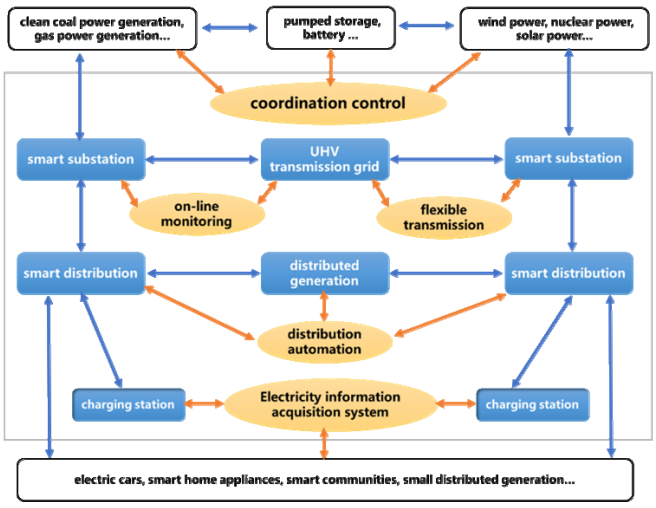

Figure 1. The conceptual architecture of smart grid

Until now, a lot of research has been done on the privacy protection of smart grid. The existing privacy protection protocols mainly focus on data aggregation [5]. The data aggregation protocol enables the operation center to obtain only the sum or average of electricity consumption of users in a region, but not the fine-grained electricity consumption data of each user. Therefore, the privacy of users is obtained. However, many existing aggregation scheme models are mainly for single-dimensional electricity consumption data,

*Corresponding Author: Chin-Feng Lai; E-mail: cinfon@ieee.org
DOI: 10.3966/160792642021012201009

which is inconsistent with real life [6]. In addition, existing data aggregation schemes have problems such as low efficiency, insufficient security, and failure to classify electricity data types, which hinder the construction and development of smart grids to a certain extent.

1.1 Our Contributions

- **The cube data structure is proposed to realize multi-dimensional data aggregation both in time and space.** Analyze the sensitivity of users in different regions to electricity prices, including the sensitivity to electricity prices in different seasons and time periods. On the basis of classification analysis, through aggregation, the total amount of demand response that a certain area or a certain type of user needs to provide can be obtained, and then analyze how many time periods the demand response volume is reliable, and the analysis result can provide a basis for formulating demand response incentive mechanisms [7]. In the previous schemes, it needs to execute the aggregation protocol multiple times to provide this information, because in these protocols, time and space are handled separately. However, in the protocol we proposed, through the design of a cube structure, the unity of space and time in electricity data aggregation is realized. Only one implementation of the aggregation protocol can provide a basis for formulating demand response incentive mechanisms.
- **A privacy protection mechanism with forward security is designed in the smart grid.** Since real-time electricity consumption data can reflect the personal behavior of users, it will bring potential privacy risks. Therefore, protecting the privacy of users is essential to the large-scale construction of smart grids. In addition, Once the key is leaked, the confidentiality and integrity of user's data will be destroyed by the adversary. Whether the user transmits false information to the operation center or the operation center sends untrue operation instructions to the user, serious consequences will occur. Therefore, in order to resist key leakage attacks, the session keys of users need to be updated periodically.
- **The performance of the proposed scheme has been improved under the premise of meeting many security requirements.** Security analysis shows that our protocol can meet multiple security features such as privacy protection, data integrity, confidentiality, and forward security. And under the premise of meeting these security goals, the performance of our scheme is also superior to other similar data aggregation schemes.

1.2 Organization

The reminder of this paper is organized as follows. A detailed description of the related work is given in section 2. System model, threat model and security goals are presented in section 3. The preliminaries required for our scheme are described in section 4. The proposed scheme is given in in section 5. Security analysis is demonstrated in section 6. And the performance analysis of the proposed scheme is presented in section 7. Finally, we conclude this paper in section 8.

2 Related Works

Set up a gateway for each community to manage all users in that community. Smart meters collect user electricity consumption data periodically. For users, these real-time electricity consumption data are closely related to user privacy. The leakage of electricity consumption data will bring security threats to users. For this reason, many privacy-preserving data aggregation schemes for smart grid have been proposed.

In [8], the authors proposed a data aggregation scheme with key evolution. In this scheme, the user and the gateway, and the gateway and the power company negotiate a session key. The key update mechanism is designed to resist key leakage attacks. However, this scheme only considers electricity consumption data as a single dimension, which is inconsistent with reality. To solve the problem, in [9], a privacy-preserving data aggregation scheme is proposed based on the paillier encryption system. In this scheme, a super-growth sequence is constructed to represent multi-dimensional electricity consumption data. But this scheme has the problem of excessive communication overhead. To improve the efficiency, in [5], aiming at a large number of users and dynamic topology networks, an effective privacy protection data aggregation scheme was proposed. This scheme uses Horner's law to express the multi-dimensional electricity consumption data with a specific polynomial, which greatly reduces the communication overhead. But this scheme cannot resist key leakage attacks. [10] proposed the first smart grid privacy data aggregation scheme against internal attacks. In this scheme, blind factor technology is used to ensure that the attacker cannot obtain real electricity consumption data. However, this scheme uses linear pair operation, which requires a relatively large amount of calculation and cannot guarantee integrity. In [11], a data aggregation scheme that can resist differential privacy attacks was proposed. In this scheme, Laplacian noise is added to the user's electricity consumption data to achieve differential privacy. in [12], a lightweight data aggregation scheme was proposed. In this scheme, only some lightweight cryptographic primitives, such as hash, are used to aggregate users' electricity consumption data.

3 System Model, Threat Model and Security Requirements

3.1 System Model

In this section, the system model, threat model and the security requirements in the proposed scheme are presented. The system model is shown in Figure 2.

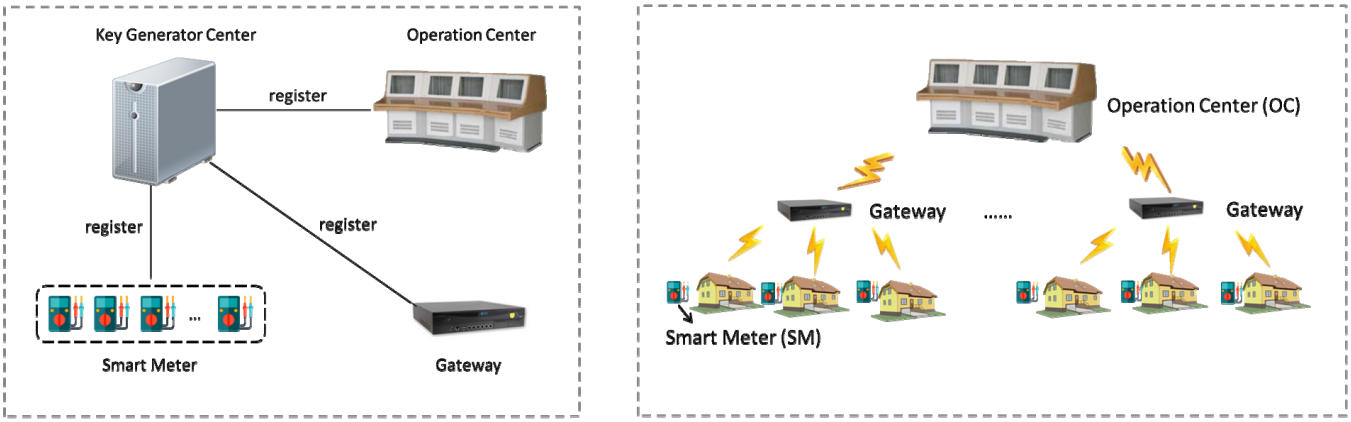

Figure 2. System model

In our system model, There are four entities and two processes. The four entities are described as following:

(1) **KGC:** KGC is completely trusted and is responsible for generating public and private key pairs for smart meters, gateways and operation centers.

(2) **SM:** A smart meter (SM) is installed in the client side, which is used to collect electricity consumption data from a user. It is assumed that the SMs are tamper-proof devices in our scheme, where users cannot modify the reading of the SMs.

(3) **Gateway:** The gateway manages SMs in a region and is responsible for maintaining the revocation list of this region. It collects the ciphertext of electricity consumption data in a region, aggregates the data on the ciphertext, and forwards the aggregation results to the operation center.

(4) **OC:** The operation center (OC) decrypts the aggregated ciphertext and analyzes the electricity consumption data of this region. After analysis, it makes instructions based on that data, such as formulating optimal energy use strategies.

The two processes are described as following:

(1) **Registration:** In this process, all entities register with KGC to obtain public and private key pairs. KGC generates public and private key pairs about the entity based on the entity's identity.

(2) **Data aggregation:** In a fixed time interval, each SM retrieves its own metering data, and applies homomorphic encryption and signature algorithms to generate encrypted and signed data, and then sends the data to the local gateway. After the gateway collects the encrypted metering data of all SMs, it adds them and sends the aggregated data to the OC. The OC can use its private key to retrieve aggregated data. In this way, the OC will obtain the total electricity consumption data of the SMs, but cannot know the specific consumption data of each SM.

3.2 Threat Model

We assume the following attacker models in our scheme:

(1) OC or gateway follows the honest but-curious model. It will execute the protocol correctly, but it may want to extract fine-grained data from the user to analyze the user's privacy and use it for some business transactions.

(2) The external adversary can monitor the channel to obtain the user's consumption data report, or destroy the data integrity to affect the stable operation of the power grid.

3.3 Security Requirements

In our scheme, the following security goals should be realized:

(1) **Privacy preservation:** Although OC can decrypt the aggregated ciphertext, it cannot acquire fine-grained consumption data of each user. Moreover, even if external adversary intercepted a customer's electricity consumption report, he could not decrypt the user's plaintext.

(2) **Authentication and data integrity:** It is necessary to make sure that accepted message comes from an authorized entity and that the modified or forged message should be refused in a timely manner.

(3) **Forward security:** The current session key is leaked will not affect the security of the previous session communication.

4 Preliminaries

In this section, a basic definition of the cryptographic primitives required for our protocol is given.

4.1 Bilinear Pairing

Let $\mathbb{G}_1$ and $\mathbb{G}_2$ be two additive cyclic groups of the same order q, where q is a large prime. $e:\mathbb{G}_1\times\mathbb{G}_2\to\mathbb{G}_2$ is a non-degenerated and efficiently computable bilinear such that:

(1) **Bilinearity:** $e(aP,bQ)=e(P,Q)^{ab}$ for $\forall P,Q\in\mathbb{G}_1$ and $\forall a,b\in\mathbb{Z}_q^*$.

(2) Non-degeneracy: $e(P,P)\neq 1_{\mathbb{G}_2}$ for $\exists P\in\mathbb{G}_1$.

(3) **Computability:** There exists an effective polynomial time algorithm to calculate the value of bilinear pairing.

4.2 Paillier Cryptosystem

The Paillier cryptosystem is homomorphic encryption which consists of the following three algorithms:

(1) **Key Generation:** Let $N=p.q$ and $\upsilon=lcm(p-1,q-1)$, where p and q are two large prime numbers, and $|p|=|q|=\kappa$, where κ is the security parameter of this cryptosystem. Define function $L(x)=(x-1)/N$, select generator $g\in\mathbb{Z}_{N^2}^*$, and calculate $\eta=(L(g^{\upsilon} \bmod N^2))^{-1}$. The public key is $pk=(N,g)$, and the private key is $sk=(\upsilon,\eta)$.

(2) **Encryption:** Given $M\in\mathbb{Z}_N$, select a random number $R\in\mathbb{Z}_N^*$ and calculate the ciphertext $C=E(M)=g^M\cdot R^N \bmod N^2$.

(3) **Decryption:** Given $C\in\mathbb{Z}_{N^2}^*$, recover the corresponding message by $M=D(C)=L(C^{\upsilon} \bmod N^2)\cdot \eta \bmod N$.

5 Our Scheme

5.1 Initialization

In system initialization, the OC inputs security parameter κ and outputs tuples $(q,P,\mathbb{G}_1,\mathbb{G}_2,e)$ by calling $Gen(\kappa)$. It then selects security parameter and calculates public key $(N=p.q,g)$ and private key (υ,η) of the Paillier cryptosystem. And chooses a secure hash function $h:\{0,1\}^*\to\mathbb{G}_1$. The OC randomly selects two number δ, ξ and $\delta>nK$, $\xi>nK$, where n is the maximum number of users in a region, K is the maximum value of each dimensional. At last, the OC publishes public parameter $\{(q,P,\mathbb{G}_1,\mathbb{G}_2,e);N,g,h,\delta,\xi\}$.

In addition, because confidential messages need to be transmitted between the user and the OC, such as demand response messages, it is necessary to establish a session key between the user and the gateway, and between the gateway and the OC. When the SM, gateway and the OC send registration request to KGC, the detailed process is shown as follows:

(1) Firstly, the KGC selects random number $\alpha\in\mathbb{Z}_q^*$, computes $P_{OC}=\alpha h(OC_{ID})$ for the OC. Then KGC computes $gsk=\alpha h(OC_{ID}\,||\,GT_{ID})$ as the gateway's private key, $gpk=gsk\cdot P$ as the gateway's public key, and grants the private key to the gateway through the secure channel, where OC_{ID} represents the identity of the OC and GT_{ID} represents the identity of the gateway.

(2) Secondly, when the gateway receives the private key gsk, it can non-interactively share a session key K_{GT-OC} with the OC. The gateway computes $K_{GT-OC}=e(gsk,h(OC_{ID}))$, and the OC computes $K_{OC-GT}=e(h(OC_{ID}\,||\,GT_{ID}),P_{OC})$. The correctness of calculating the session key is shown in Equation 1.

$$\begin{aligned} K_{GT-OC} &= e(gsk,h(OC_{ID})) \\ &= e(\alpha h(OC_{ID}\,||\,GT_{ID}),h(OC_{ID})) \\ &= e(h(OC_{ID}\,||\,GT_{ID}),\alpha h(OC_{ID})) \\ &= e(h(OC_{ID}\,||\,GT_{ID}),P_{OC}) \\ &= K_{OC-GT} \end{aligned} \quad \textbf{(1)}$$

Besides, the KGC selects random number $x_i\in\mathbb{Z}_q^*$ and computes $P_{GT}=x_i\cdot h(GT_{ID})$ for the gateway, and computes $usk_i=x_i\cdot h(U_{ID_i}\,||\,GT_{ID})$ as the private key of the SM that submitted the registration request to it and computes $upk_i=usk_i\cdot P$ as its public key, where U_{ID_i} represents the identity of the SM. Then the KGC grants the private key to the SM through the secure channel.

(3) Lastly, when the SM receives its private key usk_i, it can non-interactively share a session key K_{U_i-GT} with the gateway. The SM computes $K_{U_i-GT}=e(usk_i,h(GT_{ID}))$, and the gateway computes $K_{GT-U_i}=e(h(U_{ID_i}\,||\,GT_{ID}),P_{GT})$. The correctness of calculating the session key is shown in Equation 2.

$$\begin{aligned} K_{U_i-GT} &= e(usk_i,h(GT_{ID})) \\ &= e(x_i\cdot h(U_{ID_i}\,||\,GT_{ID}),h(GT_{ID})) \\ &= e(h(U_{ID_i}\,||\,GT_{ID}),x_i h(GT_{ID})) \\ &= e(h(U_{ID_i}\,||\,GT_{ID}),P_{GT}) \\ &= K_{GT-U_i} \end{aligned} \quad \textbf{(2)}$$

5.2 Electricity Consumption Report Generation

In our scheme, it is assumed that SM collects multi-dimensional electricity consumption data $(d_{ij1},d_{ij2}...d_{ijl})$ at time t_j. Then the all SMs' data in a region in time

period $t_1 \sim t_m$ can be presented in Figure 3. User uses Equation 3 to calculate the multi-dimensional fine-grained electricity consumption data at this moment.

$$M_{ij} = \xi^i \delta^1 d_{ij1} + \xi^i \delta^2 d_{ij2} + ... + \xi^i \delta^l d_{ijl} \quad (3)$$

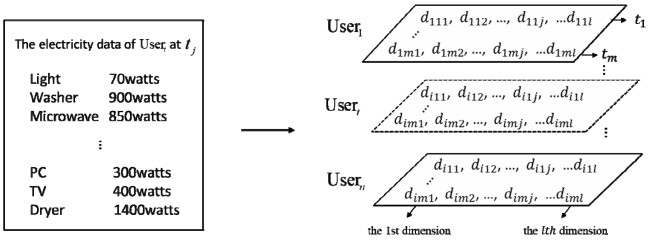

Figure 3. The cube data structure

After that, the SM chooses a random number $r_{ij} \in Z_N^*$ and calculates cipertext in the form of the Paillier cryptosystem as Equation 4.

$$C_{ij} = g^{M_{ij}} . r_{ij}^{\ N} \bmod N^2 \quad (4)$$

To authenticate the integrity of data and resist replay attack, the SM uses its private key usk_i to generate a signature $\sigma_{ij} = usk_i \cdot h(C_{ij} \| U_{ID_i} \| t_j)$. At last, the SM packages report $C_{ij} \| U_{ID_i} \| t_j \| \sigma_{ij}$ to the local gateway.

5.3 Report Aggregation

The local gateway first authenticates the signature after receiving the user's data report. The gateway uses Equation 5 to verify the signature.

$$e(P, \sum_{C_{ij} \in set} \sigma_{ij}) = \prod_{C_{ij} \in set} e(upk_i, h(C_{ij} \| IU_{ID_i} \| t_j)) \quad (5)$$

where $set = \{C_{1j}, C_{2j}, ..., C_{nj}\}$ .After the authentication is passed, the local gateway aggregates the data. The aggregating operation is performed in two parts. The first aggregation is based on the time period in Equation 6.

$$D_i = \prod_{j=1}^{m} C_{ij} \bmod N^2 \quad (6)$$

where m is the total time periods.

And the second aggregation is based on the number of users in Equation 7.

$$D = \prod_{i=1}^{n} D_i \quad (7)$$

where n is the total number of users in a region. Through the above two steps of data aggregation, we realize temporal aggregation and spatial aggregation at the same time. Finally, the local gateway uses its private key gsk to generate a signature $\sigma = gsk \cdot h(D \| GT_{ID} \| t)$. And sends the report $D \| GT_{ID} \| t \| \sigma$ to the OC.

5.4 Report Reading

When receiving the report from the gateway, the OC first verifies the validity of the report in Equation 8.

$$e(P, \sigma) = e(gpk, h(D \| GT_{ID} \| t)) \quad (8)$$

If the verification fails, ignore this packet. Otherwise, the aggregation results are expanded and substituted in the following form to obtain Equation 9.

$$\begin{aligned} D &= \prod_{i=1}^{n} D_i \\ &= \prod_{i=1}^{n} (\prod_{j=1}^{m} C_{ij} \bmod N^2) \\ &= \prod_{i=1}^{n} (\prod_{j=1}^{m} g^{M_{ij}} . r_{ij}^{\ N} \bmod N^2) \\ &= \prod_{j=1}^{m} (\prod_{i=1}^{n} g^{M_{ij}} . r_{ij}^{\ N} \bmod N^2) \\ &= \prod_{j=1}^{m} (g^{\sum_{i=1}^{n} M_{ij}}) . (\prod_{j=1}^{m} \prod_{i=1}^{n} r_{ij})^N \bmod N^2 \\ &= g^{\sum_{j=1}^{m} (M_{1j} + M_{2j} + ... + M_{nj})} . (\prod_{j=1}^{m} \prod_{i=1}^{n} r_{ij})^N \bmod N^2 \\ &= g^{\sum_{j=1}^{m} (\sum_{w=1}^{l} \xi^1 \delta^w d_{1jw} + \sum_{w=1}^{l} \xi^2 \delta^w d_{2jw} + ... + \sum_{w=1}^{l} \xi^n \delta^w d_{njw})} \cdot \\ &\quad (\prod_{j=1}^{m} \prod_{i=1}^{n} r_{ij})^N \bmod N^2 \\ &= g^{\xi^1 \sum_{w=1}^{l} \delta^w \sum_{j=1}^{m} d_{1jw} + \xi^2 \sum_{w=1}^{l} \delta^w \sum_{j=1}^{m} d_{2jw} + ... + \xi^n \sum_{w=1}^{l} \delta^w \sum_{j=1}^{m} d_{njw}} \cdot \\ &\quad (\prod_{j=1}^{m} \prod_{i=1}^{n} r_{ij})^N \bmod N^2 \end{aligned} \quad (9)$$

Let $DM = \xi^1 \sum_{w=1}^{l} \delta^w \sum_{j=1}^{m} d_{1jw} + \xi^2 \sum_{w=1}^{l} \delta^w \sum_{j=1}^{m} d_{2jw} + ... + \xi^n \sum_{w=1}^{l} \delta^w \sum_{j=1}^{m} d_{njw}$, $R = \prod_{j=1}^{m} \prod_{i=1}^{n} r_{ij}$, we can obtain $D = g^{DM} \cdot R^N \bmod N^2$. By secret key $sk = (\upsilon, \eta)$, we can acquire DM.

Let $DM_i = \sum_{w=1}^{l} \delta^w DM_{iw}$ and $DM_{iw} = \sum_{j=1}^{m} d_{ijw}$, then DM can also be expressed as $DM = \sum_{i=1}^{n} \xi^i \cdot DM_i$.

Algorithm 1 is designed to obtain $(DM_1, DM_2, ... DM_n)$, that is the total electricity consumption data of each user in all dimensions during $t_1 \sim t_m$. Based on DM_i, the electricity consumption data of each user in each dimensions DM_{iw} during $t_1 \sim t_m$ an also be derived by Algorithm 1. The detailed steps of Algorithm 1 are

shown in the following. Since the OC can only obtain the user's electricity consumption in multiple time periods, and cannot obtain the user's real-time fine-grained consumption data, the user's daily life habits cannot be derived from these data. Therefore, the protocol not only protects the privacy of users, but also allows OC to obtain the electricity consumption information it needs.

Algorithm 1. Parsing polynomials to obtain functional aggregated data

Input: DM and ξ

Output: $(DM_1, DM_2, \ldots DM_n)$

$X_0 = DM / \xi$

1. $= \sum_{i=1}^{n} \xi^i \cdot DM_i / \xi$

   $= DM_1 + \xi DM_2 + \ldots + \xi^{n-1} DM_n$

2. for $i = 1$ to n do

3. $DM_i = X_{i-1} \bmod \xi$

4. $X_i = \frac{X_{i-1}}{\xi}$

5. end for

6. return $(DM_1, DM_2, \ldots DM_n)$

5.5 Key Evolution

Each gateway maintains an area revocation list RL and checks whether the user is in the revocation list every time the smart meter requests a key update. Only when the user is not in the list, the regional gateway will update the private key for him. The gateway reselects a new random number $s_i \in \mathbb{Z}_q^*$ and recalculate the user's private key $usk_i = s_i \cdot h(U_{ID_i} \| GT_{ID})$ and public key $upk_i = usk_i \cdot P$ for him. Then the user can non-interactively share a new session key $K_{U_i - GT}$ with the gateway. Because the session key is dynamic as the key generation involve unique random number s_i, if any session key is captured by the adversary, it will not impact the secrecy of past session keys.

6 Security Analysis

Based on the ECDLP and the decisional composite residuosity assumption [13], the proposed scheme can be proven secure. In this section, the proposed scheme is proven to be secure against passive adversaries and active adversaries.

6.1 Security Against Passive Attacks

A passive adversary is someone who tries to obtain information by eavesdropping on the channel. Note that an adversary can get the public parameter $\{(q, P, \mathbb{G}_1, \mathbb{G}_2, e); N, g, h, \delta, \xi\}$ and the public key upk_i for U_{ID_i}. However, the private key usk_i for U_{ID_i} can not be duduced because solving the ECDLP problem is difficult. On the basis of [14], if $X \approx_{ploy} Y$, then the proposed scheme is secure against passive adversaries. $X \approx_{ploy} Y$ denotes that the following two tuples $X = \{(q, P, \mathbb{G}_1, \mathbb{G}_2, e); N, g, h, \delta, \xi, e(usk_i, h(GT_{ID}))\}$ and $Y = \{(q, P, \mathbb{G}_1, \mathbb{G}_2, e); N, g, h, \delta, \xi, y\}$ satsifies polynomial indistinguishability, where $y \in \mathbb{Z}_q^*$. Because of the difficulty in solving discrete logarithm problem over elliptic curves, it can be concluded that $X \approx_{ploy} Y$.

6.2 Security Against Active Attacks

In an active attack, an adversary not only acquires information by eavesdropping on the channel, but also launch modification attacks, replay attacks, forgery attacks etc. To resist active attacks, the proposed data aggregation scheme should meet the following security goals.

- **Integrity and authentication.** When the user generates a data report, he needs to sign the report with his own private key in $\sigma_{ij} = usk_i \cdot h(C_{ij} \| U_{ID_i} \| t_j)$. Then the gateway can authenticate users by Equation 10.

$$\begin{aligned} e(P, \sum_{C_{ij} \in set} \sigma_{ij}) &= e(P, \sum_{C_{ij} \in set} usk_i \cdot h(C_{ij} \| U_{ID_i} \| t_j)) \\ &= \prod_{C_{ij} \in set} e(usk_i \cdot P, h(C_{ij} \| U_{ID_i} \| t_j)) \\ &= \prod_{C_{ij} \in set} e(upk_i, h(C_{ij} \| U_{ID_i} \| t_j)) \end{aligned} \quad \textbf{(10)}$$

Besides, in our scheme, the consumption data is hashed by $h(C_{ij} \| U_{ID_i} \| t_j)$, once an external adversary tampers with the user's consumption data, it will be detected by the OC. As a result, the authentication and data integrity of the consumer's electricity consumption report are assured in the proposed scheme.

- **Privacy-preserving.** In our scheme, the user's consumption report is encrypted in $C_{ij} = g^{M_{ij}} . r_{ij}^{N} \bmod N^2$ and then is sent to the regional gateway. With the homomorphic property of Paillier encryption, the gateway can perform data aggregation operations only on the ciphertext. Since the gateway does not have the private key for decryption, the gateway cannot obtain the user's electricity consumption report. Thus, the privacy of users is guaranteed in regional gateway. Secondly, because Paillier encryption is secure under known plaintext attacks, even if the adversary intercepts the user's consumption data report, the plaintext cannot be obtained. Finally, after receiving D, the OC can only recover D as

DM and $(DM_1, DM_2, \ldots DM_n)$, so the OC can only obtain the aggregated data of all dimensions of all users in a region, and the sum of the electricity consumption data of all dimensions of a user in a period of time, but cannot obtain each user's fine-grained electricity consumption data every moment. Besides, even if the adversary hack the database of the OC, he cannot acquire the fine-grained electricity data for specific users. Therefore, the privacy of users is guaranteed in our scheme.

- **Forward security.** If the current session key is leaked will not affect the security of the previous session communication, then this protocol provides forward security. In our protocol, we calculate the previous session key in $K_{U_i-GT} = e(usk_i, h(GT_{ID}))$, where $usk_i = x_i \cdot h(U_{ID_i} \| GT_{ID})$, calculate the current session key in $K_{U_i-GT} = e(usk_i, h(GT_{ID}))$, where $usk_i = s_i \cdot h(U_{ID_i} \| GT_{ID})$. Even the current session key is compromised, as the random number is dynamically updated with every session, it will not impact the secrecy of past session keys. Thus, the proposed approach satisfies forward protection.

7 Performance Analysis

7.1 Features Comparison

In this subsection, a comparison about features of our proposed scheme with those of other schemes is given in Table 1.

Table 1. Features comparison

Schemes	Ours	[8]	[9]	[11]	[12]	[15]	[16]	[17]
Authentication	√	×	√	×	×	×	√	√
Consumer's privacy	√	√	√	√	√	√	√	√
Data integrity	√	×	√	√	√	×	√	×
Data confidentiality	√	√	√	√	√	√	√	√
Key evloution	√	√	×	×	×	×	×	×
Scalability	√	×	×	×	√	√	×	√
Forward secrecy	√	√	×	×	×	×	×	×

Only our protocol has all the features while meeting the proposed security goals. We uses batch verification to achieve fast and safe user authentication. BLS short signature is utilized to guarantee the data integrity during transmission. In addition, considering that key leakage will seriously endanger the security of the system, a key update mechanism is proposed. Regularly update the user's private key according to the time period and the revocation list, making our scheme forward-secure. Furthermore, Our protocol is not limited by the number of users and has good scalability.

7.2 Computational Complexity

In this subsection, the proposed scheme and the compared schemes are simulated on pairing-based cryptography (PBC) library to compare their performance. Execution time demonstrates the complexity of the proposal, which can be evaluated by time cost in report generation of each user and report aggregation. In our simulation, the time cost of the two phases is separately simulated. It can be seen from Figure 4 that the time cost of report generation in our scheme and in [11] is constant as the dimension of electricity data increases, and increases linearly with the dimension of electricity data in [9]. The time cost of our scheme is much smaller than the cost of the other two schemes. That's because in our scheme, multi-dimensional electricity data $(d_{ij1}, d_{ij2} \ldots d_{ijl})$ is represented as a whole in the form of $M_{ij} = \xi^i \delta^1 d_{ij1} + \xi^i \delta^2 d_{ij2} + \ldots + \xi^i \delta^l d_{ijl}$, which greatly reduces the cost of data report generation. In Figure 5, we can see that the time overhead in report aggregation increases linearly with the number of users in these schemes. It can be seen from the Figure 5 that our scheme has a slightly higher cost than the EPPA scheme. That is because our scheme can achieve multi-functional data aggregation in time and space.

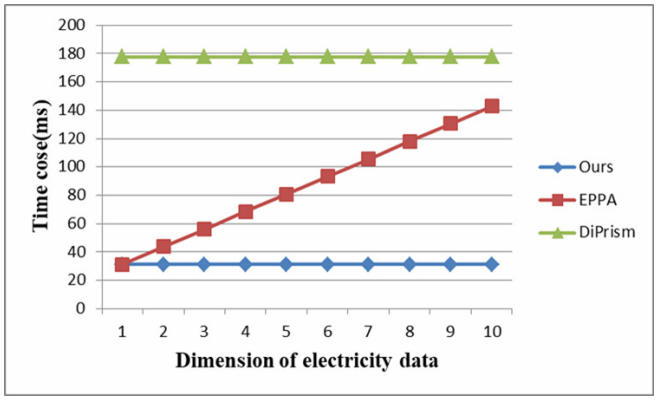

Figure 4. Time cost in Report Generation of each user

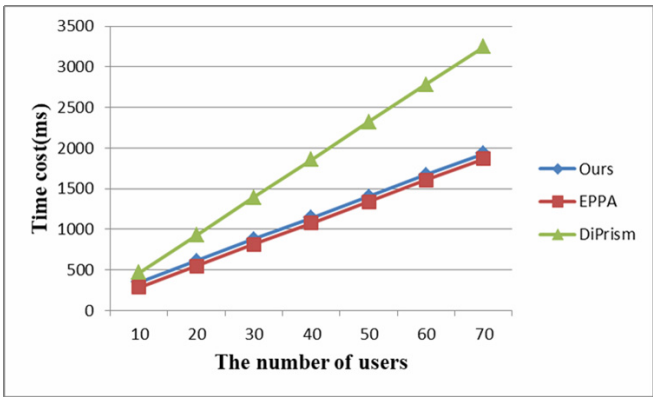

Figure 5. Time cost in Report Aggregation

8 Conclusion

By designing a cube data structure, this paper can realize spatial multi-dimensional data aggregation and temporal multi-dimensional data aggregation at the same time. In this scheme, homomorphic encryption is applied to realize the privacy of users' electricity consumption information. Besides, it also uses identity-based batch verification to ensure the integrity and authentication of users' electricity consumption report. Finally, security analysis proves that the proposed scheme meet multiple security requirements, and the comparison results of performance analysis show that this protocol is better than other similar protocols and is more practical.

References

[1] M. Z. Gunduz and R. Das, Cyber-security on smart grid: Threats and potential solutions, *Computer Networks*, Vol. 169, pp. 107094, March, 2020.

[2] K. Kimani, V. Oduol, and K. Langat, Cyber security challenges for iot-based smart grid networks, *International Journal of Critical Infrastructure Protection*, Vol. 25, pp. 36-49, June, 2019.

[3] M. H. Rehmani, A. Davy, B. Jennings, and C. Assi, Software defined networks-based smart grid communication: A comprehensive survey, *IEEE Communications Surveys and Tutorials*, Vol. 21, No. 3, pp. 2637-2670, Third Quarter, 2019.

[4] C. Wang, J. Shen, J.-F. Lai, and J. Liu, B-TSCA: Blockchain assisted trustworthiness scalable computation for V2I authentication in VANETs, *IEEE Transactions on Emerging Topics in Computing*, DOI: 10.1109/TETC.2020.2978866, March, 2020.

[5] H. Shen, M. Zhang, and J. Shen, Efficient privacy-preserving cube-data aggregation scheme for smart grids, *IEEE Transactions on Information Forensics and Security*, Vol. 12, No. 6, pp. 1369-1381, June, 2017.

[6] M. Feng, C.-F. Lai, H. Liu, R. Qi, and J. Shen, A novel identity-based broadcast authentication scheme with batch verification for wireless sensor networks, *Journal of Internet Technology*, Vol. 21, No. 5, pp. 1303-1311, September, 2020.

[7] Y. Yoldas, A. Önen, S. M. Muyeen, A. V. Vasilakos, and İ. Alan, Enhancing smart grid with microgrids: Challenges and opportunities, *Renewable and Sustainable Energy Reviews*, Vol. 72, pp. 205-214, May, 2017.

[8] H. Li, X. Lin, H. Yang, X. Liang, R. Lu, and X. Shen, EPPDR: An efficient privacy-preserving demand response scheme with adaptive key evolution in smart grid, *IEEE Transactions on Parallel and Distributed Systems*, Vol. 25, No. 8, pp. 2053-2064, August, 2014.

[9] R. Lu, X. Liang, X. Li, X. Lin, and X. Shen, Eppa: An efficient and privacy-preserving aggregation scheme for secure smart grid communications, *IEEE Transactions on Parallel and Distributed Systems*, Vol. 23, No. 9, pp. 1621-1631, September, 2012.

[10] C. I. Fan, S. Y. Huang, Y. L. Lai, Privacy-enhanced data aggregation scheme against internal attackers in smart grid, *IEEE Transactions on Industrial informatics*, Vol. 10, No. 1, pp. 666-675, February, 2014.

[11] J. Ni, K. Zhang, K. Alharbi, X. Lin, N. Zhang, and X. Shen, Differentially private smart metering with fault tolerance and range-based filtering, *IEEE Transactions on Smart Grid*, Vol. 8, No. 5, pp. 2483-2493, September, 2017.

[12] P. Gope and B. Sikdar, Lightweight and privacy-friendly spatial data aggregation for secure power supply and demand management in smart grids, *IEEE Transactions on Information Forensics and Security*, Vol. 14, No. 6, pp. 1554-1566, June, 2019.

[13] P. Paillier, Public-key cryptosystems based on composite degree residuosity classes, *EUROCRYPT '99, International Conference on the Theory and Application of Cryptographic Techniques*, Prague, Czech Republic, 1999, pp. 223-238.

[14] D. Boneh and M. Franklin, Identity-based encryption from the weil pairing, *Siam Journal on Computing*, Vol. 32, No. 3, pp. 586-615, March, 2003.

[15] X. F. Wang, Y. Mu, and R. M. Chen, An efficient privacy-preserving aggregation and billing protocol for smart grid, *Security and Communication Networks*, Vol. 9, No. 17, pp. 4536-4547, November, 2016.

[16] E. Vahedi, M. Bayat, M. R. Pakravan, and M. R. Aref, A secure ECC-based privacy preserving data aggregation scheme for smart grids, *Computer Networks*, Vol. 129, pp. 28-36, December, 2017.

[17] Y. Zhang, J. Zhao, and D. Zheng, Efficient and privacy-aware power injection over AMI and smart grid slice in future 5G networks, *Mobile Information Systems*, Vol. 2017, Article ID 3680671, January, 2017.

Biographies

Guojun Wang received a master's degree in software engineering from Nanjing University in 2015, and has been an associate professor of China Yancheng Polytechnic College since 2018. His research interests include cloud computing and security, and information security systems.

Xueya Xia received the BS degree, in 2019. 1011 She is currently working toward the ME degree at the Nanjing University of Information Science and Technology, Nanjing, China. Her current research interests include data access control and privacy preserving in smart grid.

Sai Ji received his M.S. degree from the Nanjing Aeronautics and Astronautics University (NUAA), Nanjing, China, in 2006. He works as an Associate Professor at the NUIST. His research interests are in the areas of structural health monitoring, and WSNs. Ji has published more than 20 journal/ conference papers.

Chin-Feng Lai (SM'14) received the Ph.D. degree in engineering science from National Cheng Kung University, Tainan,Taiwan, in 2008. Since 2016, he has been an Associate Professor of Engineering Science, National Cheng Kung University, Tainan. His research focuses on Internet of Things, body sensor networks, e-healthcare, mobile cloud computing, etc.