TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214
DOI: 10.26599/TST.2025.9010032

Research Article

Distributed Differential Privacy Protection with High Data Availability in Smart Grids

You Li, Yan Huo*, Xin Fan, Chengxin Niu, Jian Mao, and Tao Jing

Abstract: In smart grids, real-time electricity data uploaded by smart meters may be analyzed by an attacker with other data analytics methods, which may expose users' privacy. To ensure user privacy, differential privacy methods are often used to process data. However, these methods reduce the accuracy of the data results obtained by the center and lead to unavailability of the data. In this paper, we address this problem and propose a distributed differential privacy protection scheme. Two methods of data noise addition and data perturbation are fused and used in the protection scheme. Data accuracy is improved by optimizing the noise generation method. To address the problem of quantitatively balancing the users' privacy needs with the central analytics needs, this paper describes the needs of both through mathematical definitions, i.e., data accuracy and data privacy, and proposes a privacy budget that balances data accuracy and privacy. The performance of the proposed scheme is evaluated using the typical power data, which proves the excellent performance.

Key words: smart meter; data privacy; data accuracy; distributed differential privacy

1 Introduction

Internet of Things (IoT) has become an integral part of daily life, with smart grids as a key application. Smart grids are intelligent power networks that combine high-speed communication, data analysis, and distribution[1]. Smart meters regularly collect household appliance data to the grid's analysis center. The center then intelligently manages the allocation of electricity based on the data of the smart meter[2]. However, smart grids have faced many challenges. In 2020, attackers illegally obtained sensitive data from the European energy giant and held the company to ransom. Since the fine-grained power data collected by smart meters contains a large amount of sensitive user information, data leakage will lead to the exposure of users' privacy[3, 4]. Attackers can exploit leaked data to infer users' habits and potentially commit theft or other illegal activities, which is unacceptable to users. Thus, user data must be protected before use to reduce attack risks and protect privacy.

In smart grid, user privacy protection schemes are divided into two categories, one based on user identity privacy protection schemes and the other based on user data privacy protection schemes. For example, the

• You Li is with School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China, and also with the Aostar Information Technologies Co. Ltd, Chengdu 610094, China. E-mail: 22115007@bjtu.edu.cn.
• Yan Huo, Chengxin Niu, and Tao Jing are with School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China. E-mail: yhuo@bjtu.edu.cn; 24120096@bjtu.edu.cn; tjing@bjtu.edu.cn.
• Xin Fan is with School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China, and also with the Engineering Research Center for Forestry-Oriented Intelligent Information Processing of National Forestry and Grassland Administration, Beijing 100083, China. E-mail: fanxin@bjfu.edu.cn.
• Jian Mao is with School of Cyber Science and Technology, Beihang University, Beijing 100191, China, also with the Tianmushan Laboratory, Hangzhou 311115, China, and also with the Zhongguancun Laboratory, Beijing 100080, China. E-mail: maojian@buaa.edu.cn.
* To whom correspondence should be addressed.
Manuscript received: 2024-09-09; revised: 2024-11-29; accepted: 2025-02-28

work introduces a protection scheme that uses group blind signatures to anonymize user identities[5]. This approach effectively prevents attackers from discerning any correlation between the data and its corresponding users, rendering subsequent data analysis futile and substantially mitigating the risk of user privacy breaches. However, the privacy protection scheme that only protects the user's identity is not strong enough and needs to protect the users' data. The work proposes a privacy protection scheme, Efficient Privacy-preserving Multiple Data Aggregation Scheme (EPri-MDAS), which is based on ElGamal homomorphic encryption technology and does not require a trusted third party[6]. This method protects user privacy by encrypting data. However, this method of data encryption puts a huge computational burden on smart meters. Some works put forward to hide the real power consumption information privacy protection scheme by rechargeable batteries[7, 8]. However, this will increase the cost for the user, as rechargeable batteries need to be purchased. The works in Refs. [9, 10] used differential privacy technology to protect user privacy, which is a low-cost and efficient method. However, when the differential privacy mechanism is used to process the data, the data results may become inaccurate, and there is a contradiction between the accuracy of the data results and the data privacy. On the one hand, these schemes based on differential privacy do not take into account the balance between data accuracy and data privacy. At present, the data accuracy of most distributed differential privacy protection schemes is lower than that of centralized differential privacy schemes. On the other hand, the centralized differential privacy scheme does not take into account the reliability of data centers, which spurs distributed differential privacy. However, the impact of noise on differential privacy strategies is not yet fully understood, which leads to incomplete policy design for distributed differential privacy. In particular, the work in Ref. [10] did not discuss the influence of the number of discrete noise points N on the final result when using the distributed Laplacian mechanism.

In order to solve these problems, this work proposes a privacy protection scheme based on a distributed differential privacy mechanism. The scheme uses distributed noise injection technology to independently add noise that meets the differential privacy mechanism to the original electricity data of each smart meter, change the value of the electricity data record, and protect the privacy of users. At the same time, the scheme uses mathematical formulas to measure data accuracy and privacy, ensuring the balance of data accuracy and privacy. The main contributions of this work are as follows:

- We propose a distributed differential privacy scheme for smart grid power data upload. In our scheme, data privacy and data accuracy are described quantitatively, and a privacy protection scheme is provided that takes both data accuracy and privacy into account. Moreover, we deeply study the generation method of distributed Laplacian mechanism noise, discuss the influence of hyperparameter N (the number of discrete noise points extracted) on the accuracy of data results, and hence to optimize our scheme.
- Based on the distributed Laplacian mechanism, we propose a new perturbation algorithm that can make change to data values. Different from previous perturbation algorithms, this algorithm can keep the time order of data and realize data perturbation. In addition to increasing the uncertainty of the data, it provides useful time information for the center to facilitate the analysis.
- We evaluate the proposed scheme's performance in two aspects: Privacy and accuracy. The centralized methods are considered as the benchmarks, demonstrating the excellent performance of our proposed scheme.

The rest of this paper is organized as follows. Section 2 presents the related work on privacy protection schemes in smart grids. Section 3 describes the relevant theoretical knowledge of differential privacy technology. Section 4 introduces our privacy protection model. Section 5 is the experimental section. Section 6 summarizes our work.

2 Related Work

Smart grid user privacy protection covers the protection of user identity in the grid as well as the protection of real-time fine-grained user data collected by smart meters. The research on privacy protection in grid scenarios is summarized below.

2.1 Privacy protection of user identity

Currently, signature technology is mainly used to protect the identity of individual users. Wang et al.[11] proposed a protection scheme using ring signature technology to achieve unlimited anonymity of individual users and message integrity. Fouda et al.[12]

proposed an identity authentication protection scheme based on hash message authentication codes and Diffie–Hellman key establishment protocols to ensure secure data release. Zhu et al.[13] proposed an efficient lattice-based identity signature protocol, which enhances protocol efficiency through commitment tree structure to deal with quantum attack risk and protect user data integrity in smart grid.

2.2 Privacy protection of user data

There are three main categories of user data protection: Data encryption, differential privacy, and battery charging techniques.

In the data encryption scheme, Zhang et al.[14] proposed an authentication scheme with a low performance overhead, which combines public-private key encryption technology to achieve secure transmission of power data. However, after data decryption, the data aggregator or center can still obtain user data. From the user's perspective, the data aggregator or center is not completely trustworthy, so this scheme cannot effectively protect user privacy. To avoid the problem of untrustworthy data centers, many scholars have proposed schemes based on homomorphic Paillier encryption techniques to aggregate data without decryption[15–17]. Yan et al.[18] proposed privacy protection of smart meter data based on improved homomorphic encryption and double-blind noise addition protocol. Xu et al.[19] proposed a homomorphic encryption protection scheme that takes into account the privacy protection requirements in different trust boundary scenarios and uses multiple public keys to effectively protect user privacy. In addition, to ensure the communication security and integrity of the power data, some schemes also introduce blockchain technology. For example, Singh et al.[20] proposed a blockchain-based homomorphic encryption protection scheme, which improves security while reducing computing costs. Since the power data collected by smart meters is real-time, this means that smart meters need to upload data to the center multiple times a day, and each uploading of data to the center requires encryption processing, which will greatly consume the computing resources of smart meters and data centers.

Privacy protection schemes based on battery charging technology can achieve privacy protection with low computing overhead. Zhu et al.[7] proposed a physical privacy protection scheme using battery charging technology. The user's power data is blurred by the charging and discharging of the battery to protect user privacy. Natgunanathan et al.[8] proposed an energy-saving battery charging privacy protection scheme and used adaptive smoothing output to effectively reduce the change of output load. However, there are also some problems with the use of battery charging technology. First, under the policy of time-of-use electricity pricing, it is necessary to discuss how to bill. Second, the introduction of rechargeable battery technology requires users to purchase rechargeable batteries, which increases user costs. Third, the use of rechargeable batteries will bring fire hazards to users.

Differential privacy is widely used in data privacy protection[21–27], which is an excellent privacy protection technology. Differential privacy technology is a privacy protection method that is defined mathematically and can be proven mathematically[28–30]. Common models of differential privacy are centralized differential privacy and distributed differential privacy. In the centralized model, the differential privacy of data is completed in the data center, which is safe and reliable and can ensure the privacy of user data. In the local model, the differential privacy of the data is completed locally by the user and the data center is not trustworthy. In addition, the random response mechanism is also a type of differential privacy mechanism, which is a distributed mechanism[31]. The random response mechanism implements differential privacy by injecting perturbation noise into the dataset with a certain probability and then obtains statistical characteristics consistent with the original dataset through probabilistic statistics. Due to the powerful privacy protection performance of differential privacy technology, researchers have considered using this technology when designing user power data privacy protection schemes. Bao and Lu[32] proposed a secure data aggregation scheme, called DDPFT, that uses differential privacy technology and a fault-tolerant approach to distributed data aggregation. Fotiou et al.[33] used the smart contract technology in the blockchain combined with the differential privacy technology to implement a privacy protection scheme and used the RAPPOR[34] proposed by Google for performance verification. Zhang et al.[35] proposed a dual-response differential privacy mechanism that enhances blockchain privacy protection by synchronizing the privacy budget and historical query

information. Zheng et al.[10] proposed a data privacy protection scheme that combines the Laplace mechanism with a random scrambling algorithm. The scheme has a small performance overhead and an average data utility. At the same time, this scheme publishes user data with a smaller time interval, which overexposes user data, and its scrambling algorithm is based on the scrambling of time order, which destroys the time correlation of the data. Gai et al.[36] proposed a differential privacy protection scheme based on the k-RR mechanism. The scheme has good data availability. However, the division of sub-intervals in this scheme is affected by the number of users. When the number of users increases or decreases, the sub-interval difference needs to be redetermined, and hence the scheme has poor universality. The detailed analysis shows that the differential privacy scheme is the most suitable scheme to protect the privacy of user-power data.

3 Preliminary

3.1 Definition of differential privacy

Differential privacy is to blur data records by changing the values or attributes of data records in a dataset. When an attacker has some data records in the dataset, the attacker cannot determine the authenticity of other data records in the acquired dataset. The specific definition is as follows:

$$p[C(A)=R]/p[C(A')=R] \leqslant \mathrm{e}^{\varepsilon} \tag{1}$$

where A and A' are adjacent datasets with only one different data record, R is all the data records consisting of the two datasets, p represents the probability density distribution, $C(\cdot)$ is a mechanism, and ε is defined as the privacy budget. This definition is specifically explained as follows: When the ratio of the probability density distribution of the output R obtained by the mechanism $C(\cdot)$ of the two adjacent datasets A and A' is less than or equal to e^{ε}, $C(\cdot)$ is a differential privacy mechanism[28, 29]. Privacy budget ε is used to measure the distance between two probability density distributions, which is derived from the KL divergence. The smaller ε is, the closer the two probability density distributions are, and the more indistinguishable the two probability density distributions are, the better the privacy.

Mechanism $C(\cdot)$ is a Laplace mechanism when it fits the definition below[37]. The Laplace differential privacy mechanism is a strict one that can change the magnitude of data values.

$$C(A)=c(A)+\mathrm{Laplace}\left(\frac{\Delta c}{\varepsilon}\right) \tag{2}$$

where $c(A)$ is the value search function of dataset A, and $\mathrm{Laplace}\left(\frac{\Delta c}{\varepsilon}\right)$ is the discrete value of $\frac{\Delta c}{\varepsilon}$ with location parameter 0 and scale parameter. Δc is the sensitivity of $c(A)$ and ε is the privacy budget. Sensitivity[37] is defined as

$$\Delta c=\max_{m,n\in A,m\neq n}\|c(m)-c(n)\| \tag{3}$$

3.2 Decomposability of the Laplace distribution

In the Laplace mechanism, the position parameter of the Laplacian probability density function is set to zero to ensure that the sum of noise is zero. Based on this feature, the Laplacian distribution can be decomposed into n gamma distributions ($n>1$, $n\in\mathbb{Z}^{+}$). Before the derivation, the three main properties used in the derivation are listed.

Property 1: When the positional parameter is 0, the probability density curve of the Laplace distribution is symmetric along the y-axis, consisting of two symmetric exponential distributions. The probability density function of x with scale q is

$$p(x|0,q)=(2q)^{-1}\mathrm{e}^{\frac{-|x|}{q}} \tag{4}$$

Property 2: The random variable Y obeys the gamma distribution, $Y\sim\Gamma(\alpha,\beta)$, where α is the shape parameter and β is the inverse scale parameter. When $\alpha=1$, Y obeys an exponential distribution.

Property 3: Additivity of the gamma distribution with n random variables $X_i\sim\Gamma(\alpha_i,\beta)$, then $\sum_{i=1}^{n}X_i\sim\Gamma\left(\sum_{i=1}^{n}\alpha_i,\beta\right)$.

The derivation process is as follows: There are n random variables $X_i\sim\Gamma\left(\frac{1}{n},\beta\right)$ and n random variables $Y_i\sim\Gamma\left(\frac{1}{n},\beta\right)$. By Property 3, it follows that $\sum_{i=1}^{n}X_i\sim\Gamma(1,\beta)$, $\sum_{i=1}^{n}Y_i\sim\Gamma(1,\beta)$. By Property 2, $\sum_{i=1}^{n}X_i$ and $\sum_{i=1}^{n}Y_i$ follow an exponential distribution. By Property 1, it follows that $\left(\sum_{i=1}^{n}X_i-\sum_{i=1}^{n}Y_i\right)\sim\mathrm{Laplace}(0,\beta)$. Therefore, $\sum_{i=1}^{n}(X_i-Y_i)\sim\mathrm{Laplace}(0,\beta)$. So, the Laplace

distribution can consist of several gamma distributions subtracted from each other.

3.3 Properties of differential privacy

3.3.1 Invariance under split operations

Suppose that there exists a dataset A and $C(A)$ is a differential privacy mechanism with privacy budget ε. By dividing the data A into n mutually exclusive datasets A_1 to A_n and processing these datasets using the mechanism C, the mechanism that discloses all the results of the processing $C(A_1)$ to $C(A_n)$ is still the privacy mechanism with ε[38].

3.3.2 Invariance under multiple reprocessing

Let there be a dataset A with $C_1(A)$ as a differential privacy mechanism with ε_1 and $C_2(A)$ as an arbitrary processing mechanism. Then the dataset A is processed by both mechanisms and $C_2[C_1(A)]$ is the privacy mechanism of ε_1 [39].

3.4 Relative error and entropy increase rate

Relative error *d*: The main measure of the distance between two data, the larger the error, the farther the distance between the two data, the greater the difference. Suppose the original data is S_1 and the processed data is S_2, we define the relative error of the data as

$$d = \frac{|S_1 - S_2|}{S_1} \times 100\% \tag{5}$$

Entropy increase rate: Mainly used to measure the change of a sequence, when the sequence is more disordered, the entropy increase rate becomes larger. The formula for discrete information entropy is $E(A) = -\sum_{i=1}^{n} P(a_i)\log P(a_i)$, where a_i is an element of the data sequence A and P denotes the probability. Assuming that the information entropy of the original data sequence is E_1 and the information entropy of the processed data sequence is E_2, the entropy increase rate is given by

$$\frac{E_2 - E_1}{E_1} \times 100\% \tag{6}$$

3.5 Theoretical analysis of bit perturbation algorithm based on random response mechanism

A random response mechanism is widely used in event investigation with two attributes and is a typically distributed mechanism[37]. The principle of the random response mechanism is illustrated by an example below. Suppose there are n individuals, each of whom owns a small ball, which has two colors, red and green. There are b individuals who have one of the green balls. These n individuals are asked in turn about the color of the ball they own, but not all of them want the others to know the color of the ball in their hands, so each answer with probability p to a color different from the color of the ball in their hand. The statistics are carried out and it is known that there are b' people who have green balls and the remaining people $n - b'$ people who have red balls. Based on the results of the statistics, one can use the method of maximum likelihood estimation to be able to know the true number of people with green balls.

In the bit perturbation algorithm, the two colors of the blob are mapped as 0 and 1 on the data bits. The statistics are perturbed by transforming 0 to 1 and 0 to 1 with a probability of p. The data are then transformed into 1 and 0 into 1 with a probability of p. The true ratio of 0 and 1 is then estimated from the statistical ratio. The difference is that in the bit perturbation algorithm, it is also necessary to calculate the data error based on the estimated ratio of 0, 1, and the weights of the data bits w. Assuming a total of n samples, the proportion of ones in the m samples is b, the number of ones after counting is n_1, and the weight of the data bits is w, we can calculate the data bias error. The probability of 1 after statistics can be expressed as $P(x=1) = b(1-p) + (1-b)p$. The probability of 0 after statistics can be expressed as $P(x=0) = bp + (1-b)(1-p)$. Then we can establish the likelihood function and take logarithms, we have

$$\ln[\wedge(b)] = n_1 \ln[b(1-p) + (1-b)p] + (n - n_1)\ln[bp + (1-b)(1-p)].$$

Let $\partial \ln[\wedge(b)]/\partial b = 0$, the great likelihood estimate of b, for $p \neq 0.5$ is given by $b = (np - n_1)/n(2p-1)$. So the number of ones before the perturbation is $(np - n_1)/(2p-1)$, The number of zeros is $[n(p-1) + n_1]/(2p-1)$. Deviations can be calculated as

$$\text{error} = \left\{ \frac{(np - n_1)p}{2p-1} - \frac{[n(p-1) + n_1]p}{2p-1} \right\} w.$$

The design of the bit perturbation algorithm is based on the above theoretical analysis unfolding, it needs to be proposed that the relationship between the probability p and the privacy budget ε: $\varepsilon = \ln[(1-p)/p]$[37]. Generally speaking, the privacy

budget is usually between 0 and 1, privacy to ensure that the privacy budget for the value of p is generally 0 to 0.5. When p is greater than 0.5, the calculated privacy budget is negative, and p for the value of 0 to 0.5 presents a relationship about the odd symmetry. Therefore, the range of p in the above study is [0, 0.5).

4 Proposed Distributed Differential Privacy Scheme

4.1 Distributed privacy protection model

A smart grid is mainly divided into two parts, the communication network, and the distribution network, as shown in Fig. 1. The communication network is mainly responsible for the transmission of grid information, and the distribution network is mainly responsible for power transmission. The communication network and distribution network are connected to the smart power data center, and the center manages and coordinates the two networks to ensure that the two networks can work properly. Among them, the communication network is the key network for realizing intelligent management of the smart grid. The distribution network includes all the power equipment connected to the user end and is controlled by the smart data center, which enables accurate scheduling of power automation and intelligence.

We propose a distributed model of data accuracy and privacy. The model consists of two parts: Data noise addition and data perturbation. In the part of data noise addition, we adopt the optimized distributed Laplacian noise addition algorithm. We use the method of generating standard noise and then enlarging it to generate noise of the corresponding scale. In this way, the distribution of discrete noise points is more uniform and the noise deviation is smaller, which is conducive to improving the accuracy of data. In the data perturbation section, we design a bit perturbation algorithm that can change the size of data values. This new perturbation method does not break the correlation between data and time like previous perturbation methods and provides time information for data centers while enhancing data privacy. The model treats all data nodes except users themselves as untrustworthy and implements strict privacy protection. In this model, all users will independently complete the differential privacy processing of their smart meter data, and generate independent Laplacian noise by using the separability of Laplacian distribution proposed in Section 3 of this paper. Each user disturbs the data with a certain probability to further disturb the data and destroy the data regularity, and finally publish the data to the aggregator for data analysis by the data center, as shown in Fig. 2.

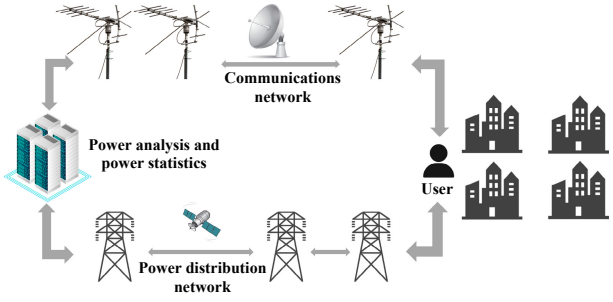

Fig. 1 Architecture diagram of the smart grid model.

Suppose there are n users within a block and the individual user dataset is $\{A_{t=1}, A_{t=2}, \ldots, A_{t=96}\}$. Take the moment t as an example, and specifically the working process of the model. Firstly, the smart meter in the block will generate discrete N-dimensional noise data locally using two subtractive gamma distributions according to the privacy budget ε and data sensitivity Δc agreed with the center, and then sum the N-dimensional noise data to get the noise data H_t^i, and then add the noise data H_t^i with the user's data a_t^i, and finally perturb it numerically using the bit perturbation algorithm. The significance of generating N-dimensional noise data here is to use this N-dimensional discrete data to characterize the result of subtracting two consecutive gamma distributions,

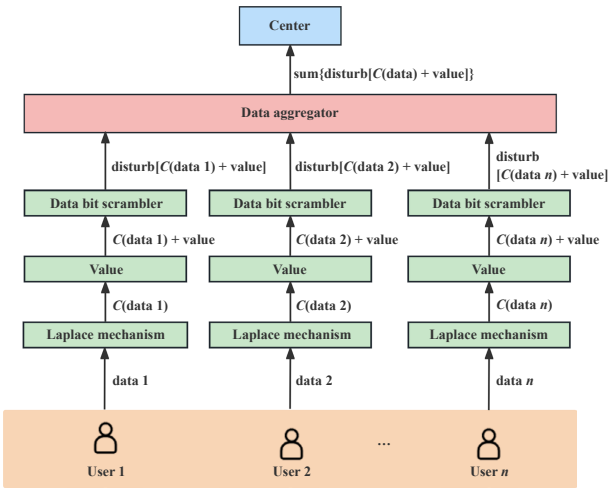

Fig. 2 Workflow of the distributed differential privacy data protection model.

which is equivalent to sampling these two consecutive gamma distributions. At moment t, the designed distributed Laplace mechanism can be expressed as $C(A_t) = c(A_t) + H_t$, where $c(A_t)$ is the search function of A_t. $H_t = \{H_t^1, H_t^2, \ldots, H_t^n\}$ is the set of n user noises, $H_t^i = \sum_{j=1}^{N} h_j$ is the sum of N-dimensional noisy data, and $\{h_1, h_2, \ldots, h_N\}$ is a discrete sampling of the $\Gamma_1\left(n^{-1}, \Delta c/\varepsilon\right) - \Gamma_1\left(n^{-1}, \Delta c/\varepsilon\right)$ distribution. Define the bit perturbation algorithm as disturb and value as a fixed constant. The final published data can be expressed as disturb$[C(A_t) + \text{value}]$. A fixed value is added to each noised data firstly to ensure that the perturbation value of each data is positive before the data is perturbed, and secondly to add a larger fixed value to the data to ensure that the overall data is greater than the perturbation value, to achieve the full perturbation of the data. A_t at each moment t is processed by disturb$[C(A_t) + \text{value}]$, and from the invariance of the split processing of differential privacy in Section 3 of the paper, it is known that each user's data A_i for one day in the block will satisfy the differential privacy mechanism of ε.

4.2 Distributed noise addition algorithm based on Laplace mechanism

The distributed noise addition algorithm (Algorithm 1) is based on the Laplace differential privacy mechanism. It leverages the decomposability of the Laplace distribution, as detailed in Section 3, to inject noise into the data in a manner that is independently and uniformly distributed. The algorithm ensures that aggregated data retains Laplace distribution characteristics. It derives discrete noise from a continuous distribution using two gamma distributions. By subtracting these two gamma distributions and sampling the resultant distribution at N points, the algorithm preserves the distribution's properties within the N-point dataset. To avoid the problem that the noise values collected during the random sampling of the noisy data are seriously deviated from the expected ones, which causes significant aggregation errors. The algorithm first collects smaller noise and then expands it, reducing the error of aggregation results and improving data accuracy under the same privacy budget. The specific algorithm flow is shown in Fig. 3. Firstly, the algorithm generates two gamma distributions, $\Gamma_1(1/n, 1)$ and $\Gamma_2(1/n, 1)$, and subtracts

Algorithm 1 Algorithm of distributed noise addition algorithm based on Laplace mechanism

Initialize: User data: electricity data

Output: Noise-added data: electricity data noise

1: Initialise the noise point N;
2: Initialise the privacy budget e;
3: n = length(electricity data);
4: sensitivity = max(electricity data) − min(electricity data);
5: scale = sensitivity/e;
6: **for** $i \in [1, n]$ **do**
7: standard noise ← gamma$_1$ − gamma$_2$;
8: result ← sum(standard noise);
9: noise data ← result × scale;
10: Noise[] ← noise data;
11: **end for**
12: **for** $i \in [0, n-1]$ **do**
13: data ← electricity data[i] + Noise[i];
14: electricity data noise[] ← data;
15: **end for**
16: **return** electricity data noise

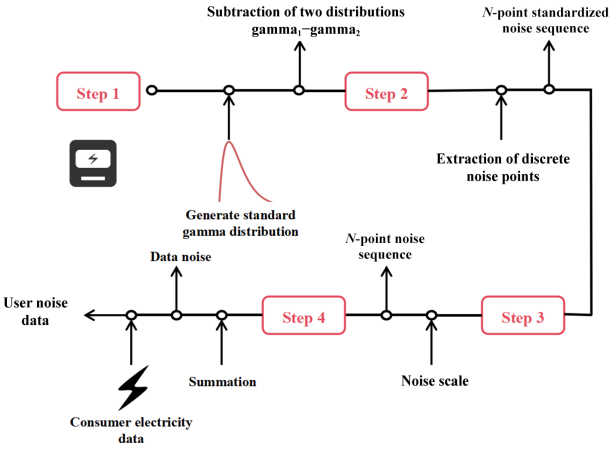

Fig. 3 Laplace noise algorithm flowchart.

the two distributions to obtain the distribution, $\Gamma_1\left(n^{-1}, 1\right) - \Gamma_2\left(n^{-1}, 1\right)$. Subsequently, a random collection of N points in the distribution is performed as the standard noise data. Next, the algorithm calculates the noise scale parameters according to $\Delta c/\varepsilon$ and expands the noise according to the noise scale to get a noise sequence of N points. Finally, the noise data is summed to obtain the noise value and added to the user power data. The pseudo-code of the algorithm is shown in Algorithm 1.

In the Laplace mechanism, a simpler way is for each user to independently sample noise from the Laplace

distribution. However, this requires a sufficient number of users, as each only samples a discrete amount of noise. Only with a large number of users can enough discrete data be generated to characterize this Laplace distribution. Therefore, only with a sufficiently large number of users within the block can it be ensured that the noise values extracted in this way are usable and that the aggregation error is acceptable. If user numbers change, aggregation error varies. In particular, fewer users increase the aggregation error. To keep the error stable, the algorithm does not extract noise in this way. Using gamma-distributed synthetic noise ensures stable aggregation error without specific requirements for the number of users in the block.

The data noise addition effect of the algorithm is demonstrated below. The algorithm uses typical electricity data[40] for the data noise addition work. The data noise addition results with a privacy budget of 0.3 are shown in Fig. 4. It can be seen that the noise addition to the data greatly destroys the correlation between the data, blurring the user's power usage at the current time. It can be seen that the algorithm is able to protect the user's privacy.

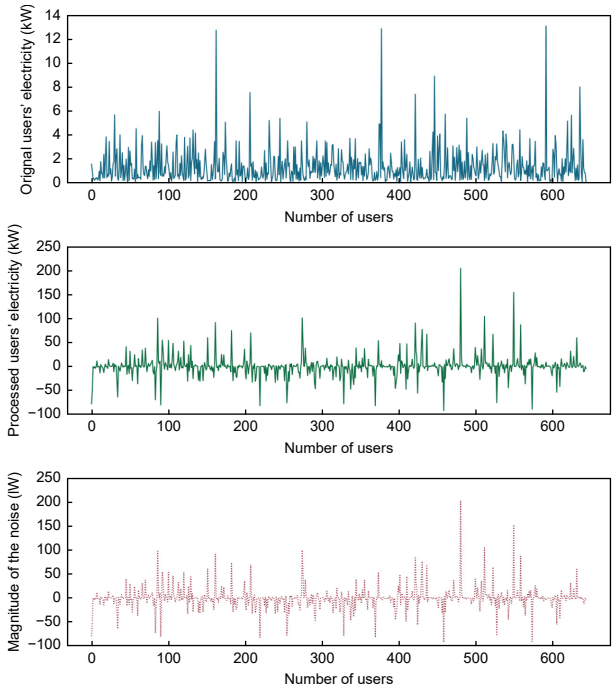

Fig. 4 Processing results of users electricity. The blue curve is the original electricity consumption in kW for 644 pieces of users' data. The green curve is the processed data after the noise addition process using $N = 100$. The red curve is the difference between the original data and the processed data, i.e., the magnitude of the noise.

4.3 Bit perturbation algorithm based on random response mechanisms

The bit perturbation algorithm (Algorithm 2) based on the random response mechanism innovatively applies the random response mechanism to the differential privacy processing of numerical values and cleverly corresponds the zeros and ones in the data to the attributes in the random response mechanism by converting the data into binary form. The data frequency obtained by using the maximum likelihood estimation is used to calculate the data deviation caused by the data perturbation, and the deviation is summed up with the aggregation result, which ensures the usability of the aggregation result. Theoretical analysis is detailed in Section 3. This perturbation algorithm changes the data values in a probabilistic

Algorithm 2 Bit perturbation algorithm based on random response mechanisms

Initialize: data, perturbation probability p, perturbation start bit x, number of perturbation bits a

Output: disturb data, sum disturb data error

```
n = length(data);
for i ∈ [0, n − 1] do
    integer part ← data[i];
    fractional part ← data[i];
    binary encode[] ← integer part;
    if probability generator < p then
        for j ∈ [x, a + x − 1] do
            binary encode[j] ← negation(binary encode[j]);
        end for
    end if
    disturb integer part ← decoder(binary encode[]);
    disturb data[] ← disturb integer part + fractional part;
end for
for i ∈ [0, n − 1] do
    disturb integer part ← disturb data[i];
    binary encode[] ← disturb integer part;
    for j ∈ [x, a + x − 1] do
        if binary encode[j] == 1 then
            number[j]++;
        end if
    end for
end for
disturb data error ← calculate(number[]);
sum disturb data error ← sum(disturb data error);
return disturb data, sum disturb data error
```

way, rather than destroying the correlation between the data and time as the existing methods do. This has the advantage of providing the center with useful time information while enhancing data privacy, helping the center better understand how regional power consumption changes over time. By adjusting perturbation bits and probabilities, the algorithm can reduce the correlation of the data over a continuous period, thus effectively countering data analysis and preventing the attacker from obtaining information about the data and the algorithm through continuous observation.

The algorithm flow is as shown in Fig. 5. First, the algorithm divides the input data into an integer part and a decimal part. The integer part is then binary encoded and converted into 8-bit binary data. The selected data bits are then perturbed to change the data size by performing an inverse operation on the data bits with a probability of p. Next, the integer and fractional parts of the perturbation are summed to obtain the final perturbed data. Subsequently, the distribution of the data is performed. At the receiving end of the data, the integer part of the data is taken out in the same way, the number of zeros and ones on the perturbed data bits are counted, and the deviation of the data is calculated based on the weights of the data bits and the method of calculating the data deviation error derived in Section 3 of this paper. Finally, the final aggregation result is obtained by adding the data deviation and the existing aggregation result. The pseudo-code of the algorithm is shown in Algorithm 2, describing the exact implementation of the algorithm.

Below is a demonstration of one effect of perturbing the data using the algorithm. Figure 6 is a plot of the effect after a two-bit perturbation with a perturbation probability of 0.3. It can be seen that the algorithm can randomly change the value of the original data to achieve data disturbance and hide the change rule of

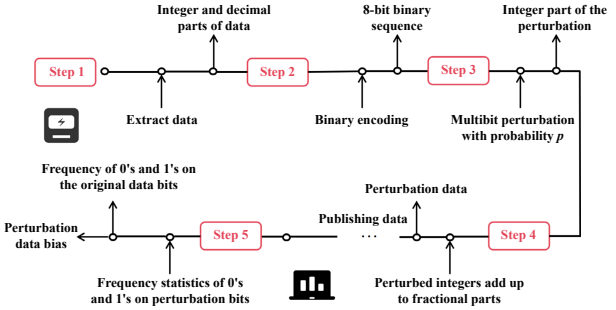

Fig. 5 Bit perturbation algorithm flowchart.

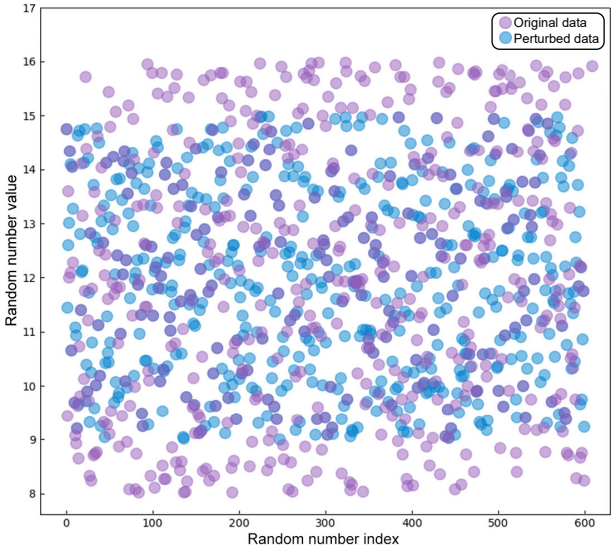

Fig. 6 Data perturbation results with Algorithm 2 (p = 0.3). The blue data points are 600 random numbers generated using a random number generator with a range of 9 to 15 retaining two decimal places, and the purple ones are after perturbation using the algorithm.

the original sequence. By selecting different perturbation probabilities and different perturbation bits, the risk of the attacker learning the data perturbation method through long-term data analysis can be reduced. Therefore, the algorithm can provide effective data privacy protection.

5 Performance Analysis

In this section, real-time electric power usage data collected by Pecan Street Energy from customers in a region of Texas for July 2015[40] is used for the experiments. The test environment is Python 3.10 and the test equipment was a Dell laptop.

Data accuracy and data privacy are used to evaluate the experimental results. The data value of each data series will change after privacy processing, but the sum of the data series after privacy processing remains within a certain error range compared with the sum of the original data series. For data accuracy, we first calculate the relative error d between the sum of the processed data series and the sum of the original data series and then calculate the data accuracy according to $100\% - d$. The smaller the added noise, the smaller the relative error and the higher the data accuracy. After privacy processing, the data sequence will become more chaotic, and the information entropy of the data sequence will increase. When more noise is added to the data series, the more chaotic the data

series is, the information entropy of the data series increases, and the data security increases. For data privacy, we calculate the relative error between the information entropy of the processed data series and that of the original data series, that is, the entropy increase rate.

5.1 Effect of hyperparameters on data accuracy

5.1.1 Effect of the number of discrete noise points

The privacy budget ε in the algorithm is a key parameter, and the value of the privacy budget directly determines the magnitude of the injected noise. When the privacy budget is smaller, the injected noise value is larger and the data accuracy decreases. To explore the data accuracy of the algorithm under different privacy budgets, a random number generator is used to generate 1000 random numbers for the experiments, the data retain two decimals, and the data range from 0 to 6. In addition, the number of extracted noise points N in the algorithm also affects the data accuracy. Therefore, during the experiments, the relationship between privacy budget and data accuracy with different N is considered. The value of each data accuracy is the average of the data accuracy obtained by repeating the experiment 20 times based on the same N value and the same privacy budget premise.

The data is visualized as shown in Fig. 7. An analysis of Fig. 7 below shows that the algorithm performs well in terms of data accuracy. When the number of the extracted noise points $N = 200$, the data accuracy stays above 90% with different privacy budgets. With the same value of N, the data accuracy increases as the privacy budget increases, which is consistent with the pre-experimental analysis. Observing the privacy budgets at different values of N shows that data accuracy does not perform best at $N = 400$ or $N = 100$ but at $N = 200$. When extracting noise points, since it is a limited number of extractions, the sum of the extracted noise values does not sum to 0 as in a continuous distribution, but rather there will be some deviation, and the sum will generally be close to 0 but not 0. Therefore, extracting too many noise points will introduce too many additional data errors, leading to a decrease in data accuracy. In extracting too few noise points, these discrete noise points cannot fully characterize the distribution and will also introduce large data errors when aggregated, leading to a decrease in data accuracy. Therefore, a compromise value of N should be chosen for data noise addition and better data accuracy has been obtained.

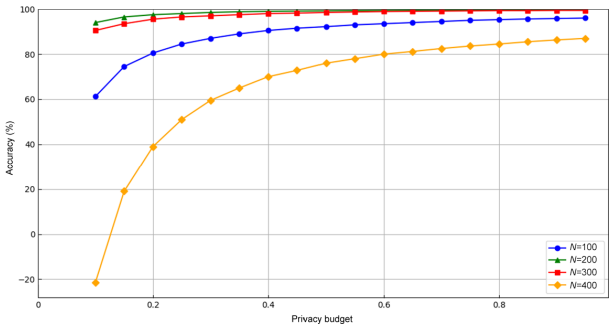

Fig. 7 Privacy budget and data accuracy under different numbers of the extracted noise points.

5.1.2 Effect of the number of data perturbations

The probability of perturbation p and the number of bits of data perturbation a in the algorithm are the key parameters that affect the data accuracy. To explore the effect of perturbation probability and the number of data perturbation bits on data accuracy, a random number generator was used in this experiment to generate 1000 random numbers for the experiment, with two decimal places reserved for the data values, and the data range is from 15 to 30. Considering the relatively small data range, only the case where the maximum number of perturbation bits of 4 is discussed. The data accuracy is the average of the data accuracy for 20 experiments with the same perturbation probability and the same number of perturbation bits. To facilitate the analysis, a line graph with privacy budget as the horizontal coordinate and data accuracy as the vertical coordinate is plotted using the data, as shown in Fig. 8. Combined with the graph, it can be seen that data accuracy decreases when the probability of perturbation increases. The data accuracy decreases when the number of perturbation bits increases. However, overall, the data accuracy of the

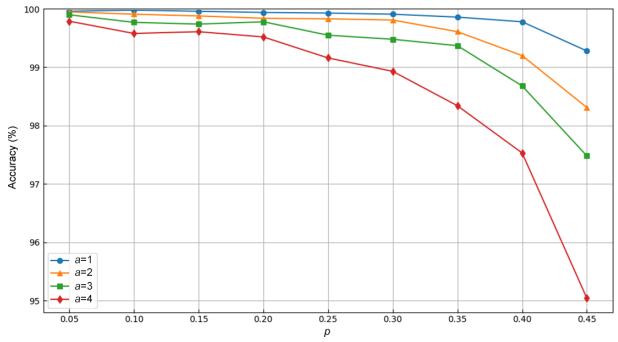

Fig. 8 Data accuracy comparisons of the perturbation algorithm under different numbers of bits of data perturbation.

algorithm is all above 90 percent and the data accuracy is excellent.

5.2 Performance analysis of distributed models

In this subsection, the performance of the model is evaluated in terms of data accuracy and privacy. Meanwhile, the performance difference between using only the Laplace noise addition and using both algorithms is compared. It is demonstrated that the superposition of the two algorithms can effectively improve the model performance. In the experiment, $n = 644$ power data from users are used for evaluation. The effects of different privacy budgets and perturbation probabilities on the model performance are investigated. In the experiments, discrete data characterizing noise with $N = 100$ are extracted and the data were perturbed with two bits of perturbation. The experiments explored the performance of the model when using only the data noise addition algorithm. As well as the performance of the model using both algorithms under different privacy budgets and perturbation probabilities is also tested. To eliminate chance error, 30 experiments are conducted with the same parameters. The experimental results are shown in Fig. 9, which shows the averaged results of 30 experiments for data privacy and accuracy.

In Fig. 9, the orange plot lines represent the data accuracy and data privacy of the model under different privacy budgets when using only the noise addition algorithm. The blue, green, and red lines show accuracy and privacy using both algorithms at different perturbation probabilities and privacy budgets. Comparing the lines shows that using both algorithms provides significantly better privacy than using only the noise addition algorithm. And the privacy-preserving performance of the model keeps increasing as the probability of perturbation increases. Data accuracy is highest when using only the noise addition algorithm and decreases slightly with increased perturbation when using both algorithms. The plots of model data accuracy overlap more and decrease less significantly than the plots of model data privacy separation. This suggests that the model achieves better privacy-preserving performance when using both algorithms while sacrificing a small portion of data accuracy compared to the model's data-noise-only algorithm. The overall trend shows data accuracy improves with higher privacy budgets and remains high overall. Data privacy decreases with higher privacy budgets but remains above 50%, even at the maximum budget, showing strong privacy performance.

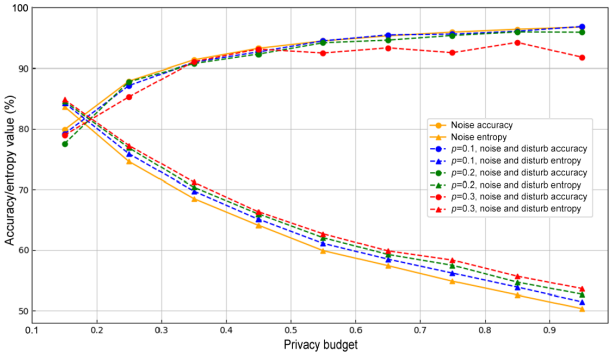

Fig. 9 Model performance analysis.

5.3 Privacy budget that balances data accuracy and privacy

In the previous subsection, it can be observed that data accuracy and privacy exhibit opposite trends with the privacy budget. When the model performs privacy protection, to balance the data accuracy and privacy, it is necessary to select an appropriate privacy budget to make the data accuracy of the model equal to the data privacy of the model. This point is the optimal privacy budget for the model. Using $n = 644$ user power data, we explore the optimal privacy budget (N = 100) for different disturbance probabilities and bits. The best privacy plot for a one-bit disturbance and probability 0.1 is analyzed below, as shown in Fig. 10. The blue color in Fig. 10 is the privacy curve and the purple color is the data accuracy curve, when the two curves intersect, i.e., the point marked orange in Fig. 10, the privacy budget is optimal. To the left of the orange point, privacy exceeds accuracy, indicating

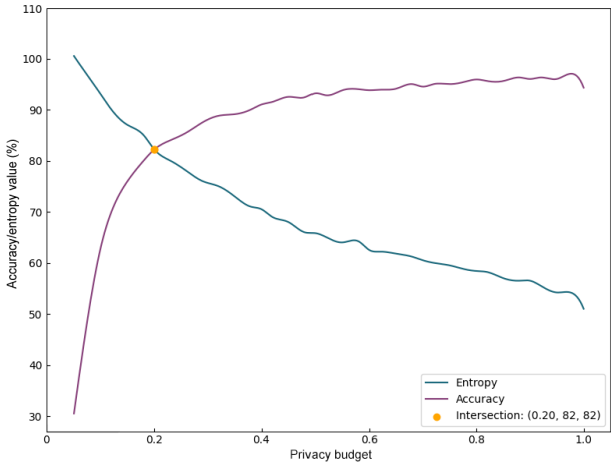

Fig. 10 Result of balancing the privacy budget.

overprotection. To the right, accuracy exceeds privacy, suggesting underprotection.

In Table 1, the optimal privacy budget and data utility for different parameter cases of the model are presented. When the curves intersect, data privacy and data utility are equal. Therefore, the data utility in Table 1 is either data accuracy or data privacy. Combining the four cases of data bit perturbation shows that the optimal privacy budget of the model is around 0.2 and the data utility of the model is around 0.82, which puts the data utility within an acceptable range.

5.4 Comparison with centralized models

This subsection provides a comparative analysis of the centralized and distributed models in terms of two aspects: Data accuracy and data privacy. Before making the comparison, the centralized model is constructed as shown in Fig. 11. The traditional Laplace mechanism as well as the bit perturbation algorithm proposed in this paper are used in the centralized model.

The same data samples are used for both models, the distributed model noise point is 644, and both model

Table 1 Relationship between perturbation probability and optimal privacy budget under different data bit perturbations.

Perturbation	Probability	Position
One data bit perturbation	0	(0.20, 0.82)
	1	(0.2, 0.82)
	2	(0.2, 0.83)
	3	(0.2, 0.83)
	4	(0.2, 0.82)
Two data bit perturbation	0	(0.20, 0.82)
	1	(0.20, 0.83)
	2	(0.21, 0.83)
	3	(0.21, 0.84)
	4	(0.20, 0.84)
Three data bit perturbation	0	(0.20, 0.82)
	1	(0.21, 0.83)
	2	(0.24, 0.83)
	3	(0.29, 0.82)
	4	(0.50, 0.73)
Four data bit perturbation	0	(0.20, 0.82)
	1	(0.23, 0.84)
	2	(0.32, 0.83)
	3	(0.44, 0.82)
	4	(0.48, 0.70)

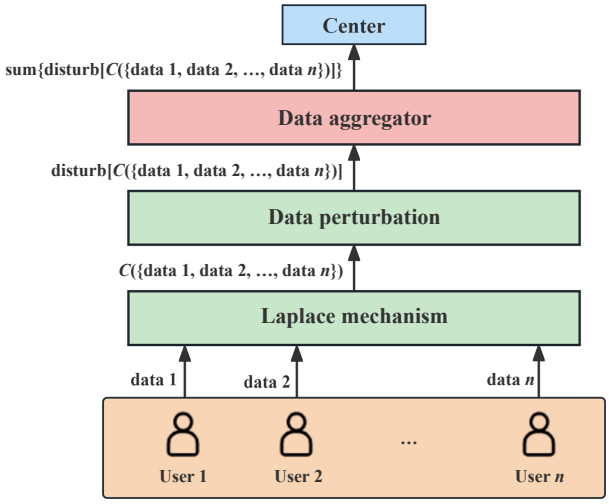

Fig. 11 Flow diagram of the centralized differential privacy model.

bit perturbation algorithms used a perturbation probability of 0.2 and two bits of perturbation. In terms of data accuracy, it is shown in Fig. 12. Figure 12 analyzes the data accuracy of the two models under different privacy budget premises. The blue scatter Fig. 12 is the data accuracy of the distributed model and the pink scatter is the data accuracy of the centralized model. From Fig. 12, it can be seen that the data accuracy of the centralized model is quite excellent, keeping the overall data accuracy above 80%, while in the distributed model, the effect of data accuracy with a privacy budget of around 0.2 is comparable to that of the centralized model. The reason for the good data accuracy of the centralized model is that the center can calculate the sum of the extracted noise before injecting the noise into the data, and if the sum of the noise is too large and unacceptable, the center can re-do the extraction of the noise until it meets the requirements. And this is not possible with

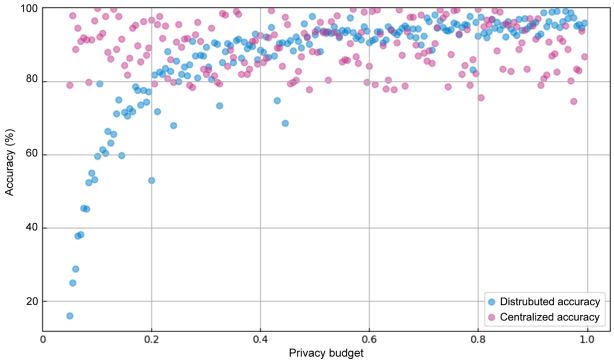

Fig. 12 Comparison of data accuracy between centralized and distributed models.

distributed models, because the noise of each individual in a distributed model is agnostic. In terms of data privacy, it is shown in Fig. 13. The data privacy performance of the model is evaluated assuming that the data center is trusted. As can be seen from Fig. 13, the distributed model is comparable to the centralized model. Thus combining the two dimensions, the distributed model outperforms the centralized model. Because the distributed model achieves good data privacy at this value and its usability is comparable to that of the centralized model. Moreover, in the distributed model, there is no need to consider whether the center is trustworthy or not, and privacy protection is more stringent.

6 Conclusion

6.1 Summary of research

This paper addresses the privacy protection of smart meter user data. After preliminary literature research and theoretical studies, a privacy protection scheme with differential privacy as the main technique is identified. Through the study of differential privacy technology and the requirements of data protection in grid scenarios, a distributed privacy protection scheme with high data accuracy is designed, and the scheme incorporates the use of two algorithms, distributed noise injection and bit perturbation, to improve the privacy protection performance of the scheme. This paper also strictly describes data privacy and data utility, and the proposed algorithm and model are comprehensively evaluated by these two metrics, which are experimentally proved to have excellent performance. Finally, this paper compares the proposed model with the centralized model, proving again through experiments and analysis that the proposed model performs better in privacy preservation.

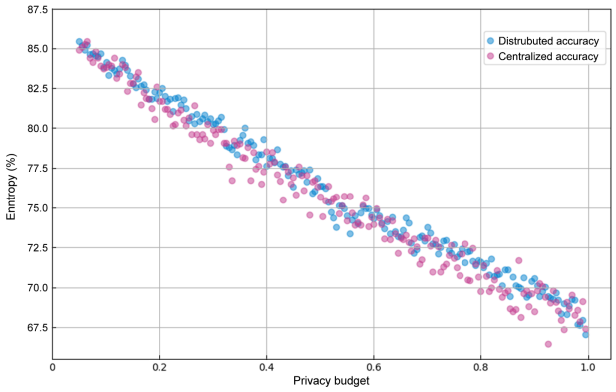

Fig. 13 Comparison of data privacy between centralized and distributed models.

6.2 Outlook for research

There are still some aspects of this research work that deserve further research and improvement. In this paper, there are some limitations in the methodology of considering all users as the same type of user. There may be various types of users in the same block, including general household users, enterprises and factories, etc., which generate significantly different power data. When performing privacy protection, users with smaller values of electricity data are prone to overprotection, while users with larger values of electricity data are prone to underprotection. Therefore, all users should not be subjected to the same intensity of privacy protection. To solve the above problems and further improve the privacy protection performance of the scheme, in future research work, we can consider adopting methods such as cluster analysis to analyze the users in the block, formulate corresponding classification rules, and carry out personalized privacy protection of different strengths for different types of users.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (Nos. U24B20117 and 62502036), the Fundamental Research Funds for the Central Universities (No. 2019JBZ001), the China Postdoctoral Science Foundation (No. 2024M750199), and the Natural Science Foundation of Zhejiang Province of China (No. LZ23F020013).

References

[1] V. C. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, and G. P. Hancke, Smart grid technologies: Communication technologies and standards, *IEEE Trans. Industr. Inform.*, vol. 7, no. 4, pp. 529–539, 2011.

[2] Y. Liu, W. Guo, C. I. Fan, L. Chang, and C. Cheng, A practical privacy-preserving data aggregation (3PDA) scheme for smart grid, *IEEE Trans. Industr. Inform.*, vol. 15, no. 3, pp. 1767–1774, 2019.

[3] M. L. Tuballa and M. L. Abundo, A review of the development of smart grid technologies, *Renew. Sustain. Energy Rev.*, vol. 59, pp. 710–725, 2016.

[4] S. Li, K. Xue, D. S. L. Wei, H. Yue, N. Yu, and P. Hong, SecGrid: A secure and efficient SGX-enabled smart grid system with rich functionalities, *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 1318–1330, 2020.

[5] W. Kong, J. Shen, P. Vijayakumar, Y. Cho, and V. Chang,

A practical group blind signature scheme for privacy protection in smart grid, *J. Parallel Distrib. Comput.*, vol. 136, pp. 29–39, 2020.

[6] J. Zhang, W. Zhang, X. Wei, and H. Liu, EPri-MDAS: An efficient privacy-preserving multiple data aggregation scheme without trusted authority for fog-based smart grid, *High-Confid. Comput.*, vol. 4, no. 4, p. 100226, 2024.

[7] L. Zhu, Z. Zhang, Z. Qin, J. Weng, and K. Ren, Privacy protection using a rechargeable battery for energy consumption in smart grids, *IEEE Netw.*, vol. 31, no. 1, pp. 59–63, 2017.

[8] I. Natgunanathan, M. B. Hossain, Y. Xiang, L. Gao, D. Peng, and J. Li, Progressive average-based smart meter privacy enhancement using rechargeable batteries, *IEEE Internet Things J.*, vol. 6, no. 6, pp. 9816–9828, 2019.

[9] L. Ou, Z. Qin, S. Liao, T. Li, and D. Zhang, Singular spectrum analysis for local differential privacy of classifications in the smart grid, *IEEE Internet Things J.*, vol. 7, no. 6, pp. 5246–5255, 2020.

[10] Z. Zheng, T. Wang, A. K. Bashir, M. Alazab, S. Mumtaz, and X. Wang, A decentralized mechanism based on differential privacy for privacy-preserving computation in smart grid, *IEEE Trans. Comput.*, vol. 71, no. 11, pp. 2915–2926, 2022.

[11] H. Wang, L. Wang, M. Wen, K. Chen, and Y. Luo, A lightweight certificateless aggregate ring signature scheme for privacy protection in smart grids, *Wirel. Pers. Commun.*, vol. 126, no. 2, pp. 1577–1599, 2022.

[12] M. M. Fouda, Z. M. Fadlullah, N. Kato, R. Lu, and X. Shen, Towards a light-weight message authentication mechanism tailored for smart grid communications, in *Proc. 2011 IEEE Conf. Computer Communications Workshops* (*INFOCOM WKSHPS*), Shanghai, China, 2011, pp. 1018−1023.

[13] L. Zhu, F. Jiang, M. Luo, and Q. Li, An efficient identity-based signature protocol over lattices for the smart grid, *High-Confid. Comput.*, vol. 3, no. 4, p. 100147, 2023.

[14] L. Zhang, L. Zhao, S. Yin, C. H. Chi, R. Liu, and Y. Zhang, A lightweight authentication scheme with privacy protection for smart grid communications, *Future Gener. Comput. Syst.*, vol. 100, pp. 770–778, 2019.

[15] 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 Trans. Parallel Distrib. Syst*., vol. 23, no. 9, pp. 1621−1631, 2012.

[16] L. Chen, R. Lu, and Z. Cao, PDAFT: A privacy-preserving data aggregation scheme with fault tolerance for smart grid communications, *Peer Peer Netw. Appl*., vol. 8, no. 6, pp. 1122−1132, 2015.

[17] W. J. Zhou, H. C. Zhu, S. Y. Yao, and T. Li, A homomorphic encryption-based privacy preserving data aggregation scheme for smart grid, in *Proc. 15th Int. Conf. Computational Intelligence and Security* (*CIS*), Macao, China, 2019, pp. 315−319.

[18] R. Yan, Y. Zheng, N. Yu, and C. Liang, Multi-smart meter data encryption scheme based on distributed differential privacy, *Big Data Mining and Analytics*, vol. 7, no. 1, pp. 131–141, 2024.

[19] W. Xu, J. Sun, R. Cardell-Oliver, A. Mian, and J. B. Hong, A privacy-preserving framework using homomorphic encryption for smart metering systems, *Sensors*, vol. 23, no. 10, p. 4746, 2023.

[20] P. Singh, M. Masud, M. S. Hossain, and A. Kaur, Blockchain and homomorphic encryption-based privacy-preserving data aggregation model in smart grid, *Comput. Electr. Eng.*, vol. 93, p. 107209, 2021.

[21] Z. He, L. Wang, and Z. Cai, Clustered federated learning with adaptive local differential privacy on heterogeneous IoT data, *IEEE Internet Things J.*, vol. 11, no. 1, pp. 137–146, 2024.

[22] H. Xu, W. Li, S. Wu, L. Zhao, and Z. Cai, APOLLO: Differential private online multi-sensor data prediction with certified performance, in *Proc. 2024 IEEE Int. Conf. Data Mining* (*ICDM*), Abu Dhabi, UAE, 2024, pp. 530−539.

[23] Z. Cai, X. Zheng, and J. Yu, A differential-private framework for urban traffic flows estimation via taxi companies, *IEEE Trans. Industr. Inform.*, vol. 15, no. 12, pp. 6492–6499, 2019.

[24] Z. Cai, X. Zheng, J. Wang, and Z. He, Private data trading towards range counting queries in Internet of Things, *IEEE Trans. Mob. Comput.*, vol. 22, no. 8, pp. 4881–4897, 2023.

[25] Z. Cai and X. Zheng, A private and efficient mechanism for data uploading in smart cyber-physical systems, *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 2, pp. 766–775, 2020.

[26] W. Zhang, Z. Xie, A. M. Vera Venkata Sai, Q. Zia, Z. He, and G. Yin, A local differential privacy trajectory protection method based on temporal and spatial restrictions for staying detection, *Tsinghua Science and Technology*, vol. 29, no. 2, pp. 617–633, 2024.

[27] X. Wang, L. Mo, X. Zheng, and Z. Dang, Streaming histogram publication over weighted sliding windows under differential privacy, *Tsinghua Science and Technology*, vol. 29, no. 6, pp. 1674–1693, 2024.

[28] C. Dwork, Differential privacy, in *Proc. 33rd Int. Colloquium on Automata, Languages, and Programming*, Venice, Italy, 2006, pp. 1−12.

[29] C. Dwork, F. McSherry, K. Nissim, and A. Smith, Calibrating noise to sensitivity in private data analysis, in *Proc. 3rd Theory of Cryptography Conf. Third Theory of Cryptography*, New York, NY, USA, 2006, pp. 265−284.

[30] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, Our data, ourselves: Privacy via distributed noise generation, in *Proc. 25th Int. Conf. Theory and Applications of Cryptographic Techniques on Advances in Cryptology*, St. Petersburg, Russia, 2006, pp. 486−503.

[31] S. L. Warner, Randomized response: A survey technique for eliminating evasive answer bias, *J. Am. Stat. Assoc.*, vol. 60, no. 309, pp. 63–69, 1965.

[32] H. Bao and R. Lu, DDPFT: Secure data aggregation scheme with differential privacy and fault tolerance, in *Proc. 2015 IEEE Int. Conf. Communications* (*ICC*), London, UK, 2015, pp. 7240−7245.

[33] N. Fotiou, I. Pittaras, V. A. Siris, G. C. Polyzos, and P. Anton, A privacy-preserving statistics marketplace using

local differential privacy and blockchain: An application to smart-grid measurements sharing, *Blockchain*: *Res. Appl.*, vol. 2, no. 1, p. 100022, 2021.
[34] U. Erlingsson, V. Pihur, and A. Korolova, RAPPOR: Randomized Aggregatable privacy-preserving ordinal response, in *Proc. 2014 ACM SIGSAC Conf. Computer and Communications Security*, Scottsdale, AZ, USA, 2014, pp. 1054–1067.
[35] K. Zhang, P. W. Tsai, J. Tian, W. Zhao, X. Cai, L. Gao, and J. Chen, Towards privacy in decentralized IoT: A blockchain-based dual response DP mechanism, *Big Data Mining and Analytics*, vol. 7, no. 3, pp. 699–717, 2024.
[36] 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, *Digit. Commun. Netw.*, vol. 8, no. 3, pp. 333–342, 2022.
[37] M. Yang, T. Guo, T. Zhu, I. Tjuawinata, J. Zhao, and K. Y. Lam, Local differential privacy and its applications: A comprehensive survey, *Comput. Stand. Interfaces*, vol. 89, p. 103827, 2024.
[38] F. D. McSherry, Privacy integrated queries: An extensible platform for privacy-preserving data analysis, in *Proc. 2009 ACM SIGMOD Int. Conf. Management of Data*, Providence, RI, USA, 2009, pp. 19–30.
[39] D. Kifer and B. R. Lin, Towards an axiomatization of statistical privacy and utility, in *Proc. 29th ACM SIGMOD-SIGACT-SIGART Symp. Principles of Database Systems*, Indianapolis IND, USA, 2010, pp. 147–158.
[40] Pecan Street Inc., Dataport, https://www.pecanstreet.org/dataport/, 2024.

You Li received the BEng degree from Shanghai Electric Power University, Shanghai, China, in 2013, and the MEng degree from North China Electric Power University, Beijing, China, in 2019. He is currently pursuing the PhD degree in communication and information system at School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China. He is also with the Aostar Information Technologies Co. Ltd., Chengdu, China. His current research interests include smart grids, resource allocation, privacy, and security.

Yan Huo received the BEng and PhD degrees in communication and information system from Beijing Jiaotong University, Beijing, China, in 2004 and 2009, respectively. He was a visiting scholar at Department of Computer Science, George Washington University, Washington, DC, USA, from 2015 to 2016. He is currently a professor at School of Electronics and Information Engineering, Beijing Jiaotong University, Beijing, China. He has served as an associate editor for *IEEE Access* and a reviewer for a number of journals, including *IEEE Wireless Communications*, *IEEE Internet of Things Journal*, *IEEE Transactions on Wireless Communications*, *IEEE Transactions on Vehicular Technology*, and *IEEE Transactions on Mobile Computing*. His current research interests include wireless communications, physical layer security, privacy protection, and edge computing.

Chengxin Niu received the BEng degree from Beijing Jiaotong University, Beijing, China, in 2024. He is currently pursuing the MEng degree at School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China. His research interests include physical layer authentication, Internet of Things, and machine learning.

Xin Fan received the BEng, MEng, and PhD degrees from Beijing Jiaotong University, Beijing, China, in 2016, 2018, and 2023, respectively. He was a visiting PhD student at Department of Electrical and Computer Engineering, George Mason University, Fairfax, VA, USA, from 2020 to 2022. He is currently an assistant professor at School of Information Science and Technology, Beijing Forestry University, Beijing, China. He is also with the Engineering Research Center for Forestry-Oriented Intelligent Information Processing of National Forestry and Grassland Administration, Beijing, China. His current research interests include wireless communications, machine learning, security and privacy, optimization, statistical signal processing, and blockchain.

Jian Mao received the BS and PhD degrees from Xidian University, Xi'an, China, in 1997 and 2004, respectively. She is currently a professor at School of Cyber Science and Technology, Beihang University, Beijing, China. She is also with the Tianmushan Laboratory, Hangzhou, China, and the Zhongguancun Laboratory, Beijing, China. Her research interests include IoT security, web security, and mobile security.

Tao Jing received the MS and PhD degrees from Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China, in 1994 and 1999, respectively. He is currently a professor at School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China. His current research interests include resource allocation in wireless communication networks, such as IoT, industrial IoT, and Internet of vehicles.