PPMM-DA: Privacy-Preserving Multi-dimensional and Multi-subset Data Aggregation with Differential Privacy for Fog-based Smart Grids

Shuai Zhao, Shuhua Xu, Song Han, Siqi Ren, Ye Wang, Zhixian Chen, Xiaoli Chen, Jianhong Lin and Weinan Liu

***Abstract*—The smart grid (SG) is a new type of grid that integrates traditional power grid with the Internet of Things (IoT) to make the entire grid system more compatible, controllable and self-healing. However, the flourishing of SG still faces some challenges in term of privacy-preserving data aggregation. Previous multi-dimensional data aggregation schemes need heavy computation operations, cannot support multi-subset data aggregation, and resist neither collusion attack among the gateway (GW) and control center (CC) nor differential attack. To solve these issues, we propose a privacy-preserving data aggregation scheme for fog-based smart grids to achieve multi-dimensional and multi-subset data aggregation. The parallel composability of differential privacy is used to reasonably allocate the privacy budget, which can provide higher data utility in multi-dimensional data aggregation. In addition, each user's multi-dimensional power consumption data will be structured as a composite data by utilizing Chinese Remainder Theorem (CRT), which will further reduce the computational overhead. Security analysis shows that our scheme can resist differential attack, eavesdropping attack, collusion attack and active attack. Evaluation of the performance also demonstrates that our scheme is more efficient in terms of computational overhead and communication overhead.**

***Index Terms*—Data utility, privacy-preserving, smart grid, fog computing, Industrial Internet of Things, multi-dimensional and multi-subset data aggregation, differential privacy.**

I. Introduction

THE SG is an advanced digital two-way flow power system, especially in Industrial Internet of Things, with self-healing, adaptive and resilient [1]–[3]. In order to achieve the real-time monitoring and optimal control of the power grid, the SG needs to collect a large number of power consumption data from users via IoT devices such as smart meters (SMs). The collection and transmission of power consumption data for SMs bring risks of individual user's privacy leakage [4]. How to protect the security and privacy of power consumption data has become a major concern [5]. Hence, the collaborative interaction and cooperation among users, devices, and environments are critical [6]. We need to ensure the security of data exchange and the privacy of the computation, etc [7], [8].

Existing works [9]–[15] allow SMs to report one type of data, and the CC can only obtain the total power consumption data for all users. However, the fog-based smart grid system has different types of devices such as refrigerator, microwave, oven, TV, etc. In order to analyze the data more deeply and optimize the data more finely, the CC needs more detailed power consumption data for each device. Some solutions [16]–[22] are dedicated to solving the fine-grained problem of multi-dimensional data aggregation. However, these schemes use heavy computation operations to encrypt the multi-dimensional data, and the collusion attack is not explicitly addressed. In addition, one of the key challenges for data aggregation is how to achieve multi-subset aggregation of multi-dimensional data in the SG, which is expected to reduce the peak-to-average ratio, balance energy supply and demand.

Another challenging issue for data aggregation is differential attack [23]–[26]. Specifically, even if an aggregation scheme is secure, once adversary obtains the sum of n and $n-1$ users' power consumption data, the private data of different user will inevitably be obtained. This issue has been studied in [27]–[34], but they only achieve differential privacy for one-dimensional data. Therefore, it remains a critical challenge for multi-dimensional data aggregation to resist differential attack, offer provable differential privacy guarantees on the aggregated statistic and improve the utility of data.

Futhermore, the above literatures adopt the traditional SG model, that is, the GW aggregates the data reports from SMs and transmits them to CC for further analysis. However, these SMs generate a large amount of data every 15 minutes, which will put a heavy burden on the cloud, cause huge network delay and endanger the privacy and security of personal data [37], [38]. Actually, the fog computing model can extend cloud computing functions from the network center to the network edge, optimize local computing capacity and make the collection, transmission and processing of big data more

This work was supported in part by the National Natural Science Foundation of China under Grant 61906167, Grant 62072403, and Grant 62072404, in part by the Natural Science Foundation of Zhejiang Province under Grant LTY21F020001, Grant LQ22F020001, and Grant LY21F020011, in part by the Pioneer and Leading Goose R&D Program of Zhejiang Province under Grant 2022C01243, in part by the Key Research and Development Program of Zhejiang Province under Grant 2020C01076, in part by the Ministry of Education Humanities and Social Sciences General Project under Grant 22JDSZ3124, and in part by the Laboratory Work Research Project of Universities in Zhejiang under Grant YB202240. Corresponding author: S. Han (hans@hzcu.edu.cn; hansongau215@gmail.com).

S. Zhao, S. Xu, S. Ren, Y. Wang, and Z. Chen are with the School of Computer Science and Technology, Zhejiang Gongshang University, Hangzhou 310018, China. S. Xu and S. Zhao contributed equally to this work.

S. Han is with the School of Computer and Computing Science, Hangzhou City University (a.k.a Zhejiang University City College), Hangzhou, 310015, China. S. Han is also with Zhejiang Ponshine Information Technology Co., Ltd., Hangzhou 311100, China.

J. Lin is with the College of Computer Science and Technology, Zhejiang University, Hangzhou 310027. X. Chen and J. Lin are with Zhejiang Ponshine Information Technology Co., Ltd., Hangzhou 311100, China.

W. Liu is with the Business & Tourism Institute, Hangzhou Vocational & Technical College, Hangzhou 310018, China.

efficient [39]–[41]. Using the fog computing model in data aggregation schemes can improve computational efficiency and reduce transmission delay [42], [58], [59].

In sum, in this paper, we present a Privacy-Preserving Multi-dimensional and Multi-subset Data Aggregation (PPMM-DA) scheme with differential privacy for fog-based smart grids. Our major contributions are summarized as below:

- First, in order to analyze the power consumption data more deeply and implement effective grid monitoring and management, our scheme achieves multi-dimensional and multi-subset data aggregation by employing the CRT, super-increasing sequence and Paillier cryptosystem.
- Second, each user's multi-dimensional power consumption data will be structured as one composite data by CRT. We can extract aggregated results of power consumption for each dimension from the aggregated ciphertext, which makes computational overhead be independent of the dimension of power consumption data, results in significant saving in computational overhead. Besides, our scheme also can resist collusion attack among the GW and CC.
- Third, in order to realize privacy-enhanced multi-dimensional data aggregation and offer provable differential privacy guarantees on the aggregated statistic, we propose two methods to achieve differential privacy, which extracted noise from Geometric distribution and Laplace distribution, respectively. Meanwhile, the parallel composability of differential privacy is used to reasonably allocate the privacy budget, which can provide higher data utility. Then the data utility of these two methods is compared and analyzed.

The roadmap of this paper is as follows. The related works are given in Section II. The problem formalization is introduced in Section III. Then we recall the Paillier cryptosystem, differential privacy, CRT as the preliminaries in Section IV. In Section V, we introduce our PPMM-DA scheme. To achieve privacy-enhanced multi-dimensional data aggregation, we propose the scheme with differential privacy in Section VI. In Section VII, the correctness and security of our scheme are analyzed. Section VIII presents performance comparison, experimental results and differential privacy comparison. Finally, we conclude this paper in Section IX.

II. RELATED WORK

Research on privacy-preserving is conducted on all phases of the information life cycle, including information collection, storage, processing, publication and destruction [9]. As a key phase of data collection and processing, data aggregation has become one of the research hotspots, and many methods have arisen. In [16], Lu et al. presented the first multi-dimensional data aggregation scheme utilizing super-incremental sequence and Paillier cryptosystem. Chen et al. [17] proposed a data aggregation scheme that enables the SM to report more than one type power consumption data, and allows the CC to perform variance analysis and one-way analysis of the variance on the power consumption data. However, they can only provide the result of global aggregation for the CC, and can not satisfy more fine-grained requirements. At the same time, they do not solve the collusion attack between the GW and CC. In [23], Lu et al. first tried to divide users into two subsets based on power consumption according to an additive homomorphism of composite order cipher set, but it does not support multi-subset data aggregation. From the perspective of multi-subset data aggregation, Li et al. [24] proposed a PPMA scheme, where the CC can obtain the power consumption data of users in different ranges, but the approximate result is obtained instead of accurate aggregated result and it does not support multi-subset data aggregation of multi-dimensional data. Furthermore, data integrity verification [57] and fault tolerance are not considered. In [25], Chien et al. improved PPMA scheme to support fault tolerance, but it still obtains approximate results and the computational overhead is expensive. In addition, the collusion attack is not explicitly addressed. Zuo et al. [56] proposed a scheme for addressing the collusion attack, but it can not resist differential attack.

Recently, many researchers have worked with differential privacy technology to resist differential attack. Bao et al. [30] proposed a data aggregation scheme using differential privacy, where aggregated data is interfered by random noise from Geometric distribution to resist differential attack. They design a novel data aggregation scheme that can flexibly support fault tolerance of malfunctioning SMs. However, their scheme require assigning random values via bi-directional interaction. Ni et al. [31] proposed a new privacy-preserving smart metering scheme for SGs, which supports fault tolerance, differential privacy and range-based filtering simultaneously. Lu et al. used CRT, Paillier encryption and one-way hash chain technique to present a lightweight privacy-preserving data aggregation scheme [36]. Shi et al. [32] presented a diverse grouping-based aggregation scheme to achieve grouping-based private stream aggregation and utilized differential privacy technique to resist differential attack. Through normalizing and modifying the confidence score vectors with a differential privacy mechanism, Ye et al. [53] proposed a one-parameter defense method to against both model inversion and membership inference attacks. Based on randomized responses, Gai et al. [51] presented an efficient data aggregation scheme satisfying local differential privacy with privacy-preserving for smart grids. However, they only achieve differential privacy for one-dimensional data. With the development of SG, we need to focus on how multi-dimensional data aggregation can resist differential attack and improve the utility of data. Lately, some new results in [52], [54], [55] achieved privacy-preserving multi-dimensional data aggregation via differential privacy and resisted differential attack. However, they do not take into consideration other attacks, such as eavesdropping attack, collusion attack and active attack.

The above schemes solve the problems for SGs in different aspects, but there are still many weaknesses. Motivated by the above-mentioned, we are aiming to design an effective privacy-preserving data aggregation scheme for fog-based smart grids, which can achieve multi-subset aggregation of multi-dimensional data, fault tolerance, resist differential attack and collusion attack, and minimize privacy leakage, network latency and loss of data availability.

This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2023.3309132

III. Problem Formalization

A. System Model

We consider a fog-based smart grid system that consists of a trusted authority (TA), a CC, some fog nodes (FNs) and SMs. The architecture of our system model is shown in Figure 1.

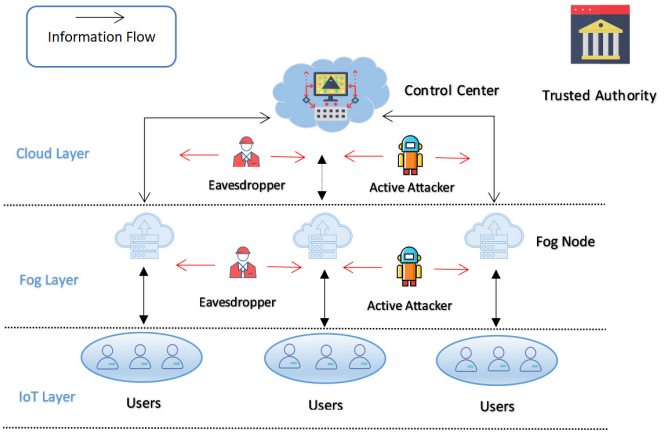


Fig. 1: System model.

- **TA:** In our system model, the TA is a only globally-trusted entity, whose duty is to bootstrap the whole system, manage and distribute key materials to other entities. After that, the TA will turn to offline.
- **CC:** The CC performs Paillier decryption, utilizes CRT and super-increasing sequence to analyze the data more deeply based on the power consumption data of each device and the corresponding number of users within specific power consumption range.
- **FN:** The FN is located in the middle layer between the Cloud layer and the IoT layer, and transfers the computing and storage functions of cloud to the edge of terminal equipment. In our system model, the FN acts as an aggregator to aggregate encrypted data from users within its coverage area.
- **User:** Each user equips with a SM (i.e., IoT node) which encrypts the collected power consumption data and transmits the encrypted report to the nearby FN.

B. Threat Model

In our threat model, we assume that the TA is trustable and other entities are honest-but-curious [35], which means that the CC, FN and users will strictly execute the protocol, but remain curious about other users private information (such as individual users private key and power consumption data). In addition, an external adversary may be lurking in the residential area, eavesdropping on the communication links between various entities or invading the database of the FN to obtain private information. In this paper, we mainly consider the following four types of security threats, i.e., the differential attack, eavesdropping attack, collusion attack and active attack, and other threats are beyond the scope of this paper.

- **Differential attack:** The adversary can analyze the aggregated difference between each multi-dimensional data set to infer device-specific power consumption data for an individual user in the fog-based smart grid.
- **Eavesdropping attack:** The adversary can eavesdrop on the communication links from the IoT layer to the Fog layer or from the Fog layer to the Cloud layer to obtain the transmitted power consumption data, and try to compromise the individual users private power consumption information.
- **Collusion attack:** The FN is allowed to collude with some users and the CC is able to collude with the FN or some users by sharing and analyzing their private information, which consists of private key, public parameters, and ciphertext of power consumption data, to obtain an individual users private information, i.e., individual power consumption data.
- **Active attack:** The adversary may invade into the FN to steal users private information. By intercepting messages, the adversary can also forge the identities of trusted users to transmit false power consumption data to compromise the integrity of data in the fog-based smart grid.

TABLE I: Main notations

Notation	Definition
$\mathbb{G}_1,\ \mathbb{G}_2$	Two multiplication cyclic groups
g	A generator of $\mathbb{G}_1$
e	The bilinear map e: $\mathbb{G}_1 \times \mathbb{G}_1 \longrightarrow \mathbb{G}_2$
H	Hash function $H : \{0,1\}^* \rightarrow \mathbb{Z}_N^*$
H_1	Hash function $H_1 : \{0,1\}^* \rightarrow \mathbb{G}_1$
$G(i),\ Y_i,\ \sigma_i$	Private key, public key and signature of $user_i$
$x_{fn},\ Y_{fn},\ \sigma_{fn}$	Private key, public key and signature of FN
x_0	Private key of CC
C_i	The ciphertext of the i-th user
C	Aggregated ciphertext
M_k	The sum of the k-th dimensional power consumption data
$\|U_j\|$	The number of users in the j-th subset
$[R_j, R_{j+1})$	The j-th subset

C. Design Goal

- **Multi-dimensional and Multi-subset:** To conduct fine-grained analysis and achieve effective grid monitoring and management, the CC should obtain power consumption data of each dimension and the corresponding number of users within specific power consumption range.
- **Privacy and Security:** The power consumption data of individual user should be protected in the system. The proposed data aggregation scheme should meet the security requirements and can resist differential attack, eavesdropping attack, collusion attack and active attack.
- **Efficiency:** Due to the limited resource of terminals, the proposed data aggregation scheme should be extremely efficient at each phase of the system.

IV. Preliminaries

In this section, we give an overview of Paillier cryptosystem [43], differential privacy [44] and CRT [49], all of which

serve as the basis of our proposed scheme. The main notations of this paper and their definitions are summarized in Table I.

A. Paillier Cryptosystem

Paillier cryptosystem is composed of three algorithms: key generation, encryption and decryption.

- **Key generation:** Given the security parameter κ, two large and independent prime numbers p, q are first chosen, where $|p| = |q| = \kappa$. Then, the RSA modulus $N = pq$ and $\lambda = lcm(p-1, q-1)$ are computed. Define a function $L(\varphi) = \frac{\varphi-1}{N}$, after choosing a generator $g = (1+N), \varphi = (L(g^{\lambda} \ mod \ N^2))^{-1} \ mod \ N$ is further calculated. The public key is (N, g), and the private key is (λ, μ).
- **Encryption:** Given a message $m \in \mathbb{Z}_N$, first, we choose a random number $r \in \mathbb{Z}_N^*$, and the ciphertext can be calculated as follow:

$$c = E(m) = g^m \cdot r^N \ mod \ N^2. \tag{1}$$

- **Decryption:** Given the ciphertext $c \in \mathbb{Z}_{N^2}^*$, the corresponding plaintext message can be computed as

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

In our proposed scheme, we utilized another form of Paillier cryptosystem. As the generator $g = (1+N)$ and $\phi(N^2) = N \cdot \phi(N) = N \cdot \lambda$, we can conclude the following equations according to Euler theorem:

$$\begin{aligned} c &= (1+N)^m \cdot r^{N \cdot \lambda} \ mod \ N^2 \\ &= (1+N)^m \cdot r^{\phi(N^2)} \ mod \ N^2 \\ &= (1+N)^m \ mod \ N^2. \end{aligned} \tag{3}$$

We can expend the power $(1+N)^m$ with Binomial theorem:

$$\begin{aligned} (1+N)^m &= \sum_{i=1}^{m} \binom{m}{i} N^i \\ &= 1 + mN \ mod \ N^2, \end{aligned} \tag{4}$$

as all items with $i \geq 2$ turn to zero.

B. Differential Privacy

In 2006 [44], Dwork first proposed a differential privacy model, which can achieve the privacy-preserving effectively by adding appropriate noise to the query or analysis result. In addition, parallel composability was proposed in [45].

Definition 1 (Differential Privacy): The aggregation function A gives $\epsilon-$differential privacy if for any data sets D_1 and D_2 differing by at most one element, and for any $O \in Range(A)$, have

$$Pr[A(D_1) \in O] \leq exp(\epsilon) \cdot Pr[A(D_2) \in O], \tag{5}$$

where the probability is taken over the randomness of A. The privacy parameter ϵ called the privacy budget, represents the privacy degree offered by the mechanism. In general, a larger perturbation noise is required for a smaller ϵ, which implies stronger privacy guarantee but worse utility of data set.

Definition 2 (Parallel Composability): Generally, when the data set is relatively complex, the parallel composability of differential privacy will be used to reasonably allocate the privacy protection budget to the algorithm, whose goal is to keep the level of privacy protection within the privacy budget and ensure the security of data. Assuming that D is a privacy database, which can be divided into n disjoint subsets $\{D_1, D_2, ..., D_n\}$. Let $\{A_1, A_2, ..., A_n\}$ be a series of mutually independent differential privacy algorithms, and the corresponding privacy protection budgets of these algorithms are $\{\epsilon_1, \ \epsilon_2, \ ... \ \epsilon_n\}$. Then, the combined algorithm $A(A_1(D_1), A_2(D_2), ..., A_n(D_n))$ will satisfy $max\{\epsilon_1, \epsilon_2, ..., \epsilon_n\}$-differential privacy.

C. Chinese Remainder Theorem

The Chinese Remainder Theorem can uniquely solve any pair of congruences that have relatively prime moduli [36], [49]. In our PPMM-DA scheme, we can use CRT to integrate each user's dimension of power consumption data into one composite data, which is described in Section V(B).

Definition 3 (Chinese Remainder Theorem): Assume that $\{d_1, d_2, ..., d_l\}$ are l integers and $\{q_1, q_2, ..., q_l\}$ are pairwise relatively prime positive integers. Then, the system of congruences $m \equiv d_k \ mod \ q_k$ for $1 \leq k \leq l$, has a unique solution

$$m \equiv d_1 Q_1 y_1 + d_2 Q_2 y_2 + ... + d_l Q_l y_l \ mod \ Q, \tag{6}$$

where $Q = \prod_{k=1}^{l} q_k$, $Q_k = \frac{Q}{q_k}$ and $y_k \cdot (\frac{Q}{q_k}) = 1 \ mod \ q_k$ for $1 \leq k \leq l$. Therefore, l integers $\{d_1, d_2, ..., d_l\}$ can be integrated into one composite integer m.

V. The PPMM-DA Scheme

In this section, we expound the PPMM-DA scheme, which includes the following phases: system initialization, user report generation, privacy-preserving report aggregation and secure report reading, the flowchart of PPMM-DA scheme is shown in Figure 2. Then, we apply differential privacy to enhanced this scheme to against differential attack in the next section.

A. System Initialization

Similar to [56], the power consumption range is classified into s continuous subsets $[R_1, R_2), [R_2, R_3), ..., [R_s, R_{s+1})$, where $R_1 = 0$, $R_{s+1} = E$ and $m_i \in [R_j, R_{j+1})$ $(j \in [1, s])$ indicates that m_i is equal or greater than R_j but less than R_{j+1}, where E denotes the maximum power consumption.

1) Step 1: The TA selects system parameter κ and two large primes p, q that satisfy $|p| = |q| = \kappa$, and gets a bilinear mapping tuple $(\mathbb{G}_1, \ \mathbb{G}_2, \ g, \ N = pq, \ e)$ by running the algorithm $Gen(\kappa)$. The public key of Paillier encryption system is (N, g) and the corresponding private key is (λ, μ).

2) Step 2: The TA generates a random number $\theta \in \mathbb{Z}_N$, and calculates

$$\theta + x_{fn} + x_0 = 0 \ mod \ \lambda. \tag{7}$$

To share the secret key θ with n users under the coverage area of FN, the TA distributes sub-secret keys to all users with a

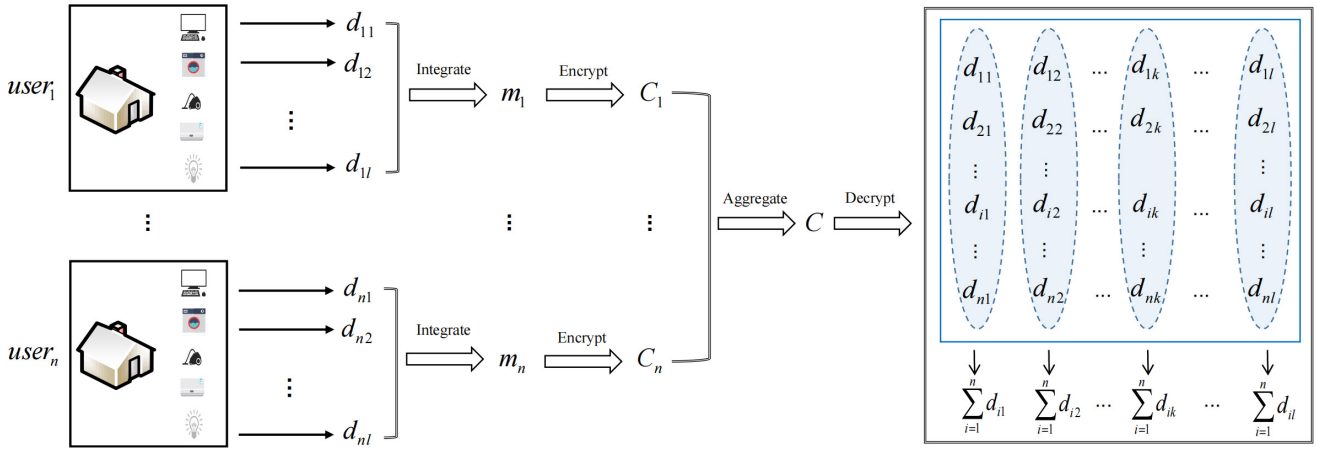


Fig. 2: Flowchart of the PPMM-DA scheme.

polynomial function of degree d as

$$G(x) = \theta + \alpha_1 x + \alpha_2 x^2 + ... + \alpha_d x^d \tag{8}$$

through employing Shamir's Secret Sharing [46]. Then, the TA computes $G(i)$, distributes $G(i)$ to $user_i$ as his private key and computes $user_i$'s public key $Y_i = g^{G(i)}$. In the secure report reading phase, θ can be recovered when the message is held by $d+1$ $(d+1 \leq n)$ or more participants. Lastly, the TA forwards x_0 to CC, sends x_{fn} to FN through a secure channel and computes the public key of FN as $Y_{fn} = g^{x_{fn}}$.

3) Step 3: Considering that there are l different kinds of devices in the fog-based smart grid system, which means that $user_i$'s power consumption data is l dimension $\{d_{i1}, d_{i2}, ..., d_{il}\}$, where each dimensional data $d_{il} \leq X$. To be noted, X denotes the maximum value of power consumption data of each dimension. Then, the TA chooses l prime numbers $\{q_1,\ q_2, ..., q_l\}$ and computes

$$\begin{cases} Q = \prod_{k=1}^{l} q_k, \\ Q_k = \dfrac{Q}{q_k},\ y_k \cdot (\dfrac{Q}{q_k}) = 1\ mod\ q_k, \\ \nu_k = Q_k \cdot y_k. \end{cases} \tag{9}$$

4) Step 4: The TA generates a set of super-increasing sequence $\{b_1, b_2, ..., b_s\}$, which satisfies

$$\begin{cases} b_i > b_0 + b_{i-1} \cdot n, \\ b_0 > n \cdot X \cdot \sum_{k=1}^{l} \nu_k. \end{cases} \tag{10}$$

In addition, the TA selects two hash functions $H : \{0,1\}^* \rightarrow \mathbb{Z}_N^*$ and $H_1 : \{0,1\}^* \rightarrow \mathbb{G}_1$, and sends $\{b_1, b_2, ..., b_s\}$ to CC.

5) Step 5: For $k \in [1, l]$, the TA publishes parameters $\{\mathbb{G}_1, \mathbb{G}_2, e, g, N, H, H_1, q_k, \nu_k, (R_1, ..., R_{s+1}), (g^{b_1}, ..., g^{b_s})\}$.

B. User Report Generation

In this phase, assuming that there are n users in a FN covered area, each $user_i$ $(i \in [1, n])$ collects and encrypts l dimensional power consumption data $\{d_{i1}, d_{i2}, ..., d_{il}\}$, and reports them to the corresponding FN periodically, e.g., every 15 minutes. The specific steps as follows:

1) Step 1: $user_i$ periodically collects power consumption data $\{d_{i1}, d_{i2}, ..., d_{il}\}$, and computes

$$m_i^{'} = d_{i1} + d_{i2} + ... + d_{il}. \tag{11}$$

If $user_i$'s total power consumption data $m_i^{'} \in [R_j, R_{j+1})$, $user_i$ is denoted to lie in the subset U_j. Then $user_i$ integrates these power consumption data $\{d_{i1}, d_{i2}, ..., d_{il}\}$ by computing

$$m_i = \sum_{k=1}^{l} d_{ik} \cdot \nu_k. \tag{12}$$

Next, $user_i$ first computes $\varrho_i = \prod_{t=1, t\neq i}^{n} \frac{t}{t-i}$ and the hash value $H(T)$, where T denotes the current timestamp, then utilizes the private key $G(i)$ and g^{b_j} to compute ciphertext

$$C_i = g^{m_i} \cdot g^{b_j} \cdot H(T)^{G(i) \cdot \varrho_i \cdot N}\ mod\ N^2. \tag{13}$$

2) Step 2: $user_i$ generates a signature σ_i with $G(i)$ as

$$\sigma_i = H(C_i || RA || T)^{G(i)}, \tag{14}$$

where RA represents the residential area.

3) Step 3: Finally, $user_i$ reports $< C_i || RA || T || \sigma_i >$ to the corresponding FN.

C. Privacy-preserving Report Aggregation

After FN receiving $n^{'}$ reports $< C_i || RA || T || \sigma_i >$ from $user_i$ $(i \in [1, n^{'}])$, where $n^{'}$ represents the number of users working normally. Then, the FN performs the following steps:

1) Step 1: The FN firstly checks the timestamp T, then verifies $n^{'}$ signatures by

$$e(g, \prod_{i=1}^{n^{'}} \sigma_i) = \prod_{i=1}^{n^{'}} e(Y_i, H(C_i || RA || T)). \tag{15}$$

2) Step 2: After successfully verifying users' signatures, the FN aggregates all ciphertexts to obtain the aggregated ciphertext.

If $d+1 \leq n^{'} \leq n$, the FN computes

$$\begin{aligned} C =& \prod_{i=1}^{n^{'}} C_i \cdot H(T)^{x_{fn}N} \\ =& g^{\sum_{i=1}^{n^{'}} m_i} \cdot g^{\sum_{j=1}^{s} b_j |U_j|}. \\ & H(T)^{(\sum_{i=1}^{n^{'}} G(i)\cdot \varrho_i + x_{fn})\cdot N} \ mod\ N^2. \end{aligned} \tag{16}$$

Note that, if $n^{'} < d+1$, the FN requests that TA randomly selects the private key $G(i)$ of $d+1-n^{'}$ users $\hat{U}$ and sends them via a secure channel. Subsequently, the FN calculates $\sum_{i\in\hat{U}} G(i)\cdot\varrho_i$ and

$$\begin{aligned} C =& \prod_{i=1}^{n^{'}} C_i \cdot H(T)^{(x_{fn}+\sum_{i\in\hat{U}} G(i)\cdot\varrho_i)N} \\ =& g^{\sum_{i=1}^{n^{'}} m_i} \cdot g^{\sum_{j=1}^{s} b_j |U_j|}. \\ & H(T)^{(\sum_{i=1}^{n^{'}} G(i)\cdot\varrho_i + x_{fn} + \sum_{i\in\hat{U}} G(i)\cdot\varrho_i)N} \ mod\ N^2. \end{aligned} \tag{17}$$

3) Step 3: The FN generates signature σ_{fn} with the private key x_{fn} as

$$\sigma_{fn} = H(C||RA||T)^{x_{fn}}. \tag{18}$$

Finally, the FN sends report $< C||RA||T||\sigma_{fn} >$ to CC.

D. Secure Report Reading

Upon receiving $< C||RA||T||\sigma_{fn} >$ from FN, the CC verifies the integrity and source authentication of aggregated ciphertext, uses Paillier decryption algorithm and CRT to process C. The detailed process is as follows:

1) Step 1: The CC first examines timestamp T, then computes and verifies signature by

$$e(g, \sigma_{fn}) = e(Y_{fn}, H(C||RA||T)). \tag{19}$$

2) Step 2: After successfully verifying the signature of FN, the CC uses its private key x_0 to compute

$$\begin{aligned} C^{'} =& C \cdot H(T)^{x_0 \cdot N} \ mod\ N^2 \\ =& g^{\sum_{i=1}^{n^{'}} m_i} \cdot g^{\sum_{j=1}^{s} b_j |U_j|}. \\ & H(T)^{(\sum_{i=1}^{n^{'}} G(i)\cdot\varrho_i + x_{fn} + x_0)\cdot N} \ mod\ N^2 \\ & \xrightarrow{\sum_{i=1}^{n^{'}} G(i)\cdot\varrho_i + x_{fn} + x_0 = 0 \ mod\ \lambda} \\ =& g^{\sum_{i=1}^{n^{'}} m_i} \cdot g^{\sum_{j=1}^{s} b_j |U_j|} \ mod\ N^2 \\ =& g^{\sum_{k=1}^{l} (\sum_{i=1}^{n^{'}} d_{ik}\cdot\nu_k)} \cdot g^{\sum_{j=1}^{s} b_j |U_j|} \ mod\ N^2 \\ =& g^{V} \ mod\ N^2, \end{aligned} \tag{20}$$

where $V = \sum_{k=1}^{l} (\sum_{i=1}^{n^{'}} d_{ik}\cdot\nu_k) + (\sum_{j=1}^{s} b_j |U_j|)$. The CC can recover V as follows:

$$V = \frac{C^{'} - 1}{N} \ mod\ N^2. \tag{21}$$

3) Step 3: By executing the Algorithm 1, the CC can recover $|U_j|$ $(j \in [1, s])$ and $M^{'}$ from V, where $|U_j|$ denotes the number of users in the j-th subset, and

$$M^{'} = \sum_{i=1}^{n^{'}} m_i \ mod\ Q = \sum_{k=1}^{l} (\sum_{i=1}^{n^{'}} d_{ik}\cdot\nu_k) \ mod\ Q. \tag{22}$$

4) Step 4: Based on the CRT, the CC can obtain

$$\begin{aligned} M_k =& M^{'} \ mod\ q_k \\ =& \sum_{k=1}^{l} (\sum_{i=1}^{n^{'}} d_{ik}\cdot\nu_k) \ mod\ q_k \\ =& \sum_{i=1}^{n^{'}} d_{ik}, \qquad (k \in [1, l]) \end{aligned} \tag{23}$$

where M_k is the sum of power consumption data of the k-th dimension in the area covered by FN.

Algorithm 1: The algorithm to recover the number of users in each subset and $M^{'}$

Input: b_j $(j \in [1, s])$ and V;
Output: $|U_j|, M^{'}$;
for $j = s; j \geq 1; j--$ **do**
  $|U_j| = \frac{V - V \ mod\ b_j}{b_j}$;
  $V = V - (b_j \cdot |U_j|)$;
$M^{'} = V \ mod\ Q$;
Return $\{|U_1|, |U_2|, ..., |U_s|, M^{'}\}$;
End procedure

VI. Privacy-enhanced Scheme With Differential Privacy Technique

In PPMM-DA scheme, although users' multi-dimensional power consumption data are encrypted by Paillier cryptosystem, the adversary still can launch differential attack to threaten their privacy. For example, the adversary launches two queries on two datasets D_k and $D_k^{'}$, which represent the k-th dimensional datasets and these two datasets differing on $user_i$'s data. Let A be a sum aggregation query operation, the corresponding results are $A(D_k)$ and $A(D_k^{'}) = A(D_k) + d_{ik}$, and it is possible for the adversary to gain $user_i$'s k-th dimensional power consumption data d_{ik} by computing $A(D_k) - A(D_k^{'})$. To avoid device-specific power consumption data of individual user leaking, we use differential privacy technique [44] to enhance privacy. Moreover, each dimension of power consumption data under the FN coverage area is regarded as an independent set, and the power consumption data in different sets have no intersection, which satisfies the parallel composability of differential privacy mentioned in Section IV(B). In order to realize privacy-enhanced multi-dimensional data aggregation, we propose two methods to achieve differential privacy, which extracted noises from Geometric distribution and Laplace distribution, respectively.

A. Geometric Distribution

Applying Geometric distribution to generate noises was first put forward by Ghosh et al. [50]. Specifically, noises are chosen from a symmetric Geometric distribution $Geom(\alpha)$ with $0 < \alpha < 1$, where α can be seen as a discrete approximation of Laplace distribution $Lap(\tilde{\lambda})$, i.e., $\alpha \approx exp(-\frac{1}{\tilde{\lambda}})$. The probability density function of Geometric distribution $Geom(\alpha)$ is

$$Pr[x] = \frac{1-\alpha}{1+\alpha}\alpha^{|x|}. \tag{24}$$

Formally, the sensitivity of aggregation function A is

$$\triangle A = max_{D,D'}||A(D) - A(D^{'})||_1, \tag{25}$$

for all datasets D and $D^{'}$ differing on at most one element, then by adding geometric noises r_1, r_2, which are randomly chosen from $Geom(exp(-\frac{\epsilon}{\triangle A}))$, to the original aggregated result, the perturbed aggregated result can achieve ϵ-differential privacy, i.e., for any integer $O \in Range(A)$,

$$Pr[A(D) + r_1 = O] \leq exp(\epsilon) \cdot Pr[A(D^{'}) + r_2 = O]. \tag{26}$$

The specific steps, which extracted noise from Geometric distribution to achieve differential privacy, are as follows:

1) Step 1: As shown in the user report generation phase, each $user_i$ computes C_i and σ_i. Then, $user_i$ sends report $< C_i||RA||T||\sigma_i >$ to FN.

2) Step 2: The FN calculates the aggregated data $\tilde{C}$ as follows. Due to $A(D_k) = \sum_{i=1}^{n^{'}} d_{ik}$, which D_k represents the k-th dimensional dataset in the area covered by FN, then $|A(D_k) - A(D_k^{'})| \leq X$ holds for any two datasets D_k and $D_k^{'}$ differing on at most one element.

Therefore, we can set $\triangle A_k = X$. The FN chooses random noises $\tilde{d}_k$ $(k \in [1, l])$ from $Geom(exp(-\frac{\epsilon_k}{\triangle A_k}))$ to implicitly add them to C.

$$\begin{aligned}
\tilde{C} =& C \cdot g^{\sum_{k=1}^{l}(\tilde{d}_k \cdot \nu_k)} \\
=& g^{\sum_{i=1}^{n^{'}}(\sum_{k=1}^{l} d_{ik} \cdot \nu_k) + \sum_{k=1}^{l}(\tilde{d}_k \cdot \nu_k)} \cdot g^{\sum_{j=1}^{s} b_j |U_j|} \cdot \\
& H(T)^{(\sum_{i=1}^{n^{'}} G(i)\varrho_i + x_{fn})N} \ mod \ N^2 \\
=& g^{\sum_{k=1}^{l}((\sum_{i=1}^{n^{'}} d_{ik} + \tilde{d}_k) \cdot \nu_k)} \cdot g^{\sum_{j=1}^{s} b_j |U_j|} \cdot \\
& H(T)^{(\sum_{i=1}^{n^{'}} G(i)\varrho_i + x_{fn})N} \ mod \ N^2 \\
=& g^{\sum_{k=1}^{l}(\tilde{M}_k \cdot \nu_k)} \cdot g^{\sum_{j=1}^{s} b_j |U_j|} \cdot \\
& H(T)^{(\sum_{i=1}^{n^{'}} G(i)\varrho_i + x_{fn})N} \ mod \ N^2.
\end{aligned} \tag{27}$$

3) Step 3: The FN generates signature σ_{fn} and returns report $< \tilde{C}||RA||T||\sigma_{fn} >$ to CC.

4) Step 4: Upon receiving the report $< \tilde{C}||RA||T||\sigma_{fn} >$ from FN, the CC verifies the signature of FN and decrypts the aggregated data $\tilde{C}$ as shown in the secure report reading phase. Finally, the CC obtains the sum of power consumption data of each dimension after perturbing $\tilde{M}_k = \sum_{i=1}^{n^{'}} d_{ik} + \tilde{d}_k$ $(k \in [1, l])$ and the number of users $|U_1|, |U_2|, ..., |U_s|$ in each subset, where $\tilde{M}_k$ is the sum of power consumption data of k-th dimension after perturbing.

B. Laplace Distribution

The Laplace distribution has been widely used to achieve ϵ-differential privacy by adding Laplace noise Lap($\tilde{\lambda}$) to the output of a query. The noise Lap($\tilde{\lambda}$) is sampled from Laplace distribution, whose probability density function is

$$Pr[x] = \frac{1}{2\tilde{\lambda}} exp(-\frac{|x|}{\tilde{\lambda}}). \tag{28}$$

We assume that the differential privacy aggregation function A answers a query on two datesets D and $D^{'}$, which are different on one single element. Then, we have

$$\begin{aligned}
\frac{Pr[A(D)=O]}{Pr[A(D^{'})=O]} &= \frac{\frac{1}{2\tilde{\lambda}} exp(-\frac{|O-A(D)|}{\tilde{\lambda}})}{\frac{1}{2\tilde{\lambda}} exp(-\frac{|O-A(D^{'})|}{\tilde{\lambda}})} \\
&\leq exp(\frac{|A(D)-A(D^{'})|}{\tilde{\lambda}}) \leq exp(\frac{\triangle A}{\tilde{\lambda}}).
\end{aligned} \tag{29}$$

Here ΔA denotes the global sensitivity of A, which is the maximum change of A between two neighboring datasets D and $D^{'}$, that is, $\Delta A = max_{D \simeq D^{'}} |A(D) - A(D^{'})|$, where $D \simeq D^{'}$ denotes that D and $D^{'}$ are neighboring. Let $\epsilon = \frac{\Delta A}{\tilde{\lambda}}$, we have $Pr[A(D) \in O] \leq exp(\epsilon) \cdot Pr[A(D^{'}) \in O]$, i.e., adding Laplace noise $Lap(\tilde{\lambda})$ to a query result for achieving ϵ-differential privacy, where $\tilde{\lambda}$ denotes noise scale and $\tilde{\lambda} = \frac{\Delta A}{\epsilon}$. Furthermore, the distribution of $Lap(\tilde{\lambda})$ is infinitely divisible. Specially, for every integer $\xi \geq 1$,

$$Lap(\tilde{\lambda}) = \sum_{i=1}^{\xi} G_i(\xi, \tilde{\lambda}) - \sum_{i=1}^{\xi} G_i^{'}(\xi, \tilde{\lambda}), \tag{30}$$

where $G_i(\xi, \tilde{\lambda})$ and $G_i^{'}(\xi, \tilde{\lambda})$ are two random variables having Gamma distribution with probability density function:

$$Pr(x, \xi, \tilde{\lambda}) = \frac{\frac{1}{\tilde{\lambda}}^{1/\xi}}{\Gamma(\frac{1}{\xi})} x^{\frac{1}{\xi}-1} exp(-\frac{x}{\tilde{\lambda}}), \tag{31}$$

where $x > 0$ and $\Gamma(\frac{1}{\xi})$ is the Gamma function evaluated at $1/\xi$. In our scheme, A_k is a aggregation function which calculates the sum of power consumption of k-th dimension in the area covered by FN, and ΔA_k is the maximum power consumption change of A_k in the k-th dimension. If the number of users is n, for each $user_i$, we can add $G_{ik}(n, \tilde{\lambda}_k) - G_{ik}^{'}(n, \tilde{\lambda}_k)$ to its measurement d_{ik} before reporting, where $\tilde{\lambda}_k = \frac{\Delta A_k}{\epsilon_k}$. Thus, the sum of power consumption of k-th dimension is

$$\begin{aligned}
&\sum_{i=1}^{n} d_{ik} + \sum_{i=1}^{n} (G_{ik}(n, \tilde{\lambda}_k) - G_{ik}^{'}(n, \tilde{\lambda}_k)) \\
=& \sum_{i=1}^{n} d_{ik} + Lap(\tilde{\lambda}_k) \qquad (k \in [1, l]).
\end{aligned} \tag{32}$$

In this way, ϵ_k-differential privacy is satisfied.

The process of extracting noise from Laplace distribution to achieve differential privacy is as follows:

1) Step 1: Each $user_i$ generates Laplace noise $\tilde{d}_{ik} = G_{ik}(n^{'}, \tilde{\lambda}_k) - G_{ik}^{'}(n^{'}, \tilde{\lambda}_k)$ $(i \in [1, n^{'}], k \in [1, l])$, which is random value independently sampled from the Gamma distribution, and adds this noise to d_{ik}. The perturbed power consumption data can be expressed as $\tilde{m}_i = \sum_{k=1}^{l}(d_{ik} + \tilde{d}_{ik}) \cdot \nu_k$. Next,

the $user_i$ computes $\tilde{C}_i = g^{\tilde{m}_i} \cdot g^{b_j} \cdot H(T)^{G(i)\cdot\varrho_i \cdot N} \ mod\ N^2$ and σ_i following the same processes as shown in the user report generation phase. Then, the $user_i$ reports $< \tilde{C}_i || RA || T || \sigma_i >$ to FN.

2) Step 2: The FN verifies signatures of $n^{'}$ users following the same processes as shown in the privacy-preserving report aggregation phase. After verifying users' signatures, the FN computes the aggregated data $\tilde{C}$. Especially, we let $\tilde{d}_k = \sum_{i=1}^{n^{'}} \tilde{d}_{ik}$ and the FN aggregates all the received ciphertexts $\tilde{C}_i$ $(i \in [1, n^{'}])$ as

$$\begin{aligned}\tilde{C} =& \prod_{i=1}^{n^{'}} \tilde{C}_i \cdot H(T)^{x_{fn}N} \\ =& g^{\sum_{i=1}^{n^{'}} \sum_{k=1}^{l} (d_{ik}+\tilde{d}_{ik})\cdot\nu_k} \cdot g^{\sum_{j=1}^{s} b_j |U_j|} \cdot \\ & H(T)^{(\sum_{i=1}^{n^{'}} G(i)\cdot\varrho_i + x_{fn})\cdot N} \ mod\ N^2 \\ =& g^{\sum_{k=1}^{l} ((\sum_{i=1}^{n^{'}} d_{ik}+\tilde{d}_k)\cdot\nu_k)} \cdot g^{\sum_{j=1}^{s} b_j |U_j|} \cdot \\ & H(T)^{(\sum_{i=1}^{n^{'}} G(i)\cdot\varrho_i + x_{fn})\cdot N} \ mod\ N^2 \\ =& g^{\sum_{k=1}^{l} \tilde{M}_k \cdot \nu_k} \cdot g^{\sum_{j=1}^{s} b_j |U_j|} \cdot \\ & H(T)^{(\sum_{i=1}^{n^{'}} G(i)\cdot\varrho_i + x_{fn})\cdot N} \ mod\ N^2.\end{aligned} \tag{33}$$

3) Step 3: The FN generates signature σ_{fn} and returns report $< \tilde{C} || RA || T || \sigma_{fn} >$ to CC following the same processes as shown in the privacy-preserving report aggregation phase.

4) Step 4: Upon receiving the report $< \tilde{C} || RA || T || \sigma_{fn} >$ from FN, the CC verifies the signature of FN and decrypts the aggregated data $\tilde{C}$ following the same processes as shown in the secure report reading phase. Finally, the CC obtains the sum of power consumption of each dimension after perturbing $\tilde{M}_k = \sum_{i=1}^{n^{'}} d_{ik} + \tilde{d}_k$ and the number of users $|U_1|, |U_2|, ..., |U_s|$ in each subset.

VII. CORRECTNESS AND SECURITY ANALYSIS

A. Proof of Correctness

1) Formula (20): With the CC's private key x_0 as well as the Lagrange interpolation polynomial [46], this formula eliminates the term containing $H(T)$ in C. We assume that

$$V = \sum_{i=1}^{n^{'}} m_i + \sum_{j=1}^{s} b_j |U_j| \tag{34}$$

and obtain

$$C = g^V \cdot H(T)^{(\sum_{i=1}^{n^{'}} G(i)\cdot\varrho_i + x_{fn})\cdot N} \ mod\ N^2. \tag{35}$$

According to the Lagrange interpolation polynomial, we have

$$G(x) = \sum_{i=1}^{d+1} (\prod_{t=1, t\neq i}^{d+1} \frac{t-x}{t-i}) G(i). \tag{36}$$

Therefore, when $d+1 \leq n^{'} \leq n$,

$$\sum_{i=1}^{n^{'}} G(i)\varrho_i = \sum_{i=1}^{d+1} (\prod_{t=1, t\neq i}^{d+1} \frac{t-0}{t-i}) G(i) = G(0) = \theta, \tag{37}$$

$$\sum_{i=1}^{n^{'}} G(i)\varrho_i + x_{fn} + x_0 = 0 \ mod\ \lambda. \tag{38}$$

We assume $\theta + x_{fn} + x_0 = \beta \cdot \lambda$, and then we can get

$$\begin{aligned} C^{'} = C \cdot H(T)^{x_0 \cdot N} &= g^V \cdot H(T)^{\beta\cdot\lambda\cdot N} \ mod\ N^2 \\ &= g^V \cdot (H(T)^{\beta})^{N\cdot\lambda} \ mod\ N^2 \\ &\xrightarrow{r = H(T)^{\beta}} \\ &= g^V \cdot r^{N\cdot\lambda} \ mod\ N^2. \end{aligned} \tag{39}$$

Therefore, we have $C^{'} = g^V \ mod\ N^2$.

2) Formula (21): From the description in our preliminaries, we can get

$$\begin{aligned} C^{'} &= g^V \ mod\ N^2 \\ &= (1+N)^V \ mod\ N^2 \\ &= 1 + N \cdot V \ mod\ N^2 \\ \Longrightarrow V &= \frac{C^{'} - 1}{N} \ mod\ N^2. \end{aligned} \tag{40}$$

3) Algorithm 1: This algorithm has one for loop, which recovers $|U_j|$ and $M^{'}$. Since

$$\sum_{i=1}^{n^{'}} m_i = \sum_{i=1}^{n^{'}} \sum_{k=1}^{l} d_{ik} \cdot \nu_k < \sum_{i=1}^{n^{'}} \sum_{k=1}^{l} X \cdot \nu_k < b_0 \tag{41}$$

and

$$\begin{aligned} & b_1 \cdot |U_1| + b_2 \cdot |U_2| + ... + b_{s-1} \cdot |U_{s-1}| \\ & < b_{s-1} \cdot |U_1| + b_{s-1} \cdot |U_2| + ... + b_{s-1} \cdot |U_{s-1}| \\ & \leq b_{s-1} \cdot n, \end{aligned} \tag{42}$$

we can get

$$\begin{aligned} & \sum_{i=1}^{n^{'}} m_i + b_1 \cdot |U_1| + b_2 \cdot |U_2| + ... + b_{s-1} \cdot |U_{s-1}| \\ & < b_0 + b_{s-1} \cdot n < b_s. \end{aligned} \tag{43}$$

Therefore, we can obtain $(V - V \ mod\ b_s)/b_s = (b_s \cdot |U_s|)/b_s = |U_s|$, and by using the same method, we can recover all $|U_1|, |U_2|, ..., |U_s|$. Then, we can obtain

$$M^{'} = \sum_{i=1}^{n^{'}} m_i \ mod\ Q = \sum_{k=1}^{l} (\sum_{i=1}^{n^{'}} d_{ik} \cdot \nu_k) \ mod\ Q, \tag{44}$$

$$M_k = M^{'} \ mod\ q_k = \sum_{i=1}^{n^{'}} d_{ik} \quad (k \in [1, l]). \tag{45}$$

B. Security Analysis

In this subsection, we analyze and prove that our proposed scheme can achieve all security goals, that is, our scheme is secure against the differential attack, eavesdropping attack, collusion attack and active attack.

1) The privacy of individual user's power consumption data is protected from differential attack: According to the threat model mentioned above, the adversary may intend to learn the private power consumption data of individual user through differential attack. Our proposed schene can guarantee differential privacy for avoiding leakage of fine-grained private power consumption data. On the one hand, with the differential privacy technique, we demonstrate that the aggregated data for k-th dimension, which adds noise extracted from Geometric distribution, can achieve ϵ_k-differential privacy. For example, assuming the adversary obtains two perturbed aggregated data $A(D_k) = v + \tilde{d}_k$ and $A(D_k^{'}) = w + \tilde{d}_k^{'}$, where v and w are two adjacent aggregation, $\tilde{d}_k$ and $\tilde{d}_k^{'}$ are the corresponding geometric noises from $Geom(exp(-\frac{\epsilon_k}{\triangle A_k}))$. Since $|v-w| \leq X$, for integer O, we have

$$\begin{aligned}\eta &= \frac{Pr[v+\tilde{d}_k = O]}{Pr[w+\tilde{d}_k^{'} = O]} = \frac{Pr[\tilde{d}_k = O-v]}{Pr[\tilde{d}_k^{'} = O-w]} \\ &= \frac{\frac{1-\alpha}{1+\alpha}\cdot\alpha^{|O-v|}}{\frac{1-\alpha}{1+\alpha}\cdot\alpha^{|O-w|}} = \alpha^{|O-v|-|O-w|}.\end{aligned} \tag{46}$$

Because

$$-|v-w| \leq |O-v| - |O-w| \leq |v-w|, \tag{47}$$

and $0 < \alpha < 1$, $\alpha \approx exp(-\frac{\epsilon_k}{X})$, we can obtain

$$\begin{aligned}\alpha^X \leq \alpha^{|v-w|} \leq \eta \leq \alpha^{-|v-w|} \leq \alpha^{-X} \\ (exp(-\frac{\epsilon_k}{X}))^X \leq \eta \leq (exp(-\frac{\epsilon_k}{X}))^{-X} \\ exp(-\epsilon_k) \leq \eta \leq exp(\epsilon_k).\end{aligned} \tag{48}$$

On the other hand, we demonstrate that the aggregated data for k-th dimension, which adds noise extracted from the Laplace distribution, can also achieve ϵ_k-differential privacy. For example, assuming the adversary obtains two perturbed aggregated data $A(D_k) = v + \tilde{d}_k$ and $A(D_k^{'}) = w + \tilde{d}_k^{'}$, where v and w are two adjacent aggregation, $\tilde{d}_k$ and $\tilde{d}_k^{'}$ are the corresponding noises from Laplace distribution with the scale of $\tilde{\lambda}_k = \frac{\Delta A_k}{\epsilon_k}$. Therefore, we have

$$\begin{aligned}\frac{Pr[A(D_k)=O]}{Pr[A(D_k^{'})=O]} &= \frac{\frac{1}{2\tilde{\lambda}_k}exp(-\frac{|O-A(D_k)|}{\tilde{\lambda}_k})}{\frac{1}{2\tilde{\lambda}_k}exp(-\frac{|O-A(D_k^{'})|}{\tilde{\lambda}_k})} \\ &= exp(\frac{|A(D_k)-A(D_k^{'})|}{\tilde{\lambda}_k}) \\ &\leq exp(\frac{\Delta A_k}{\tilde{\lambda}_k}) = exp(\epsilon_k).\end{aligned} \tag{49}$$

In both methods, converting the perturbed power consumption data into ciphertext $\tilde{C}_i$ can also satisfy ϵ_k-differential privacy. According to the differential privacy axiom proposed by Daniel Kifer in [47], differential invariance is defined: transformation invariance, which shows dataset that satisfies ϵ-differential privacy can satisfy ϵ-differential privacy after encryption. Therefore, $\tilde{M}_k$ achieves ϵ_k-differential privacy in these two methods. Additionally, we regard each dimension of power consumption data as an independent set, and the power consumption data in different sets have no intersection, which satisfies the parallel composability of differential privacy mentioned in Section IV. Consequently, the perturbed aggregated result of multi-dimensional data after adding noise satisfies $max\{\epsilon_1, \epsilon_2, ..., \epsilon_l\}$-differential privacy.

2) The privacy of individual user's power consumption data is protected from eavesdropping attack: The adversary may obtain the private power consumption data of individual user by eavesdropping on communication links, such as the communication links from users to FN or from FN to CC. Due to users' private power consumption data are transmitted on each communication link in ciphertext, even if the adversary obtains C_i or C, he cannot obtain private power consumption data of individual user. On the one hand, we consider an adversary who has obtained a report by eavesdropping on the communication between users and FN and obtains $< C_i||RA||T||\sigma_i >$, where $C_i = g^{m_i}\cdot g^{b_j}\cdot H(T)^{G(i)\cdot\varrho_i\cdot N}\ mod\ N^2$. Let $m = m_i + b_j$ and $r_i = H(T)^{G(x_i)\cdot\varrho_i}$, then ciphertext $C_i = g^m \cdot r_i^N\ mod\ N^2$ is still a legal ciphertext of Paillier cryptosystem. Since Paillier cryptosystem is indistinguishable under the chosen plaintext attack (IND-CPA) secure, the adversary cannot decrypt $user_i$'s ciphertext C_i and get his private data m_i, let alone obtain $user_i$'s fine-grained data, such as the power consumption data of k-th dimension d_{ik}.

On the other hand, we consider that the adversary has obtained the report by eavesdropping on the communication from FN to CC. However, he can only get the aggregated data and ciphertexts of all users' power consumption data. Similarly, the FN aggregates these ciphertexts and computes the aggregated ciphertext C, which has the same form as individual report C_i. Since C is still a valid ciphertext of Paillier cryptosystem, and $\sum_{i=1}^{n^{'}} m_i$ and d_{ik} are transparent to the adversary. Therefore, our proposed scheme can protect the private power consumption data of individual user from eavesdropping attack.

3) The privacy of individual user's power consumption data is protected from collusion attack: On the one hand, if the FN colludes with some users, they can obtain their secret parameters $\{G(i), \varrho_i\}$. In our scheme, the TA employs the Shamir's Secret Sharing and computes a polynomial function of degree d, which is

$$G(x) = \theta + \alpha_1 x + \alpha_2 x^2 + ... + \alpha_d x^d, \tag{50}$$

and then divides θ into n shares $G(i)$ to n users as their private keys, and any d users or fewer than d users cannot obtain θ. Supposing an extreme situation occurs, the adversary successfully compromises $d+1$ users and obtains their corresponding private keys $\{G(1), G(2), ..., G(d+1)\}$. However, due to

$$\theta + x_{fn} + x_0 = 0\ mod\ \lambda, \tag{51}$$

and private key x_0 is kept secretly by CC, even if the FN colludes with $d+1$ users, x_0 cannot be recovered. Therefore, we can conclude that no matter how many users collude with FN, they cannot obtain the private data of other users. On the other hand, the private key of Paillier cryptosystem λ is kept secretly by TA, and even if the CC colludes with FN or some users by sharing and analyzing their information (such as ciphertexts, private keys and public information), they cannot obtain the private data of other users. Hence, we can conclude that our proposed scheme can protect the privacy of individual user's power consumption data from collusion attack.

4) The privacy of individual user's power consumption data is protected from active attack: It is important to protect users' private data from active attack, such as message forgery and replay attack. We will elaborate that the report can be authenticated by FN and CC, which is indeed sent by a legitimate entity and cannot be changed during transmission.

Source authentication. The FN needs to verify the identity of user before aggregating the data to ensure that the report was sent by the legitimate user who has not been tampered with. Similarly, the CC must perform identity verification before decrypting data. Therefore, our scheme can guarantee that the received message actually come from the legitimate entity and resist message forgery attack.

Data integrity and Replay attack. On the one hand, we consider the communication link from IoT layer to Fog layer. When $user_i$ sends a report $< C_i||RA||T||\sigma_i >$ to nearby FN, the FN verifies if $e(g, \sigma_i) = e(Y_i, H(C_i||RA||T))$ holds. Only legitimate $user_i$ can generate valid ciphertext $C_i = g^{m_i} \cdot g^{b_j} \cdot H(T)^{G(i)\cdot \varrho_i \cdot N} \ mod \ N^2$ and signature $\sigma_i = H(C_i||RA||T)^{G(i)}$ by $G(i)$ and timestamp T. At the same time, an external adversary cannot make any modifications on the encrypted data C_i and report $< C_i||RA||T||\sigma_i >$, which can be guaranteed by FN to detect whether the report has been tampered during transmission or not based on the equation $e(g, \sigma_i) = e(Y_i, H(C_i||RA||T))$. On the other hand, the data integrity of the report from FN to CC can also be achieved. The FN generates signature $\sigma_{fn} = H(C||RA||T)^{x_{fn}}$ with x_{fn} and timestamp T. In addition, the T is used in the report $< C||RA||T||\sigma_{fn} >$. After the FN and CC receive the report, they check T to detect replay attack. Therefore, our scheme can satisfy data integrity and resist message replay attack.

VIII. Performance Evaluation

In this section, we compare our proposed scheme with [13], [17], [24], [25], [31], [51]–[56] from functions, computational and communication overheads, and errors in the fog-based smart grid system.

A. Functional Comparison

The functional comparison is shown in Table II. As shown in Table II, our proposed scheme achieves multi-dimensional and multi-subset data aggregation, supports fault tolerance, resists differential attack, eavesdropping attack, collusion attack and active attack. The scheme [13] can against eavesdropping attack and active attack using Boneh-Goh-Nissim public key cryptography. In the scheme [17], the CC can only obtain aggregated results of multi-dimensional data. The scheme [31] combines differential privacy technology to resist differential attack for one-dimensional data, realizes fault tolerance and resists eavesdropping attack. The scheme [53] can against both differential and active attacks with a differential privacy mechanism. The scheme [51] achieves fault tolerance and resists differential attack via randomized responses. However, these schemes [31], [51], [53] can only obtain aggregated result of one-dimensional data and cannot perform multi-subset data aggregation. Some schemes [52], [54], [55] realize privacy-preserving multi-dimensional data aggregation via differential privacy and resisted differential attack. However, these schemes [52], [54], [55] cannot resist eavesdropping attack, collusion attack and active attack. In schemes [24], [25], the CC can perform the multi-subset data aggregation, but SMs only report one type of data to CC. Also, they cannot resist differential attack. Zuo et al. [56] come up with a privacy-preserving multi-dimensional and multi-subset data aggregation for addressing the collusion attack. However, this scheme also cannot resist differential attack.

B. The Comparison of Computational Overhead

In this subsection, we compare the computational overhead of our proposed scheme with these of [17], [24], [25], [56] in terms of each user, GW, CC and the total. As authentication is not considered in schemes [24], [25], we will not discuss the computational overhead of signature here. Let T_e be the time of exponential operation in $\mathbb{Z}_{N^2}$, and T_{mul} be the time of multiplication operation in $\mathbb{Z}_{N^2}$. Our experiment is conducted on a laptop with 64-bits Windows 10 Enterprise operating system, the Intel(R) Core(TM) i7-4510U CPU and 8 GB memory. The experiment results show that $T_e = 1.7$ ms and $T_{mul} = 0.16$ ms. In addition, we assume that there are n users and each user's power comsumption data is l dimension. The comparison of computational overhead for [17], [24], [25], [56] and ours is depicted in Table III.

Chen et al.'s scheme [17]: In the user report generation phase, each user requires $(l+1)T_e + lT_{mul}$ to generate the ciphertext. In the privacy-preserving report aggregation phase, GW requires $(n–1)T_{mul}$ to generate the aggregated ciphertext. In the secure report reading phase, the CC requires T_e to decrypt the aggregated data. Thus, the computational overhead of scheme [17] is $(nl+n+1)T_e+(nl+n-1)T_{mul}$ in total.

Li et al.'s scheme [24]: In the user report generation phase, each user requires $2T_e+T_{mul}$ to generate the ciphertext. In the privacy-preserving report aggregation phase, the GW requires $(n-1)T_{mul}$ to generate the aggregated ciphertext. In the secure report reading phase, the CC requires $T_e + T_{mul}$ to decrypt the aggregated data. Therefore, the computational overhead of scheme [24] is $(2n + 1)T_e + 2nT_{mul}$ in total.

Chien et al.'s scheme [25]: In the user report generation phase, each user requires $4T_e + 2T_{mul}$ to generate the ciphertext. In the privacy-preserving report aggregation phase, the GW requires $2(n - 1)T_{mul}$ to generate the aggregated ciphertext. In the secure report reading phase, the CC requires $2T_e$ to decrypt the aggregated data. Therefore, the computational overhead of scheme [25] is $(4n+2)T_e+(4n-2)T_{mul}$ in total.

Zuo et al.'s scheme [56]: In the user report generation phase, each user requires $3T_e+T_{mul}$ to generate the ciphertext. In the privacy-preserving report aggregation phase, the GW requires $2(n - 1)T_{mul}$ to generate the aggregated ciphertext. In the secure report reading phase, the CC requires nT_e+T_{mul} to decrypt the aggregated data. Therefore, the computational overhead of scheme [56] is $4nT_e + (3n - 1)T_{mul}$ in total.

Our proposed scheme: In the user report generation phase, each user requires $2T_e + T_{mul}$ to generate the ciphertext. Particularly, we structure multi-dimensional data as a composite data, which can reduce the computational overhead of encryption significantly. In the privacy-preserving report aggregation phase, the GW requires $(n - 1)T_{mul}$ to generate the aggregated ciphertext. In the secure report reading phase, the CC requires $T_e + T_{mul}$ to decrypt the aggregated data. Therefore, the computational overhead of our proposed scheme is $(2n + 1)T_e + 2nT_{mul}$ in total.

On the one hand, we compare our proposed scheme with [17], [24], [25], [56] for different number of dimensions l in Figure 3. Specifically, we vary the dimension of power consumption data from $\{5, 10, 15, 20, 25, 30, 35, 40, 45, 50\}$ and assume there are 100 users in the system. From the Figure 3, we can find that the computational overhead of scheme [17] increases linearly with l increases, while that of our proposed scheme is independent of l in the user report generation phase.

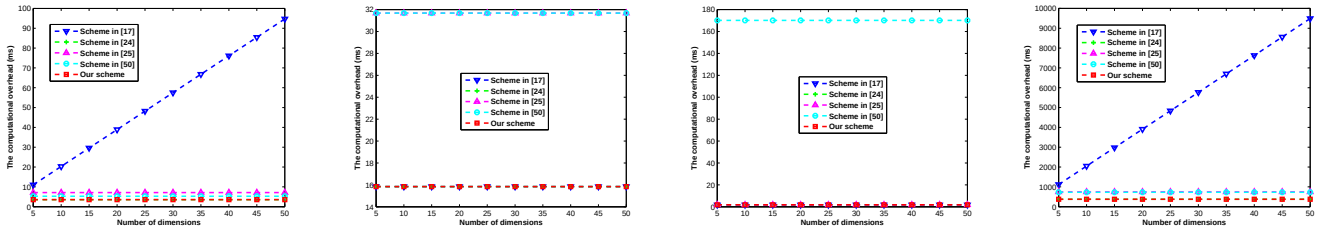
This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2023.3309132

TABLE II: Functional comparison

Feature	[13]	[17]	[24]	[25]	[31]	[51]	[52]	[53]	[54]	[55]	[56]	Ours
Multi-dimensional	×	√	×	×	×	×	√	×	√	√	√	√
Multi-subset	×	×	√	√	×	×	×	×	×	√	√	√
Multi-dimensional and multi-subset	×	×	×	×	×	×	×	×	×	√	√	√
Differential privacy	×	×	×	×	√	√	√	√	√	√	×	√
Fault tolerance	×	×	×	√	√	√	×	×	×	×	√	√
Eavesdropping attack	√	√	√	√	×	×	×	×	×	×	√	√
Collusion attack	×	×	√	×	×	×	×	×	×	×	√	√
Active attack	√	√	×	×	√	×	×	√	×	×	√	√

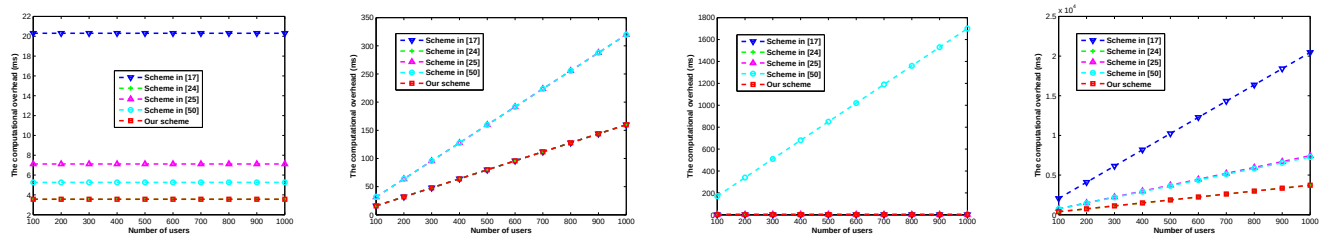
TABLE III: The comparison of computational overhead

Participant	Scheme in [17]	Scheme in [24]	Scheme in [25]	Scheme in [56]	Our scheme
Each user	$(l+1)T_e + lT_{mul}$	$2T_e + T_{mul}$	$4T_e + 2T_{mul}$	$3T_e + T_{mul}$	$2T_e + T_{mul}$
The GW	$(n-1)T_{mul}$	$(n-1)T_{mul}$	$2(n-1)T_{mul}$	$2(n-1)T_{mul}$	$(n-1)T_{mul}$
The CC	T_e	$T_e + T_{mul}$	$2T_e$	$nT_e + T_{mul}$	$T_e + T_{mul}$
The total	$(nl+n+1)T_e+(nl+n-1)T_{mul}$	$(2n+1)T_e+2nT_{mul}$	$(4n+2)T_e+(4n-2)T_{mul}$	$4nT_e+(3n-1)T_{mul}$	$(2n+1)T_e+2nT_{mul}$



(a) Each user (b) The GW (c) The CC (d) The total

Fig. 3: The variation of computational overheads for different number of dimensions when $n = 100$.



(a) Each user (b) The GW (c) The CC (d) The total

Fig. 4: The variation of computational overheads for different number of users when $l = 10$.

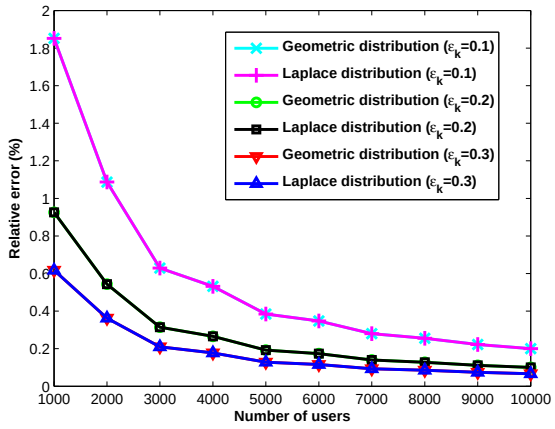


Fig. 5: The comparison of relative errors on Geometric distribution and Laplace distribution under different ϵ_k.

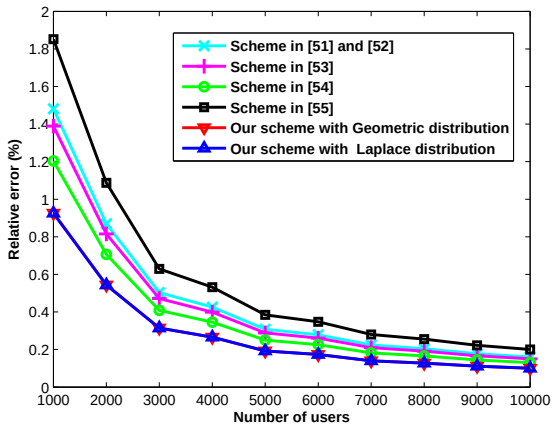


Fig. 6: The comparison of relative errors for schemes [51]–[55] and ours when $\epsilon_k = 0.2$.

On the other hand, we compare our proposed scheme with [17], [24], [25], [56] for different user numbers n in Figure 4. Specifically, we vary the the number of users from $\{100, 200, 300, 400, 500, 600, 700, 800, 900, 1000\}$ and assume each user's power comsumption data is 10 dimensions. Figure 4 demonstrates that the computational overhead of our proposed scheme is lower than these of [17], [25], [56].

C. The Comparison of Communication Overhead

The communication overhead is closely related to the size of messages transmitted between entities. We utilize the Paillier encryption algorithm to encrypt users' power consumption data and use the bilinear aggregate signature [48] to realize authentication between entities. Consequently, the size of a ciphertext is 2048 bits, if we choose the security parameter $\kappa = 1024$ bits.

Chen et al.'s scheme [17]: In the user report generation phase, $user_i$ generates a ciphertext C_i and sends it to FN, where $C_i \in \mathbb{Z}_{N^2}$. Therefore, the communication overhead from $user_i$ to FN is $S_i = 2048$ bits, and the communication overhead from FN to CC is $S_{FN} = 2048$ bits in the privacy-preserving report aggregation phase.

Li et al.'s scheme [24]: In the user report generation phase, $user_i$ generates a ciphertext C_i and sends it to FN, where $C_i \in \mathbb{Z}_{N^2}$. Therefore, the communication overhead from $user_i$ to FN is $S_i = 2048$ bits, and the communication overhead from FN to CC is also $S_{FN} = 2048$ bits in the privacy-preserving report aggregation phase.

Chien et al.'s scheme [25]: $user_i$ generates two ciphertexts $C1_i, C2_i \in \mathbb{Z}_{N^2}$ and sends them to FN in the user report generation phase. Therefore, the communication overhead from $user_i$ to FN is $S_i = 4096$ bits, and the communication overhead from FN to CC is also $S_{FN} = 4096$ bits in the privacy-preserving report aggregation phase.

Zuo et al.'s scheme [56]: $user_i$ generates two ciphertexts $C_i^a, C_i^b \in \mathbb{Z}_{N^2}$ and sends them to FN in the user report generation phase. Therefore, the communication overhead from $user_i$ to FN is $S_i = 4096$ bits, and the communication overhead from FN to CC is also $S_{FN} = 4096$ bits in the privacy-preserving report aggregation phase.

Our proposed scheme: $user_i$ generates a ciphertext C_i and sends it to nearby FN in the user report generation phase, where $C_i \in \mathbb{Z}_{N^2}$. Therefore, the size of $user_i$'s report is calculated as $S_i = 2048$ bits. Subsequently, in the privacy-preserving report aggregation phase, the aggregated result C is sent to CC. As a result, the size of FN's report is calculated as $S_{FN} = 2048$ bits.

From the above comparison, we can draw the conclusion that our proposed scheme achieves privacy-preserving multi-dimensional and multi-subset data aggregation in lower communication overhead than [25], [56].

D. The Comparison of Error

We use the relative error to analyze the error caused by extracting noise from Geometric distribution or Laplace distribution. This error can be measured by comparing the difference between the original aggregated data M_k and the perturbed aggregated data $\tilde{M}_k$. The mathematical expectation of relative error is calculated as $E(error) = \frac{E|\tilde{M}_k - M_k|}{M_k}$. We take the k-th dimensional power consumption of a user as an example, and set the failure rate of SMs is 0.8%. The error analysis is as follows.

1) Geometric distribution error analysis: In the privacy-enhanced scheme, we add noise $\tilde{d}_k$ extracted from Geometric distribution to the original aggregated result M_k in the k-th dimension. Then, the perturbed aggregated data after adding noise is $M_k + \tilde{d}_k$, denoted as $\tilde{M}_k$. According to the mathematical expectation of relative error, we can obtain

$$\widetilde{E}(\zeta_k) = \frac{E|\tilde{M}_k - M_k|}{M_k} = \frac{E|\tilde{d}_k|}{M_k}, \tag{52}$$

where $\tilde{d}_k \sim Geom(exp(-\frac{\epsilon_k}{\Delta A_k}))$ and

$$\begin{aligned} E|\tilde{d}_k| &= \sum_{\tilde{d}_k=-\infty}^{\infty} |\tilde{d}_k| \cdot Pr[\tilde{d}_k] \\ &= \sum_{\tilde{d}_k=-\infty}^{\infty} |\tilde{d}_k| \cdot \frac{1-\alpha}{1+\alpha}\alpha^{\tilde{d}_k} \\ &= \frac{2}{1+\alpha} \cdot \sum_{\tilde{d}_k=1}^{\infty} \tilde{d}_k(1-\alpha) \cdot \alpha^{\tilde{d}_k} \\ &= \frac{2}{1+\alpha} \cdot (\sum_{\tilde{d}_k=1}^{\infty} \tilde{d}_k \cdot \alpha^{\tilde{d}_k} - \sum_{\tilde{d}_k=1}^{\infty} \tilde{d}_k \cdot \alpha^{\tilde{d}_k+1}) \\ &= \frac{2}{1+\alpha} \cdot \sum_{\tilde{d}_k=1}^{\infty} \alpha^{\tilde{d}_k} = \frac{2}{1+\alpha} \cdot \frac{\alpha}{1-\alpha} \\ &= \frac{2\alpha}{1-\alpha^2}. \quad (\because 0 < \alpha < 1) \end{aligned} \tag{53}$$

As $\alpha = exp(-\frac{\epsilon_k}{\Delta A_k})$, therefore, the mathematical expectation of relative error of Geometric distribution is

$$\widetilde{E}(\zeta_k) = \frac{E|\tilde{d}_k|}{M_k} = \frac{2exp(-\frac{\epsilon_k}{\Delta A_k})}{M_k(1 - exp(-\frac{2\epsilon_k}{\Delta A_k}))}. \tag{54}$$

2) Laplace distribution error analysis: In the privacy-enhanced scheme, we add noise $\tilde{d}_{ik}$ extracted from Laplace distribution to d_{ik}. Then, the perturbed aggregated data after adding noise is $M_k + \tilde{d}_k$, denoted as $\tilde{M}_k$. According to the mathematical expectation of relative error, we can obtain

$$\widetilde{E}(\zeta_k) = \frac{E|\tilde{M}_k - M_k|}{M_k} = \frac{E|\tilde{d}_k|}{M_k}, \tag{55}$$

where $\tilde{d}_k \sim Lap(\frac{\Delta A_k}{\epsilon_k})$ and

$$\begin{aligned} E|\tilde{d}_k| &= 2\int_0^{+\infty} \tilde{d}_k \cdot Lap(\frac{\Delta A_k}{\epsilon_k}) d(\tilde{d}_k) \\ &= 2\int_0^{+\infty} \tilde{d}_k \cdot \frac{\epsilon_k}{2\Delta A_k} \cdot e^{-\frac{\tilde{d}_k \cdot \epsilon_k}{\Delta A_k}} d(\tilde{d}_k) \\ &= \frac{\Delta A_k}{\epsilon_k}. \end{aligned} \tag{56}$$

Therefore, the mathematical expectation of relative error of Laplace distribution is

$$\widetilde{E}(\zeta_k) = \frac{E|\tilde{d}_k|}{M_k} = \frac{\Delta A_k}{\epsilon_k \cdot M_k}. \tag{57}$$

3) The comparison of utility: The utility of privacy-preserving data aggregation scheme with differential privacy is mainly affected by the relative error of the aggregated result. Therefore, from the perspective of utility of privacy-preserving data aggregation scheme, we conduct experiment to evaluate and compare the utility here.

On the one hand, we compare relative errors of Geometric distribution with these of Laplace distribution under different ϵ_k. We extract noises from Geometric distribution and Laplace distribution, respectively, and add them to the real power consumption data in the fog-based smart grid. Specifically, we vary the privacy budgets of the k-th dimension ϵ_k from $\{0.1, 0.2, 0.3\}$, and vary the number of users n from $\{1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000\}$. The comparison of relative errors on Geometric distribution and Laplace distribution under different ϵ_k are depicted in Figure 5. As shown in Figure 5, when the number of users is $n = 5000$, the privacy budget is $\epsilon_k = 0.2$, the relative error of privacy-preserving data aggregation scheme which extracted noise from Geometric distribution is 0.192307564%, and the relative error of privacy-preserving data aggregation scheme which extracted noise from Laplace distribution is 0.192307692%. The relative errors of these two methods are very close and are difficult to distinguish. At the same time, we can find that a higher level of privacy needs sacrifice more data utility as a cost. That is, for the same number of users n, the increase of privacy-preserving intensity will introduce more noise resulting in a greater loss of data utility.

On the other hand, we compare our proposed scheme with the state-of-the-art schemes [51]–[55] in term of relative error. We set $\epsilon_k = 0.2$ and vary the number of users n from $\{1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000\}$. The comparison of relative errors for schemes [51]–[55] and ours when $\epsilon_k = 0.2$ are depicted in Figure 6. According to the Figure 6, when the number of users n increases, the relative error will decrease while the data utility will grow, and the relative error is kept within 0.2% when $n \geq 5000$. Therefore, we can draw the conclusion that our proposed scheme can maintain higher data utility under the same privacy budget, which means that our proposed scheme introduces lower noises while resisting differential attack.

IX. Conclusion

In this paper, for fog-based smart grids, we proposed a privacy-preserving multi-dimensional and multi-subset data aggregation scheme with differential privacy. Even if the user's power consumption data is multi-dimensional, users can also be divided into different subsets according to their power consumption. As a result, the CC can perform fine-grained analysis on user's power consumption data. Security analysis demonstrates that our proposed scheme can resist differential attack, eavesdropping attack, collusion attack and active attack. The experiment results show that our proposed scheme is more efficient at computational overhead and communication overhead. In future, we will study tariff problem [60], general transactive energy (TE) retailing problem [61], user access control using lightweight face verification [62], trusted entities problem by leveraging the Trusted Execution Environment (TEE) [63] to meet the requirements of fair pricing, intelligent electronic devices and user trust.

X. Acknowledgement

The authors sincerely thank the editors and all the anonymous reviewers for their valuable comments which have helped to enhance the quality of the paper. This research was supported in part by the advanced computing resources provided by the Supercomputing Center of Hangzhou City University. This research was also supported by a joint funding program by Hangzhou Hyperchain Technology Co., Ltd.

References

[1] S. Han, M. Xie, H. Chen, and Y. Ling, "Intrusion detection in cyber-physical systems: Techniques and challenges," *IEEE Systems Journal*, vol. 8, no. 4, pp. 1052-1062, 2014.

[2] W. Pan, H. Ming, Carl K. Chang, Z. Yang and D. Kim, "ElementRank: Ranking java software classes and packages using a multilayer complex network-based approach," *IEEE Transactions on Software Engineering*, vol. 47, pp. 2272-2295, 2021.

[3] W. Pan, H. Ming, D. Kim, and Z. Yang, "Pride: Prioritizing documentation effort based on a pagerank-like algorithm and simple filtering rules," *IEEE Transactions on Software Engineering*, vol. 49, pp. 1118-1151, 2023.

[4] S. Han, J. Lin, S. Zhao, G. Xu, S. Ren, D. He, L. Wang and L. Shi, "Location privacy-preserving distance computation for spatial crowdsourcing," *IEEE Internet of Things Journal*, vol. 7, pp. 7550-7563, 2020.

[5] X. Pang, Z. Wang, Z. He, P. Sun, M. Luo, J. Ren and K. Ren, "Towards class-balanced privacy preserving heterogeneous model aggregation," *IEEE Transactions on Dependable and Secure Computing*, https://doi.org/10.1109/TDSC.2022.3183170, 2022.

[6] X. Pang, Z. Wang, J. Li, R. Zhou, J. Ren and Z. Li, "Towards online privacy-preserving computation offloading in mobile edge computing," *in 2022 IEEE INFOCOM*, pp. 1179-1188, 2022.

[7] Y. Zhang, X. Jia, B. Pan, J. Shao, L. Fang, R. Lu and G. Wei, "Anonymous multi-hop payment for payment channel networks," *IEEE Transactions on Dependable and Secure Computing*, https://doi.org/10.1109/TDSC.2023.3262681, 2023.

[8] X. Jia, Z. Yu, J. Shao, R. Lu, G. Wei and Z. Liu, "Cross-chain virtual payment channels," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 3401-3413, 2023.

[9] F. Li, H. Li, B. Niu, and J. Chen, "Privacy computing: Concept, computing framework and future development trends," *Engineering*, vol. 5, no. 6, pp. 1179-1192, 2019.

[10] Y. Liu, W. Guo, C. Fan, L. Chang, and C. Cheng, "A practical privacy-preserving data aggregation (3PDA) scheme for smart grid," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 3, pp. 1767-1774, 2019.

[11] H. J. Jo, I. S. Kim, and D. H. Lee, "Efficient and privacy-preserving metering protocols for smart grid systems," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1732-1742, 2016.

[12] P. Gope and B. Sikdar, "An efficient data aggregation scheme for privacy-friendly dynamic pricing-based billing and demand-response management in smart grids," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 3126-3135, 2018.

[13] D. He, N. Kumar, S. Zeadally, A. Vinel, and L. T. Yang, "Efficient and privacy-preserving data aggregation scheme for smart grid against internal adversaries," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2411-2419, 2017.

[14] C. Guo, X. Jiang, K.-K. R. Choo, X. Tang, and J. Zhang, "Lightweight privacy preserving data aggregation with batch verification for smart grid," *Future Generation Computer Systems*, vol. 112, pp. 512-523, 2020.

[15] Z. Guan, Y. Zhang, L. Wu, J. Wu, J. Li, Y. Ma, and J. Hu, "APPA: An anonymous and privacy preserving data aggregation scheme for fog-enhanced IoT," *Journal of Network and Computer Applications*, vol. 125, pp. 82-92, 2019.

[16] 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, 2012.

[17] Y. Chen, J. Martínez-Ortega, P. Castillejo, and L. López, "A homomorphic-based multiple data aggregation scheme for smart grid," *IEEE Sensors Journal*, vol. 19, no. 10, pp. 3921-3929, 2019.

[18] O. R. Merad Boudia, S. M. Senouci, and M. Feham, "Elliptic curve-based secure multidimensional aggregation for smart grid communications," *IEEE Sensors Journal*, vol. 17, no. 23, pp. 7750-7757, 2017.

[19] Q. Kong, R. Lu, M. Ma, and H. Bao, "A privacy-preserving sensory data sharing scheme in internet of vehicles," *Future Generation Computer Systems*, vol. 92, pp. 644-655, 2017.

[20] Y. Ming, X. Zhang, and X. Shen, "Efficient privacy-preserving multi-dimensional data aggregation scheme in smart grid," *IEEE Access*, vol. 7, pp. 32907-32921, 2019.

[21] 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, 2017.

[22] O. R. Merad-Boudia and S. M. Senouci, "An efficient and secure multidimensional data aggregation for fog-computing-based smart grid," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6143-6153, 2021.

[23] R. Lu, K. Alharbi, X. Lin, and C. Huang, "A novel privacy-preserving set aggregation scheme for smart grid communications," in *2015 IEEE GLOBECOM*, pp. 1-6, 2015.

[24] S. Li, K. Xue, Q. Yang, and P. Hong, "PPMA: Privacy-preserving multisubset data aggregation in smart grid," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 2, pp. 462-471, 2018.

[25] H. Chien and C. Su, "A fault-tolerant and flexible privacy-preserving multisubset data aggregation in smart grid," in *International Conference on Computational Science/Intelligence & Applied Informatics*, pp. 165–175, 2019.

[26] G. Ács and C. Castelluccia, "I have a dream! (differentially private smart metering)," in *International Workshop on Information Hiding*, pp. 118-132, 2011.

[27] J. Won, C. Y. T. Ma, D. K. Y. Yau, and N. S. V. Rao, "Privacy-assured aggregation protocol for smart metering: A proactive fault-tolerant approach," *IEEE/ACM Transactions on Networking*, vol. 24, no. 3, pp. 1661-1674, 2016.

[28] V. Rastogi and S. Nath, "Differentially private aggregation of distributed time-series with transformation and encryption," in *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*, pp. 735-746, 2010.

[29] N. Gai, K. Xue, B. Zhu, J. Yang, J. Liu and D. He, "An efficient data aggregation scheme with local differential privacy in smart grid," *Digital Communications and Networks*, https://doi.org/10.1016/j.dcan.2022.01.004, 2022.

[30] H. Bao and R. Lu, "A new differentially private data aggregation with fault tolerance for smart grid communications," *IEEE Internet of Things Journal*, vol. 2, no. 3, pp. 248-258, 2015.

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

[32] Z. Shi, R. Sun, R. Lu, L. Chen, J. Chen, and X. S. Shen, "Diverse grouping-based aggregation protocol with error detection for smart grid communications," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2856-2868, 2015.

[33] S. Han, S. Zhao, Q. Li, C. -H. Ju, and W. Zhou, "PPM-HDA: Privacy-preserving and multifunctional health data aggregation with fault tolerance," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 9, pp. 1940-1955, 2016.

[34] S. Zhao, F. Li, H. Li, R. Lu, S. Ren, H. Bao, J. Lin, and S. Han, "Smart and practical privacy-preserving data aggregation for fog-based smart grids," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 521-536, 2021.

[35] S. Han, H. Ding, S. Zhao, S. Ren, Z. Wang, J. Lin, and S. Zhou, "Practical and robust federated learning with highly scalable regression training," *IEEE Transactions on Neural Networks and Learning Systems*, https://doi.org/10.1109/TNNLS.2023.3271859, 2023.

[36] R. Lu, K. Heung, A. H. Lashkari, and A. A. Ghorbani, "A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced IoT," IEEE Access, vol. 5, pp. 3302-3312, 2017.

[37] A. Kumari, S. Tanwar, S. Tyagi, N. Kumar, M. S. Obaidat, and J. J. P. C. Rodrigues, "Fog computing for smart grid systems in the 5G environment: Challenges and solutions," *IEEE Wireless Communications*, vol. 26, no. 3, pp. 47-53, 2019.

[38] J. Wang, L. Wu, K. R. Choo, and D. He, "Blockchain-based anonymous authentication with key management for smart grid edge computing infrastructure," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 1984-1992, 2020.

[39] Y. Yu, R. Chen, H. Li, Y. Li, and A. Tian, "Toward data security in edge intelligent IIoT," *IEEE Network*, vol. 33, no. 5, pp. 20-26, 2019.

[40] A. A. Sadri, A. M. Rahmani, M. Saberikamarposhti, and M. Hosseinzadeh, "Fog data management: A vision, challenges, and future directions," *Journal of Network and Computer Applications*, vol. 174, pp. 102882, 2021.

[41] Z. Wang, D. Jiang, F. Wang, Z. Lv, and R. Nowak, "A polymorphic heterogeneous security architecture for edge-enabled smart grids," *Sustainable Cities and Society*, vol. 67, pp. 102661, 2021.

[42] J. Zhang, Y. Zhao, J. Wu, and B. Chen, "Lvpda: A lightweight and verifiable privacy-preserving data aggregation scheme for edge-enabled iot," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4016-4027, 2020.

[43] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *International Conference on the Theory and Applications of Cryptographic Techniques*, pp. 223-238, 1999.

[44] C. Dwork, "Differential privacy," in *Encyclopedia of Cryptography and Security, 2nd Ed*, pp. 338-340, Springer, 2011.

[45] H. Li, J. Cui, X. Meng, and J. Ma, "IHP: Improving the utility in differential private histogram publication," *Distributed Parallel Databases*, vol. 37, no. 4, pp. 721-750, 2019.

[46] C. Yang, T. Chang, and M. Hwang, "A(t,n) multi-secret sharing scheme," *Applied Mathematics and Computation*, vol. 151, no. 2, pp. 483-490, 2004.

[47] D. Kifer and B. Lin, "Towards an axiomatization of statistical privacy and utility," in *Proceedings of the Twenty-ninth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pp. 147-158, 2010.

[48] D. Boneh, C. Gentry, B. Lynn, and H. Shacham, "Aggregate and verifiably encrypted signatures from bilinear maps," in *International Conference on the Theory and Applications of Cryptographic Techniques*, pp. 416-432, 2003.

[49] C. Guo and C. Chang, "A novel threshold conference-key agreement protocol based on generalized chinese remainder theorem," *International Journal of Network Security*, vol. 17, no. 2, pp. 165-173, 2015.

[50] A. Ghosh, T. Roughgarden, and M. Sundararajan, "Universally utility-maximizing privacy mechanisms," *SIAM Journal on Computing*, vol. 41, no. 6, pp. 1673-1693, 2012.

[51] N. Gai, K. Xue, B. Zhu, J. Yang, J. Liu, and D. He, "An efficient data aggregation scheme with local differential privacy in smart grid," *Digital Communications and Networks*, vol. 8, no. 3, pp. 333-342, 2022.

[52] Y. Sei, J. Andrew, H. Okumura, and A. Okumura, "Privacy-preserving collaborative data collection and analysis with many missing values," *IEEE Transactions on Dependable and Secure Computing*, vol. 8, pp. 2158-2173, 2023.

[53] D. Ye, S. Shen, T. Zhu, B. Liu, and W. Zhou, "One parameter defense‐defending against data inference attacks via differential privacy," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 1466-1480, 2022.

[54] L. Chen, L. Zeng, Y. Mu, and L. Chen, "Global combination and clustering based differential privacy mixed data publishing," *IEEE Transactions on Knowledge and Data Engineering*, https://doi.org/10.1109/TKDE.2023.3237822, 2023.

[55] T. Wang, W. Yang, X. Ma, and B. Wang, "Event-set differential privacy for fine-grained data privacy protection," *Neurocomputing*, vol. 515, pp. 48-58, 2023.

[56] X. Zuo, L. Li, H. Peng, S. Luo, and Y. Yang, "Privacy-preserving multidimensional data aggregation scheme without trusted authority in smart grid," *IEEE Systems Journal*, vol. 15, no. 1, pp. 395-406, 2021.

[57] G. Xu, H. Bai, J. Xing, T. Luo, N. N. Xiong, X. Cheng, S. Liu, and X. Zheng, "SG-PBFT: A secure and highly efficient distributed blockchain PBFT consensus algorithm for intelligent Internet of vehicles," *Journal of Parallel and Distributed Computing*, vol. 164, pp. 1-11, https://doi.org/10.1016/j.jpdc.2022.01.029, 2022.

[58] G. Xu, W. Dong, J. Xing, W. Lei, J. Liu, L. Gong, M. Feng, X. Zheng, and S. Liu, "Delay-CJ: A novel cryptojacking covert attack method based on delayed strategy and its detection," *Digital Communications and Networks*, https://doi.org/10.1016/j.dcan.2022.04.030, 2022.

[59] T. Hussain, B. Yang, H. U. Rahman, A. Iqbal, and F. Ali, "Improving source location privacy in social internet of things using a hybrid phantom routing technique," *Computers & Security*, vol. 123, pp. 102917, 2022.

[60] S. Wang, X. Tan, T. Liu, and D. H. K. Tsang, "Aggregation of demand-side flexibility in electricity markets: Negative impact analysis and mitigation method," *IEEE Transactions on Smart Grid*, vol. 12, no. 1, pp. 774-786, 2021.

[61] X. Tan, A. Leon-Garcia, Y. Wu, and D. H. K. Tsang, "Posted-price retailing of transactive energy: An optimal online mechanism without prediction," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 1, pp. 5-16, 2020.

[62] J. Lin, C. Ye, W. Liu, S. Ren, Y. Wang, W. Ma and Y. Ding, "A lightweight face verification based on adaptive cascade network and triplet loss function," *Wireless Communications and Mobile Computing*, https://doi.org/10.1155/2022/3017149, 2022.

[63] M. Wang, K. He, J. Chen, R. Du, B. Zhang and Z. Li, "PANDA: Lightweight non-interactive privacy-preserving data aggregation for constrained devices," *Future Generation Computer Systems*, vol. 131, pp. 28-42, 2022.