IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 31, 2016, accepted January 16, 2017, date of publication February 2, 2017, date of current version March 28, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2662940

Utility-Privacy Tradeoff Based on Random Data Obfuscation in Internet of Energy

ZHITAO GUAN[1], (Member, IEEE), GUANLIN SI[1], JUN WU[2], (Member, IEEE), LIEHUANG ZHU[3], (Member, IEEE), ZIJIAN ZHANG[3], (Member, IEEE), AND YINGLONG MA[1], (Member, IEEE)

[1] School of Control and Computer Engineering, North China Electric Power University, Beijing, China
[2] College of Information Security Engineering, Shanghai Jiao Tong University, Shanghai, China
[3] School of Computer, Beijing Institute of Technology, Beijing, China

Corresponding author: J. Wu (junwuhn@sjtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61402171, in part by the China National Key Research and Development Program under Grant 122016YFB0800301, and in part the Fundamental Research Funds for the Central Universities under Grant 2015ZD12 and Grant 2016MS29.

ABSTRACT Internet of Energy is considered as a promising approach to solve the problems of energy crisis and carbon emission. It needs to collect user's real-time data for optimizing the energy utilization. However, such data may disclose user's privacy information. Previous works usually adopt specific obfuscation value to mask user's data and counteract the deviation through data aggregation; these works can preserve the data privacy effectively, but most of them consider less about the data-utility (precision). In this paper, we propose a utility-privacy tradeoff scheme based on random data obfuscation in Internet of Energy. In the proposed scheme, we adopt random data-obfuscation to mask the real-time data and realize the fault-tolerance during data aggregation, and the random obfuscation value obeys the Laplace distribution. We use the signal-to-noise ratio to quantify the level of utility; we measure the level of privacy through information entropy. Based on these two Indicators, we balance the utility-privacy tradeoff by calculating the optimal parameters of the Laplace distribution. The analysis shows that our scheme can meet the security requirement, and it also has better performance than that of other popular methods.

INDEX TERMS Internet of energy, utility-privacy tradeoff, random data obfuscation, fault-tolerance.

I. INTRODUCTION

Internet of Energy is a pluralistic energy network. This complex energy network takes power system as the core, and combines natural gas, transportation and some other systems by the technologies of Internet and renewable energy generation [1]. As shown in figure.1, Internet of Energy can be divided into energy network and information network Energy from various users turns into electricity and interacts with the power plant through the energy transfer network and the system model of information network likes the Internet of things [2] to some degree. As the issues of environmental pollution and energy crisis are becoming increasingly serious, Internet of Energy that supports the large-scale use of renewable energy sources has been given broad intensive attention. Internet of Energy underlines the use of Internet and various distributed energy comparing with smart grid [3], therefore, Internet of Energy must pay attention to various energy management [4] and big data analysis [5].

There are various users in Internet of Energy such as distributed energy providers, energy consumers, and electric vehicle users and so on. The control center needs to collect data related to the energy consumption from each kind of users to make the power generation plan and determine the efficient energy trading scheme [6]. However, such data may disclose user's privacy information [7]. For example, an adversary can deduce the company's energy production efficiency, production cycle and other commercial secrets by analyzing the real-time bidding data of the distributed energy provider. The adversary can also infer user's house behavior and living habits by observing the real-time power consumption curve of the specific user. What's more, the adversary can also learn of the user's recent outdoor activities and planning arrangements by analyzing the real-time battery charging and discharging status of electric vehicle or position message.

Therefore, how to design a secure privacy-aware data collection scheme is one of the most popular research topics

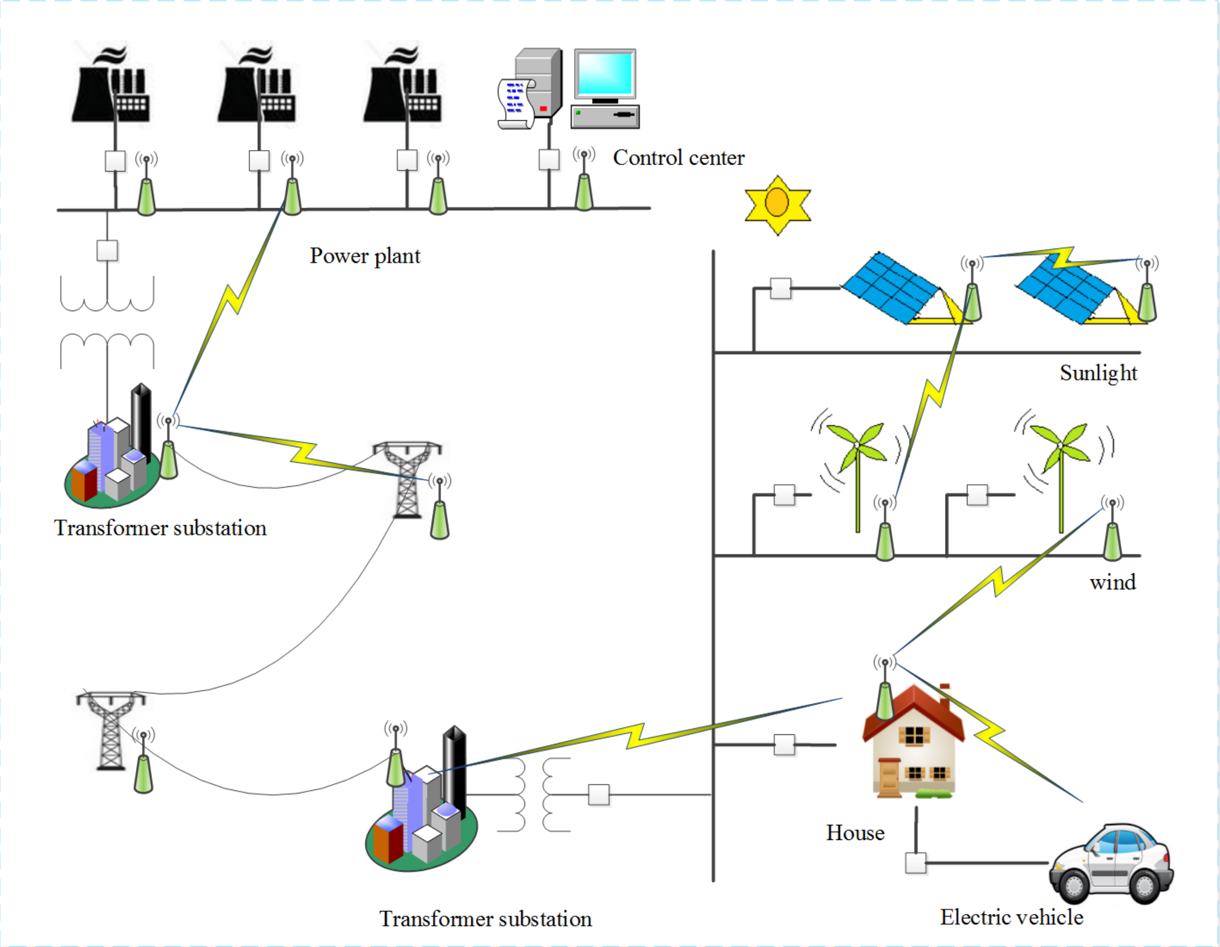


FIGURE 1. The architecture of internet of energy.

in Internet of Energy. Because Internet of Energy underlines the use of Internet and various distributed energy, privacy problems in Internet of Energy are more complex than smart grid. Existing research methods can be divided into two main categories.

The first category is protecting user's identity through anonymity or pseudonym. User's attributes can be classified into identity information, quasi-identifier, and sensitive information. Given an anonymity table, if the attributes in the table have not been properly treated, an adversary may deduce the relationship between user's identity and sensitive information according to the user's quasi-identifier such as the age and gender. Although there are k-anonymity algorithm [8] and l- diversity algorithm [9] to address the disadvantages of identity-protection scheme, it is very difficult to find a credible party to complete the secure anonymity work.

The second category is using data aggregation to protect the user's real-time data, in which the homomorphic encryption and data-obfuscation are used. In fact, these two methods are often used together. Data-obfuscation can be divided into non-random data obfuscation and random data-obfuscation.

Non-random data obfuscation refers that the obfuscation value of each user is assigned by the trusted third party so that the sum of all obfuscation values after data aggregation is zero. However, when there happens the data loss that may be caused by the malfunction of a user's meter, the result of data aggregation will be wrong. Therefore, data processing scheme which supports fault-tolerance is necessary [10].

Random data-obfuscation means that the obfuscation value of each user is generated by themselves. Since the random obfuscation value generated by each user is irrelevant, even if a meter is malfunctioning, it will not impact the normal data aggregation operation. Therefore, the computational burden will decrease for taking no account of fault-tolerance.

However, there is a tradeoff between data-utility and data privacy. For data-utility, final obfuscated curve should share the same trend with the original load curve. For data privacy, data can be only accessed by authorized users with fine-grained policies. If the random obfuscation value chosen by the user is unreasonable, the data-utility will be greatly reduced. Therefore, designing a reasonable obfuscation value generation algorithm is the key of the whole scheme, which is the focus of our paper.

We summarize our contributions as follows:

1. We adopt random data-obfuscation to mask the real-time data and realize the fault-tolerance during the data aggregation, and the random obfuscation value obeys the Laplace distribution.
2. We use the signal-to-noise ratio to quantify the level of utility; we measure the level of privacy through information entropy. Based on these two Indicators, we resolve the utility-privacy tradeoff by calculating the optimal parameters of the Laplace distribution.
3. We prove the feasibility of our scheme, analyze the relevant parameters through simulation experiment and compare with other similar schemes in computational cost and error rate.

The rest of this paper is organized as follows. Section 2 introduces the related work. In section 3, some preliminaries are given. Section 4 shows the system model and design goals. In section 5, our scheme is stated. In section 6, security analysis is given. In Section 7, the performance of our scheme is evaluated. In Section 8, the paper is concluded.

II. RELATED WORK

We elaborate related works about privacy-preserving in Internet of Energy from two aspects. One is the specific strategy aimed at protecting user's privacy; the other one is aimed at resolving related problems caused by privacy-preserving.

For the specific strategy aimed at protecting user's privacy, we can divide it into three categories.

The first category is the scheme based on household battery [11]–[13]. Internet of Energy and the household battery provide users with electricity at the same time. When the household consumption curve goes high, the battery discharges. Otherwise, it charges. In this way, we can hide the user's real-time data to protect user privacy. The disadvantage is that charging and discharging the battery may collide with dynamic electricity price.Jiyun Yao and Parv Venkitasubramaniam [14] try to use the revealing state approach and rate distortion bounds to realize the optimal tradeoff between privacy and cost savings, but it is still limited to the capacity of battery.what's more, the battery of electric vehicle is often used to facilitate the demand response, Weifeng Zhong *et al.* [15] propose a battery strategy on stability and robustness of demand response.

The second category is protecting user's identity through anonymity [16] or pseudonym [17]. User's attributes can be classified into identity information, quasi-identifier, and sensitive information. Given an anonymity table, if the attributes in the table have not been properly treated, an adversary may deduce the relationship between user's identity and sensitive information according to the user's quasi-identifier such as the age and gender. Although there are k-anonymity algorithm [8] and l- diversity algorithm [9] to address the disadvantages of identity-protection scheme, finding a credible party to complete the secure anonymity work is very difficult. Virtual ring is also a common solution to mask the user's identity and Jun Long *et al.* [18] propose an energy-efficient and sink-location privacy enhanced scheme through ring based routing to mask the identity and route.

The third category is using data aggregation to protect the user's real-time data. The common solutions contain homomorphic encryption [19], [20] and data obfuscation [21]. In fact, these two solutions are often used together. Data obfuscation can be divided into non-random data obfuscation and random data-obfuscation. For the non-random data obfuscation, if there happens the data loss caused by the malfunction of a user's meter, the result of data aggregation will go wrong, and we have to face the problem about fault-tolerance. Le Chen *et al.* [22] try to use the third party to realize the fault-tolerance, but the security of the third party is difficult to guarantee. Zhiguo Shi *et al.* [23] propose the DG-APED scheme to resolve the problems caused by malfunctioning *SM*s, which is based on grouping. When there are several malfunctioning *SM*s, it will aggregate the data by grouping, and drop the group which contains the damaged *SM*. However, the error rate is not ideal and searching the damaged member needs to spend much computational cost. Song Han *et al.* [24] propose the PPM-HDA to support fault-tolerance, but using Pollard's lambda method to compute the discrete logarithm is complex. Jongho Won *et al.* propose a fault-tolerance scheme based on future ciphertext [25], but the effect is not very ideal because of the limited buffering. Jin He *et al.* [26] achieve fault-tolerance among virtual middlebox failure for cloud computing security. Besides, differential privacy is also an important problem during the data aggregation and many schemes [27]–[29] have been proposed to realize the differential privacy.

For the strategy aimed at resolving related problems caused by privacy-preserving, we can divide it into four categories.

The first category is about the tradeoff between privacy-preserving and limited computing ability [30]–[32]. For any cryptographic-based encryption scheme, there is no doubt that the complexity of the encryption algorithm comparing to original one will increase. Theoretically, the better is the level of privacy-preserving, the higher is the computational cost. However, smart meter installed on the user side has limited computing ability and can't run overly complex encryption algorithm. Asmaa Abdallah *et al.* [33] propose a light-weight scheme based on *NTRU* to encrypt the data during message transmission.

The second category is about the tradeoff between privacy-preserving and security authentication [34]. Privacy-preserving and security authentication are two related problems. Theoretically, the better is the level of privacy-preserving, the more difficult is security authentication.

The third category is about the tradeoff between privacy and billing [35], [36]. As far as we know, data collected from users can be divided into two kinds. One kind is the real-time data for creating power generation plan and dynamic price; the other kind is non-real-time data which is used for billing. However, to some degree, the non-real-time data for billing may also be related to user's privacy, but protecting these

TABLE 1. Related work.

Privacy-preserving	Household battery	[11],[12],[13],[14],[15]
	Identity-preserving	Anonymity [16] Pseudonym [17] k-anonymity[8] L- diversity [9] Virtual ring [18]
	Data aggregation	Homomorphic encryption[19][20] Data obfuscation[21] Fault-tolerance [22],[23],[24],[25],[26] Differential privacy [27], [28], [29]
Related-problem	Lightweight computing	[30], [31], [32], [33]
	Authentication	[34]
	Privacy and billing	[35], [36], [37]
	Privacy and utility	[38], [39],[40] [41]

data may impact normal billing. Lei Yang *et al.* [37] turn this problem into an equivalent problem, which can be solved by using only the current observations. However, it is still limited to the battery capability.

The fourth category is about the tradeoff between privacy and utility [38], [39], [41]. If users adopt the random data-obfuscation scheme to protect their electricity consumption data, when the random obfuscation value chosen by user is unreasonable, it will impact the accuracy of aggregated result and reduce the utility of the data. Therefore, designing a reasonable random number generation algorithm to set obfuscation value for each user is necessary. Lalitha Sankar *et al.* [40] use the theory of rate distortion to quantify the tradeoff between the utility (mean square distortion) and privacy (information leakage).

In table. 1, the summary of the related work is given. Our scheme belongs to the tradeoff between privacy and utility and can also realize fault-tolerance in a lightweight way. Comparing with other schemes for the utility-privacy tradeoff such as [38] and [39], we propose two new indicators to measure the utility (*SNR*) and privacy (Information entropy). Besides, we also calculate the value of related parameters through simulation experiment. Comparing with fault-tolerance schemes such as [23] and [24], our scheme can realize fault-tolerance in a lightweight way.

III. PRELIMINARIES

In table. 2, the notations used in preliminaries are listed.

A. PAILLIER CRYPTOSYSTEM

Paillier cryptosystem is an asymmetric encryption algorithm, which has the additive homomorphism property. It includes three procedures: key generation, encryption and decryption.

1) Key generation: Chooses two prime numbers p, q with the same length and calculates $n = pq$. g is a generator of cyclic group $Z^*_{n^2}$, and $\gcd(L(g^\lambda \bmod n^2), n) = 1$. The public key is (n, g), and the private key is λ.

$$\lambda = lcm(p-1, q-1) \tag{1}$$

TABLE 2. Notions in preliminaries.

Acronym	Descriptions
SM	Smart meter
CC	Control center
N	The number of SMs
$\sigma^{2}{}_{s}$	The variance of signal
$\sigma^{2}{}_{N}$	The variance of noise
DA	Data aggregation device
KIC	Key initialization center
SNR	Signal-to-noise ratio
$H(b)$	Information entropy

2) Encryption phase: For the plain text $m \in Z_n$ we can select a random number $r < n$. Then, the ciphertext can be calculated as follows:

$$C = g^m r^n \bmod n^2 \tag{2}$$

3) Decryption phase: After receiving the cipher text C, the receiver can get the plain text m with the secret key λ by the following formula

$$m = \frac{L(C^\lambda \bmod n^2)}{L(g^\lambda \bmod n^2)} \bmod n \tag{3}$$

B. SIGNAL-TO-NOISE RATIO

Signal-to-noise ratio (*SNR*) is a common ratio and it is often used to measure the performance of electronic system. *SNR* is calculated as follows:

$$SNR = 10\lg \frac{P_S}{P_N} \tag{4}$$

P_S represents the power which produces the normal signal and P_N represents the power which produces the noise. The higher is the *SNR*, the stronger is the signal. In this paper, we take the entropy power of normal *SM*s' data as P_S and take the entropy power of obfuscation value as P_N. For convenience, we take e as the base of the logarithmic function. Then, we can measure the error rate of our scheme through *SNR*.

$$SNR = 10\ln \frac{P_s}{P_N} = 10\ln \frac{\sigma_s^2}{\sigma_N^2} \tag{5}$$

C. INFORMATION ENTROPY

Information entropy is a concept in information theory, which is used to measure the level of system chaos. The larger is the information entropy, the higher is the level of system chaos. The nature of privacy-preserving is to increase the level of system chaos. Therefore, we can use information entropy to measure the level of privacy-preserving.

For the information entropy of continuous variables, we have

$$H(x) = -\int_{-\infty}^{+\infty} f(x)\log_2 f(x)dx \tag{6}$$

$f(x)$ denotes the probability density function of variable x.

D. LAPLACE DISTRIBUTION

The Laplace distribution is a continuous probability distribution in probability theory. The probability density function of variable obeying Laplace distribution can be described as follows:

$$f(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}} \tag{7}$$

The expectation and variance are calculated as follows:

$$E(x) = \mu \tag{8}$$

$$\sigma^2 = 2b^2 \tag{9}$$

μ is the position parameter, which has the highest probability of occurrence. b is the scale parameter and it affects the ranges as the figure.2. The greater is the value of b, the larger is the variance. In this paper, we adopt random data-obfuscation which obeys the Laplace distribution to mask user's real-time data, and the level of privacy-preserving and level of data-utility are related to the value of b.

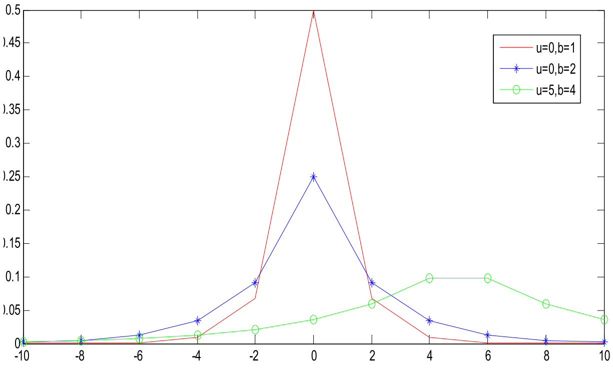


FIGURE 2. Probability density function of Laplace distribution.

IV. MODELS AND GOALS

As shown in figure.1, energy from various users turns into electricity and interacts with the power plant through the energy transfer network. Real-time data collected from different users transports to the control center through the data transmission network. The system model of information network is showed as follows:

A. SYSTEM MODEL

The information network in our paper is comprised of the control center (*CC*), the key initialization center (*KIC*), the data aggregation device (*DA*), and various users.

1) Users: We divide all the users into distributed energy providers, energy consumers and electric vehicle users. They all need to upload their real-time data to the control center for the energy optimization through smart meter(*SM*). As the real-time data is related to user's privacy, the data must be encrypted by the *SM* before sending to the control center.

2) Data aggregation device: The data aggregation device is responsible for collecting all the data sent by *SM*s, calculating the sum of real-time data by running the homomorphic algorithm and uploading the sum to the control center.

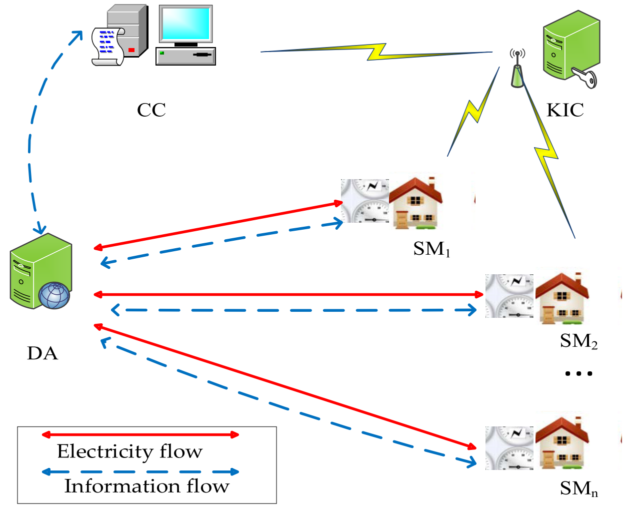


FIGURE 3. System model.

3) Key initialization center: The key initialization center is responsible for initializing all of the keys for *SM*s and *CC*. It publishes the encryption parameters by running the paillier cryptosystem and sends the decrypted key to *CC*. In particular, the key initialization center is also responsible for setting reasonable parameters for the random obfuscation function.

4) Control center: The control center can acquire the summary of real–time data from *DA*. With these data, control center can get the trend of power consumption and create the power generation plan or dynamic price immediately.

B. ADVERSARIAL MODEL

We assume that smart meter installed on the user side is vulnerable to external attacks. The communication channel is not secure and adversary may eavesdrop on the channel. *CC* and *DA* are honest-but-curious. That is to say, they do not destroy or modify user's data, but always attempt to snoop the user's private information through the background knowledge. What's more, *CC* may conspire with several smart meters to increase the probability of successful attack.

C. DESIGN GOALS

Considering the above scenarios, our design goals can be divided into three aspects.

1) Privacy-preserving: A residential user's data is inaccessible to any other users. The outside adversary, *DA* or *CC* should not acquire the real-time data of users even if they try to conspire with each other.

2) Data-utility: We adopt data aggregation based on random data-obfuscation to protect user's electricity consumption data. However, if the random obfuscation value chosen by user is unreasonable, the accuracy of aggregated result will decrease largely. Therefore, we need to guarantee the data-utility during the data-obfuscation.

3) Fault-tolerance: For the conentional data aggregation scheme, if there is a malfunctioning *SM*, the data aggregation can't run in the right way. Therefore, we must ensure our

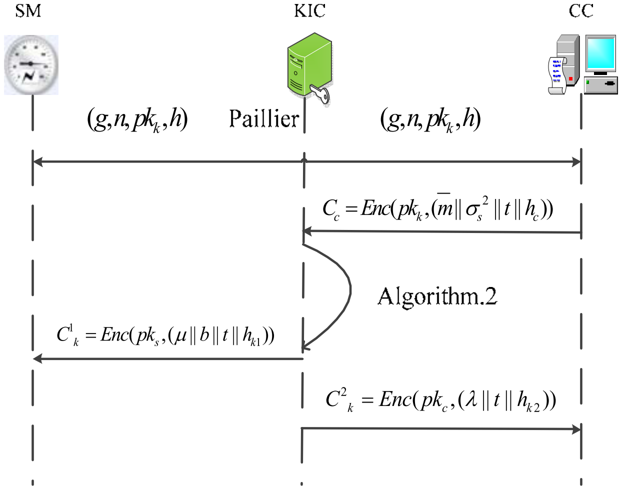

SM
KIC
CC
(g,n,pk_k,h)
Paillier
(g,n,pk_k,h)
$C_c = Enc(pk_k,(\overline{m} \| \sigma_s^2 \| t \| h_c))$
Algorithm.2
$C^1_k = Enc(pk_s,(\mu \| b \| t \| h_{k1}))$
$C^2_k = Enc(pk_c,(\lambda \| t \| h_{k2}))$

FIGURE 4. System initialization.

data aggregation can run normally, even if there are several malfunctioning *SM*s in a user group.

V. OUR SCHEME

A. SYSTEM INITIALIZATION

In table. 4, the notations used in system initialization are listed. As figure.4 shows, firstly, the *KIC* generates (g, n) by running the paillier encryption algorithm and creates (pk_k, sk_k) by running *RSA*. It chooses two big primes pq, and computes $n = pq$. g is an integer chosen from $Z^*_{n^2}$, where $\gcd(L(g^\lambda \bmod n^2), n) = 1$ and $L(x) = \frac{x-1}{x}$. $(pk_c, sk_c)(pk_k, sk_k)(pk_d, sk_d)$ are the public and private keys for *CC*, *KIC* and *DA* by running *RSA*. (pk_{si}, sk_{si}) represents the public and private key of SM_i, which is created by the valid manufacturer. $h : \{0, 1\}^* \rightarrow \{0, 1\}^\Delta$ represents a classical hash function and Δ denotes the length of message digest. Then, the *KIC* publishes (g, n, pk_k, h) using a common channel, and calculates $\lambda = lcm(p - 1, q - 1)$ as the decrypted key.

Secondly, when *KIC* receives the message C_c, it will create reasonable parameters μ and b by running the **Algorithm 2**, where C_c denotes the concatenation of the average electricity data $\overline{m}$, variance σ_s^2, timestamp t and hash value h_c, which is encrypted by the *KIC*'s public key pk_k. $h_c = h(\overline{m}||\sigma_s^2||t)$ and $||$ denotes the concatenation operation. $(\overline{m}, \sigma_s^2)$ is calculated by *CC* according to the historical data set. **Algorithm 2** represents the detailed solution to calculate the optimal parameter of Laplace distribution and we will analyze this algorithm below.

Thirdly, *KIC* calculates $h_{k1} = h(\mu||b||t)$ and sends C_k^1 which denotes the concatenation of μ, b, t and h_{k1} encrypted by the *SM*'s public key pk_{si} to *SM*. Similarly, it sends C_k^2 which denotes the concatenation of λ, t and h_{k2} encrypted by the *CC*'s public key pk_c to *CC*. $h_{k2} = h(\lambda||t)$. we show the detailed algorithm in table.3.

B. THE OPTIMAL VALUE OF μ AND b

As we analyze before, the parameters of Laplace distribution μ and b are related to the utility-privacy tradeoff. Reasonable values of μ and b can protect user's privacy while don't impact the data-utility. Therefore, *KIC* needs to calculate the optimal value of μ and b based on the average $\overline{m}$ and variance σ_s^2 in the current time period.

TABLE 3. System initialization.

```
Algorithm 1 System initialization
KIC. Input: $C_c$                    Output: $C^1_k, C^2_k, g, n, pk_k$
(1)publish $(g,n,pk_k) \longleftarrow$ Paillier cryptosystem and RSA
(2)waiting $C_c$ derived from CC
(3) $Dec(sk_k,C_c)$ and verify $h_c$
(4) $(\mu,b) \longleftarrow$ Algorithm 2
(5) $C^1_k = Enc(pk_s,(\mu \| b \| t \| h_{k1}))$; $C^2_k = Enc(pk_c,(\lambda \| t \| h_{k2}))$
(6)send $C^1_k$ to SM; send $C^2_k$ to CC
CC. Input: historical data set   Output: $C_c$
(1) $(\overline{m},\sigma_s^2) \longleftarrow$ historical data set
(2) $h_c = h(\overline{m} \| \sigma_s^2 \| t)$; $C_c = Enc(pk_k,(\overline{m} \| \sigma_s^2 \| t \| h_c))$
(3)send $C_c$ to KIC
```

TABLE 4. Notations in the system initialization.

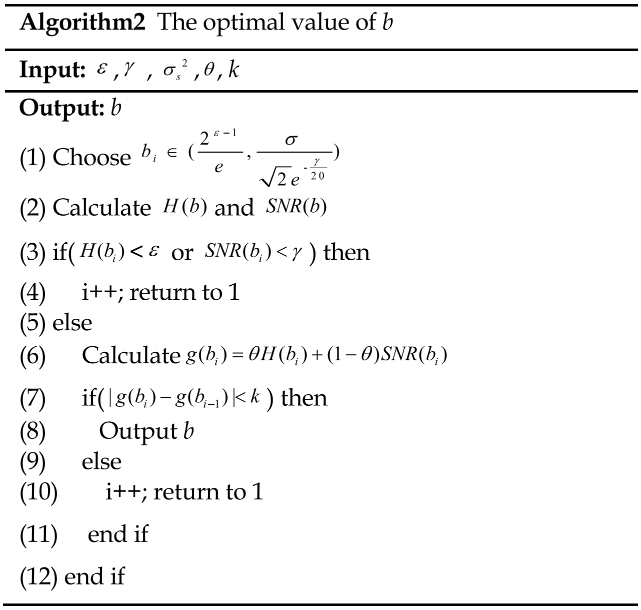
Acronym	Description
t	Time stamp
λ	Decrypted key of *CC*
b	Scale parameter of Laplace distribution
μ	Position parameter of Laplace distribution
h_{k1}	Hash value of message sent from *KIC* to *SM*
h_{k2}	Hash value of message sent from *KIC* to *DA*
h_c	Hash value of message sent from *CC* to *KIC*
C_c	Ciphertext sent from *CC* to *KIC*
$pk_c, pk_d, pk_{si}, pk_k$	Public keys of *CC*, *DA* , SM_i and *KIC*
$sk_c, sk_d, sk_{si}, sk_k$	Private keys of *CC*, *DA*, SM_i and *KIC*

μ denotes the expectation of Laplace distribution and is related to the performance of obfuscation. If the value of μ is too small, the random obfuscation value generated by Laplace distribution can't obfuscate the original real-time data. However, if the value of μ is too large, the random obfuscation value will grow large, which has a bad impact on the final result. To obfuscate the original data while has a lower error rate, we can set $\mu \approx \overline{m}$.

We adopt the information entropy $H(b)$ in the information theory to measure the level of privacy-preserving. The larger is $H(b)$, the higher is level of privacy-preserving. Therefore, $H(b)$ has the lower bound and we can use ε to denote. Then, we calculate the lower bound of b as follows:

$$
\begin{aligned}
H(b) &= -\int_{-\infty}^{+\infty} f(x)\log_2 f(x)dx \\
&= -\frac{1}{2b}\int_{-\infty}^{+\infty} e^{-\frac{|x-\mu|}{b}} \log_2 \frac{1}{2b} e^{-\frac{|x-\mu|}{b}} dx \geq \varepsilon \\
&\Rightarrow b > \frac{2^{\varepsilon-1}}{e}
\end{aligned}
\tag{10}
$$

TABLE 5. The optimal value of *b*.



```
Algorithm2 The optimal value of b
Input: $\varepsilon, \gamma, \sigma_s^2, \theta, k$
Output: b
(1) Choose $b_i \in (\frac{2^{\varepsilon-1}}{e}, \frac{\sigma}{\sqrt{2}e^{-\frac{\gamma}{20}}})$
(2) Calculate $H(b)$ and $SNR(b)$
(3) if( $H(b_i) < \varepsilon$ or $SNR(b_i) < \gamma$ ) then
(4)     i++; return to 1
(5) else
(6)     Calculate $g(b_i) = \theta H(b_i) + (1-\theta) SNR(b_i)$
(7)     if( $|g(b_i) - g(b_{i-1})| < k$ ) then
(8)         Output b
(9)     else
(10)        i++; return to 1
(11)    end if
(12) end if
```

We use the signal-to-noise ratio $SNR(b)$ to measure the error rate which represents the level of data-utility. The larger is $SNR(b)$, the better is level of data-utility. Thus, $SNR(b)$ has the lower bound and we can use γ to denote. Then, we can calculate the higher bound of b as follows:

$$SNR(b) = 10|\ln \frac{\sigma^2}{2b^2}| \geq \gamma \Rightarrow b < \frac{\sigma}{\sqrt{2}e^{-\frac{\gamma}{20}}} \tag{11}$$

We use θ to denote the weight of privacy and $(1-\theta)$ denotes the weight of utility. Generally, we can set $\theta = 0.5$, which means privacy-preserving has the same weight with data-utility. Then, the optimal objective function can be calculated as follows:

$$\begin{aligned} g(b) &= \theta H(b) + (1-\theta) SNR(b) \\ &= -\frac{\theta}{2b} \int_{-\infty}^{+\infty} e^{-\frac{|x-\mu|}{b}} \log_2 \frac{1}{2b} e^{-\frac{|x-\mu|}{b}} dx \\ &\quad + 10(1-\theta) \ln \frac{\sigma^2}{2b^2} \end{aligned} \tag{12}$$

k represents the infinitesimal of difference between $g(b_i)$ and $g(b_{i-1})$, which is set according to the historical data set. $g(b_i)$ and $g(b_{i-1})$ are two values calculated at different times. We analyze the detailed algorithm for the best value of b in table.5.

C. DATA AGGREGATION

In table. 6, the notations used in system initialization are listed.

1) SMART METER

After receiving the parameters of Laplace distribution μ and b, each *SM* generates random obfuscation value x_i following by $laplace(\mu, b)$ and encrypts his real-time data m_i as follows:

$$c_i = g^{m_i + x_i} r_i^n \bmod n^2 \tag{13}$$

TABLE 6. Notations in the optimal value of μ and *b*.

Acronym	Description
θ	Weight of privacy
k	Infinitesimal about $g(b)$
$g(b)$	The optimal objective function
ε	The lower bound of *H(b)*
γ	The lower bound of *SNR(b)*
$\overline{m}$	Average of plaintext(original data)

Then, *SM* calculates the hash value $h_{si} = h(id||c_i||t)$ and encrypts the concatenation of id, c_i, t and h_{si} by the *DA*'s public key pk_d as formula (14). At last, it sends C_{si} to the *DA* through data transmission internet.

$$C_{si} = Enc(pk_d, (id||c_i||t||h_{si})) \tag{14}$$

2) DATA AGGREGATION DEVICE

After receiving the data from all the *SM*s, *DA* decrypts the C_{si} using his private key sk_d and checks h_{si} to guarantee the message integrity. Then, it counts the number of functioning *SM*s denoted by N and multiplies all the encrypted data c_i as follows:

$$\begin{aligned} c_{sum} &= \prod_{i=1}^{N} c_i = g^{\sum_{i=1}^{N}(m_i + x_i)} r_i^n \bmod n^2 \\ &\approx g^{\sum_{i=1}^{N} m_i + N\mu} (r_1 r_2 .. r_N)^n \bmod n^2 \end{aligned} \tag{15}$$

Then, *DA* calculates the hash value $h_D = h(c_{sum}||N||t)$ and encrypts the concatenation of c_{sum}, N, t and h_D by the *CC*'s public key pk_c as formula (16). At last, it sends C_D to the *CC* through data transmission internet.

$$C_D = Enc(pk_c, (c_{sum}||N||t||h_D)) \tag{16}$$

3) CONTROL CENTER

After receiving the data from *DA*, *CC* decrypts the C_D using his private key sk_c and checks h_D to guarantee the message integrity. Then, uses the decrypted key λ to calculate the summary of users' data as follows

$$m_{sum} = \frac{L(C_{sum}^{\lambda} \bmod n^2)}{L(g^{\lambda} \bmod n^2)} \bmod n - N\mu \tag{17}$$

Then, it will calculate the new average $\overline{m_t} = \frac{m_{sum}}{N}$ and compare it with the previous one. If the difference is bigger than β which is a parameter set by *CC*, *CC* will send the new average $\overline{m_t}$ to *KIC* for resetting.

It is worth nothing that $\overline{m_t'}$ is the previous average. If the difference of two averages is smaller than β, *KIC* doesn't have to reset the parameter of Laplace distribution. We show the major process in figure.5.and describe the detailed algorithm in table.8 Related notions during the data aggregation are presented in table.7.

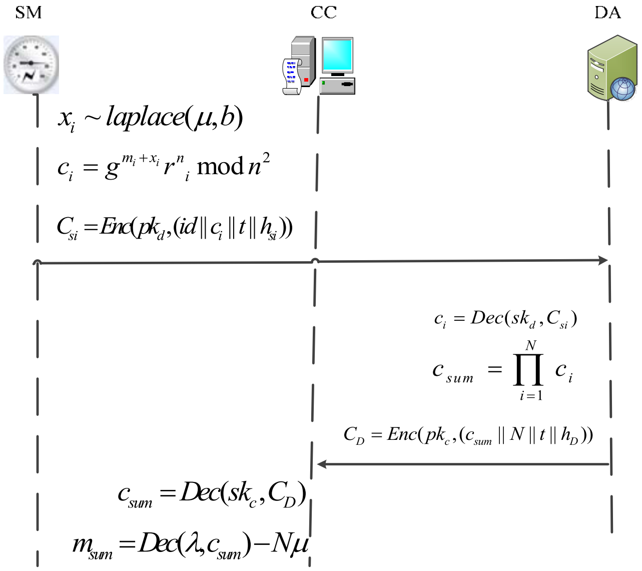

SM
CC
DA
$x_i \sim laplace(\mu,b)$
$c_i = g^{m_i + x_i} r^n{}_i \bmod n^2$
$C_{si} = Enc(pk_d,(id \| c_i \| t \| h_{si}))$
$c_i = Dec(sk_d, C_{si})$
$c_{sum} = \prod_{i=1}^{N} c_i$
$C_D = Enc(pk_c,(c_{sum} \| N \| t \| h_D))$
$c_{sum} = Dec(sk_c, C_D)$
$m_{sum} = Dec(\lambda, c_{sum}) - N\mu$

FIGURE 5. Data aggregation.

TABLE 7. Notions in data aggregation.

Acronym	Description
x_i	Random obfuscation value
C_D	Ciphertext sent from *DA*
C_{si}	Ciphertext sent from SM_i
m_i	Plaintext (original electricity data)
c_i	Ciphertext(encrypted electricity data)
m_{sum}	Sum of *N SM*'s electricity data
C_{sum}	Sum of *N SM*'s encrypted data
h_{si}	Hash value of message sent from SM_i
h_D	Hash value of message sent from *DA*
β	The higher bound of difference between two $\overline{m}$

VI. SECURITY ANALYSIS

In table. 9, the notations used in system initialization are listed.

A. PRIVACY-PRESERVING

To prove the feasibility of our scheme, we analyze the security of our scheme using an adversary game. As *CC* only has the decrypted key λ and the aggregated result C_{sum}, so he can't snoop a single user's privacy information without c_i supplied by *DA*. Similarly, *DA* can't snoop user's privacy information without the help of *CC*. Therefore, in our adversary game, we take the collusion between *DA* and *CC* as the assumed adversary. For the standardization and formalization of our proof, we give a symbolic definition as follows:

A represents the assumed challenger. *B* represents the random oracle. *A* can communicate with *B* to confirm the plaintext space, ciphertext space and encryption algorithm. However, *B* doesn't tell *A* the related key. For this paper, the related key is the random obfuscation value chosen by each user. What's more, *A* knows several context of ciphertext, which means that *CC* or *DA* can get several users' real-time data by observing user's activities or conspiring with several users for the real situation. Then, we can give a definition of privacy challenge for the conspiracy attack:

TABLE 8. Data aggregation.

```
Algorithm 3 Data aggregation
SM. Input: $\mu, b, m_i$  Output: $C_{si}$
(1) $x_i = GenRan(laplace(\mu,b)); r_i = GenRan() \cap \{r_i < n\}$ ;
(2) $c_i = g^{m_i + x_i} r^n{}_i \bmod n^2; h_{si} = h(id \| c_i \| t)$
(3) $C_{si} = Enc(pk_d,(id \| c_i \| t \| h_{si}))$
(4) send $C_{si}$ to $DA$
DA. Input: $C_{si}$    Output: $C_D$
(1) $c_i = Dec(sk_d, C_{si})$ and verify $h_{si}$
(2) $N = count(id_1, id_2, ...id_N); c_{sum} = \prod_{i=1}^{N} c_i; h_D = h(c_{sum} \| N \| t)$
(3) $C_D = Enc(pk_c,(c_{sum} \| N \| t \| h_D))$
(4)send $C_D$ to $CC$
CC. Input: $C_D$   Output: $m_{sum}$
(1) $c_{sum} = Dec(sk_d, C_D)$ and verify $h_D$
(2) $m_{sum} = Dec(\lambda, C_{sum}) - N\mu$
(3) $\overline{\mathrm{m}_t} = \frac{m_{sum}}{N}$;
(4)if( $|\overline{\mathrm{m}_t} - \overline{m_t}'| > \beta$ )then
(5)send $\overline{m_t}'$ to $KIC$ for resetting
```

Definition 1 (Privacy Challenge): The challenger A sends a piece of ciphertext to random oracle B. Then, B decrypts the ciphertext and returns the plaintext to A. Next, A and B repeat above steps several times, and we call this behavior "Cryptography Training". When A is satisfied with the training result, he will send two different messages (m_0, m_1) *to B for encryption. After encrypting the two messages denoted by* $E_k(m_0)$ *and* $E_k(m_1)$*, B toss a coin to confirm the value of* $e \in \{0, 1\}$ *and sends* c^* *to A.*

$$c^* = \begin{cases} E_k(m_0) & e = 0 \\ E_k(m_1) & e = 1 \end{cases} \tag{18}$$

If A can apply a decision algorithm to guess the right value of c^**, we say he wins the challenge. Otherwise, he loses the game, which means our scheme is privacy-friendly.*

Definition 2 (δ-Privacy): The random obfuscation algorithm satisfies δ-Privacy if A doesn't have decision algorithm to confirm the value of c^* *with the probability larger than δ.*

We define a advantage function *Adv* to represent the probability of success, the detailed formula is showed as follows:

$$Adv = |\Pr[0 \leftarrow A(c^* = E_k(m_0)] - \Pr[0 \leftarrow A(c^* = E_k(m_1))]| \tag{19}$$

$\Pr[0 \leftarrow A(c^* = E_k(m_0)]$ represents the probability of correct answer. $\Pr[0 \leftarrow A(c^* = E_k(m_1))]$ represents the probability of wrong answer. Therefore, if *Adv* is close to δ, we have

$$\Pr[0 \leftarrow A(c^* = E_k(m_0)]{=}\Pr[0 \leftarrow A(c^* = E_k(m_1))] \pm \delta \tag{20}$$

Therefore, A is not sure which one is the right answer because of similar probability.

Theorem 1: If random obfuscation values obey laplace(μ, b) during the data aggregation, the scheme can defend the conspiracy attack launched by the honest-but-curious entity.

Proof:

1) According to the relationship between the adversary game and our scheme, we have

$$E_k(m_0) = m_0 + x_0, E_k(m_1) = m_1 + x_1 \tag{21}$$

2) As the random obfuscation values obey the Laplace distribution, $x_i \sim laplace(\mu, b)$. therefore,

$$\begin{aligned}\Pr[0 \leftarrow A(c^* = E_k(m_0)] &= \Pr(|x - x_0| < \zeta)\\ &= \int_{x_0-\zeta}^{x_0+\zeta} \frac{1}{2b} e^{-\frac{|x-\mu|}{b}} dx \end{aligned}\tag{22}$$

$$\begin{aligned}\Pr[0 \leftarrow A(c^* = E_k(m_1))] &= 1 - \Pr[0 \leftarrow A(c^* = E_k(m_0))]\\ &= 1 - \int_{x_0-\zeta}^{x_0+\zeta} \frac{1}{2b} e^{-\frac{|x-\mu|}{b}} dx \end{aligned}\tag{23}$$

Note that ζ denotes the infinitesimal amount of variable x.

3) According to formula (22) and (23), we can calculate *Adv* as follows:

$$\begin{aligned} Adv &= |\Pr[0 \leftarrow A(c^* = E_k(m_0)] - \Pr[0 \leftarrow A(c^* = E_k(m_1))]|\\ &= 1 - 2\int_{x_0-\zeta}^{x_0+\zeta} \frac{1}{2b} e^{-\frac{|x-\mu|}{b}} dx \end{aligned}\tag{24}$$

4) Therefore, we can calculate the parameter

$$\delta = 1 - 2\int_{x_0-\zeta}^{x_0+\zeta} \frac{1}{2b} e^{-\frac{|x-\mu|}{b}} dx. \tag{25}$$

We complete the proof.

B. DATA-UTILITY

As we analyzed before, *SNR* can be used to measure the level of data-utility in the system and it's value is related to the value of b. Therefore, if we set the lower bound of *SNR* when we are calculating the value of b, the random value following *laplace*(μ, b) doesn't impact the data-utility in theory.

Theorm 2: If random obfuscation values obey laplace(μ, b) distribution and SNR $> \gamma$, scheme based on the random data-obfuscation can guarantee the data-utility.

Proof:

1) As the random obfuscation values in our scheme obey the *laplace*(μ, b) distribution, we have

$$E(\mathrm{x}_i) = \mu, \sigma_i^2 = 2b^2 \tag{26}$$

2) After the data aggregation carried by *DA*, we have

$$\begin{aligned} c_{sum} &= \prod_{i=1}^{N} c_i = g^{\sum\limits_{i=1}^{N}(m_i+x_i)} r_i^n \bmod n^2\\ &\approx g^{\sum\limits_{i=1}^{N} m_i + N\mu} (r_1 r_2 .. r_N)^n \bmod n^2 \end{aligned}\tag{27}$$

3) After the decryption carried by *CC* using the decrypted key λ, we have

$$\begin{aligned} m_{sum} &= \frac{L(C_{sum}^{\lambda} \bmod n^2)}{L(g^{\lambda} \bmod n^2)} \bmod n - N\mu\\ &= \sum_{i=1}^{N} m_i + \sum_{i=1}^{N} x_i - N\mu \end{aligned}\tag{28}$$

4) According to the Wiener-khinchin law of large Numbers, if variables are mutal independent and obey the same distribution, we have

$$\lim_{N\to\infty} \Pr\{|\frac{1}{N}\sum_{i=1}^{N} x_i - \frac{1}{N}\sum_{i=1}^{N} E(x_i)| < \tau\} = 1 \tag{29}$$

5) As the random obfuscation values in our scheme are mutal independent and obey the *laplace*(μ, b) distribution, and the amount of *SM*s in Internet of Energy is very huge. Therefore, we have

$$\begin{aligned} &\lim_{N\to\infty} \Pr\{|\frac{1}{N}\sum_{i=1}^{N} x_i - \mu| < \tau\} = 1\\ &\Rightarrow \lim_{N\to\infty} \Pr\{|\sum_{i=1}^{N} x_i - N\mu| < \tau\} = 1\\ &\Rightarrow m_{sum} = \sum_{i=1}^{N} m_i + \tau \end{aligned}\tag{30}$$

6) Because we set the *SNR* as *SNR* $> \gamma$, we limit the range of b. As b is the scale parameter and is related to τ.therefore, the range of τ has been restricted, which means

$$m_{sum} \approx \sum_{i=1}^{N} m_i. \tag{31}$$

We complete the proof.

C. FAULT-TOLERANCE

Data aggregation scheme often involves fault-tolerance. When there is a malfunctioning *SM*, the final result will go wrong.

Theorm 3: If random obfuscation values are mutal independent and obey laplace(μ, b) distribution, scheme based on the random data-obfuscation can realize the fault-tolerance.

TABLE 9. **Notions in Security analysis.**

Acronym	Description
A	Adversary
B	Random oracle
Adv	Advantage function
δ	Infinitesimal about Adv
m_{sum1}	Sum of (*N-K*) *SMs*' electricity data
ζ	Infinitesimal about lapace random number
τ	Infinitesimal parameter of Wiener-khinchin law

Proof: The first 5 steps of the proof are same with that of **Theorm 2**.

6) If there are K malfunctioning *SM*s, we still have

$$\lim_{N\to\infty} \Pr\{|\sum_{i=1}^{N-K} x_i - (N-K)\mu| < \tau\} = 1$$

$$\Rightarrow m_{sum1} = \sum_{i=1}^{N-K} m_i + \tau \tag{32}$$

7) Therefore, when there are several *SM*s damaged, we still have

$$m_{sum1} \approx \sum_{i=1}^{N-K} m_i \quad m_{sum} \approx \sum_{i=1}^{N} m_i \tag{33}$$

We complete our proof

VII. PERFORMANCE EVALUATION

In table. 10, the notations used in system initialization are listed.

A. THE LOWER BOUND OF INFORMATION ENTROPY

The information entropy represents the level of privacy-preserving. Therefore, as the lower bound of information entropy, ε should represent original level of system chaos for the real-time data. We assume that user's real-time data follows the normal distribution, then, ε can be calculated as follows:

$$\varepsilon = -\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_s} e^{-\frac{(x-\mu)^2}{\sigma_s^2}} \log_2 \frac{1}{\sqrt{2\pi}\sigma_s} e^{-\frac{(x-\mu)^2}{\sigma_s^2}} dx \tag{34}$$

In our paper, we use two hundred users' real-time data collected at 00:00 to evaluate μ and σ_s. According to our analysis, $\mu \approx 0.148$, $\sigma_s^2 \approx 0.03$. Therefore, we can calculate the value of $\varepsilon = -0.8424$ and $b > 0.132$.

B. THE LOWER BOUND OF SNR

The *SNR* represents the accuracy of the final aggregated result. To set the lower bound of *SNR*, we need to confirm the relationship between *SNR* and error rate.

Given the data set collected from two hundred users, we calculate *SNR* and error rate by changing the value of b to get different results of obfuscation. We show the relationship between *SNR* and error rate as follows:

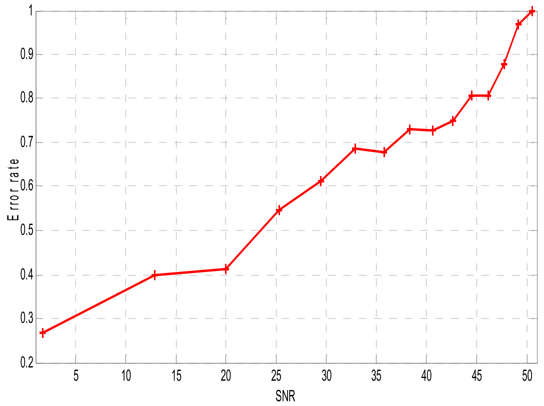


FIGURE 6. **Relationship between error rate and SNR.**

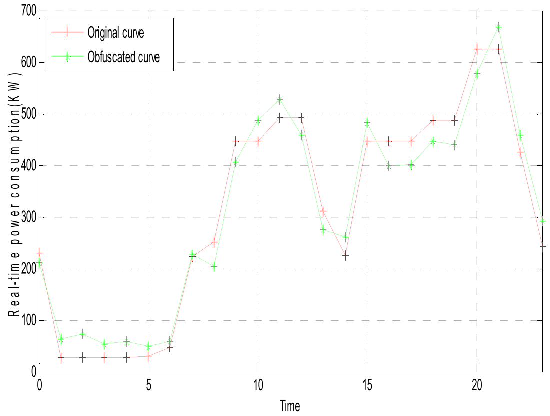


FIGURE 7. **Real-time load curve.**

Through the figure.6, we can set the lower bound of *SNR* according to the acceptable error rate.when the value of *SNR* is close to 50.8849, the error rate will be closed to 1. Therefore, $\gamma \approx 50.8849$ and $b < 1.533$.

C. EXPERIMENTAL EVALUATION

Accroding to the previous analysis based on two hundred users' real-time data, we can draw the real-time load curve and obfuscated load curve as figure.7 from 00:00 to 23:00.

We set the weight of privacy $\theta = 0.5$. For the real-time data at 00:00, we have

$$\varepsilon = -0.8424, \quad \gamma = 50dB, \ \theta = 0.5,$$
$$\mu = \overline{\mathrm{m}} \approx 0.148, \quad N = 200$$

Inputting these parameters into **Algorithm 2,** we can get the optimal value of b which satisfies $b = 1.34$, and the random obfuscation value follows the Laplace distribution $laplace(0.148, 1.34)$ at 00:00. Values of μ and b at other times are calculated in the same way.

As shown in figure.7, we can find the obfuscated load curve is very close to the original load curve, which doesn't deviate from the main trend. Therefore, *CC* can create right power plan and dynamic price according to the main trend in different times, while it has no information of user's real-time data. The tradeoff between data-utiliy and privacy-preserving can be resolved in an optimal way.

Because our scheme can also realize the fault-tolerance, we will compare our scheme with other schemes which satisfy fault-tolerance in computational cost.

PPM-HDA scheme is proposed by Song Han *et al.* [21], which is a novel scheme to realize the fault-tolerance. It adopts Boneh-Goh-Nissim cryptosystem to encrypt the real-time data and use Pollard's lambda method to compute the final discrete logarithm. DG-APED scheme is proposed by Zhiguo Shi [20] to realize the fault-tolerance. It encrypts data based on grouping and drops a group if the group contains malfunctioning *SM*s.

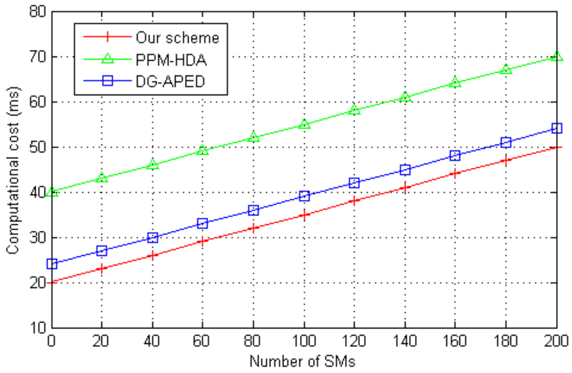


FIGURE 8. Computational cost in normal situation.

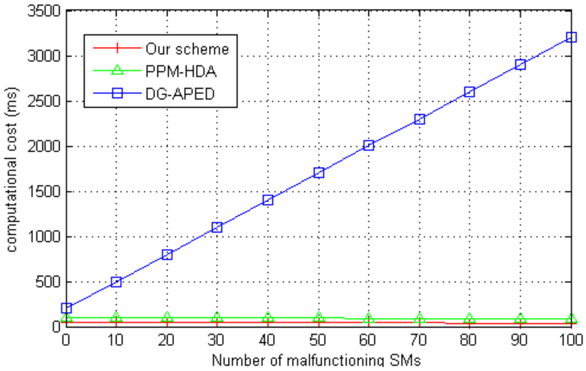


FIGURE 9. Computational cost considering fault-tolerance.

We compare with DG-APED and PPM-HAD for computational cost at figure.8 and figure.9, and show the detailed calculating formulas in table.11. We compare our scheme with DG-APED and PPM-HDA at computational cost under the assumption that all the *SM*s run normally. Then, we compare our scheme with DG-APED and PPM-HDA at computational cost considering the malfunctioning *SM*s. From the two pictures, we can find that our scheme has the smallest computational cost whether or not there are malfunctioning *SM*s.

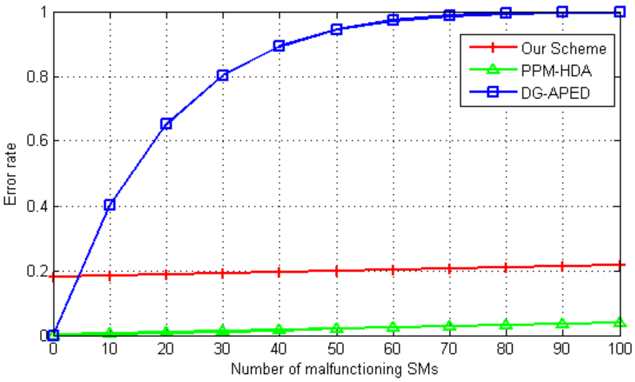


FIGURE 10. The error rate to the ratio of malunctioning SM (b=1.533).

TABLE 10. Notions in performance evaluation.

Acronym	Description
$T_{\exp}$	Time of exponentiation $\approx 1.6ms$
T_{mul}	Time of multiplication $\approx 0.15ms$
T_{ran}	Time of generation of random number $\approx 12ms$
T_{lap}	Time of generation of Laplace parameter $\approx 0.03(N\sigma)^{\frac{1}{2}} e^{\frac{\gamma}{40}}$
T_{pro}	Pollard's lambda method $\approx 0.048\sqrt{N(2^{\alpha}-1)}$
N	Number of *SM*s
K	Number of malfunctioning *SM*s
d	Number of cloud servers
L	Number of *SM*'s types
ω	Number of groups in each type
α	Satisfies $m \in \{0,1,2...2^{a}-1\}$

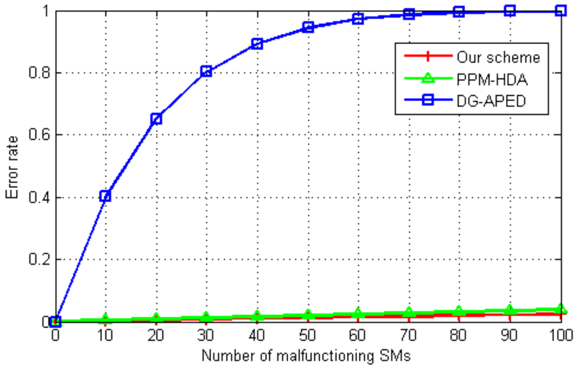


FIGURE 11. Error rate to the ratio of malunctioning *SM* (b=0.132).

In addition, we compare with the other two schemes in the error rate for different numbers of malfunctioning *SM*s and show them according to different values of b. As shown in figure.10, figure.11 and figure.12, with the increase in the number of malfunctioning *SM*s, the error rate of DG-APED keeps increasing and tends to 1 when the number of malfunctioning *SM*s tends to one hundred. However, for our scheme and PPM-HDA, the error rate almost keeps constant. When we set $b = 1.533$, which is the higher bound of b, the error rate of our scheme tends to 0.2. When we set $b = 0.132$, which is the lower bound of b, the error rate of our scheme tends to zero. When we set $b = 1.34$, which is the optimal value of b, the error rate of our scheme tends to 0.1.

Therefore, for the optimal value of b, our scheme has a lower error rate comparing with DG-APED. Although the

TABLE 11. Computational cost.

protocol	Number of *SMs*	Number of malfunctioning *SMs*
Our scheme	$4T_{\exp}+(N+1)T_{mul}+T_{ran}+T_{lap}$	$4T_{\exp}+(N-K+1)T_{mul}+T_{ran}+T_{lap}$
PPM-HDA	$(2d+4)T_{\exp}+(N+d)T_{mul}+2T_{pro}$	$(2d+4)T_{\exp}+(N-K+d)T_{mul}+2T_{pro}$
DG-APED	$7T_{\exp}+(N+5)T_{mul}+T_{ran}+T_{pro}$	$7T_{\exp}+(N+5)T_{mul}+T_{ran}+(\frac{(K+1)\omega}{2}+L)T_{pro}$

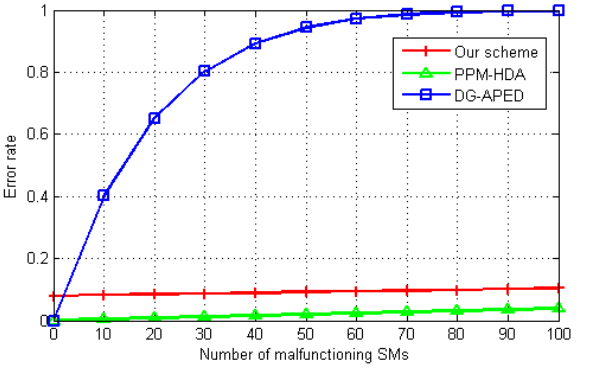


FIGURE 12. Error rate to the ratio of malunctioning *SM* (*b*=1.34).

error rate of our scheme is a little higher than PPM-HDA, this doesn't impact the data-utility shown at figure.7 and our scheme has larger advantages than PPM-HDA in computational cost. Thus, our scheme has better performance than that of other popular methods.

VIII. CONCLUSION

In this paper, we propose a utility-privacy tradeoff method based on random data-obfuscation in Internet of Energy. Random data-obfuscation is adopted to mask the real-time data and realize the fault-tolerance. In addition, we use information entropy to measure the level of privacy-preserving and *SNR* to measure the level of data-utility. Based on these two indicators, we calculate the optimal parameters of Laplace distribution for balancing the utility-privacy tradeoff. At last, we prove the feasibility of our scheme and compare with other fault-tolerance schemes in computational cost and error rate. As we consider less about security authentication in this paper. Therefore, in future, we will work on resolving the tradeoff between privacy and authentication in depth.

REFERENCES

[1] K. Wang *et al.*, "A survey on energy Internet: Architecture, approach, and emerging technologies," *IEEE Syst. J.*, to be published, doi: 10.1109/JSYST.2016.2639820.

[2] M. Dong, K. Ota, and A. Liu, "RMER: Reliable and energy-efficient data collection for large-scale wireless sensor networks," *IEEE Internet Things J.*, vol. 3, no. 4, pp. 511–519, Aug. 2016.

[3] N. Bui, A. P. Castellani, P. Casari, and M. Zorzi, "The Internet of energy: A Web-enabled smart grid system," *IEEE Netw.*, vol. 26, no. 4, pp. 39–45, Jul. 2012.

[4] K. Wang, Z. Ouyang, R. Krishnan, L. Shu, and L. He, "A game theory-based energy management system using price elasticity for smart grids," *IEEE Trans. Ind. Informat.*, vol. 11, no. 6, pp. 1607–1616, Dec. 2015.

[5] H. Jiang, K. Wang, Y. Wang, M. Gao, and Y. Zhang, "Energy big data: A survey," *IEEE Access*, vol. 4, pp. 3844–3861, 2016.

[6] K. Zhang *et al.*, "Incentive-driven energy trading in the smart grid," *IEEE Access*, vol. 4, pp. 1243–1257, 2016.

[7] S. Davies, "Internet of energy," *Eng. Technol.*, vol. 16, no. 5, pp. 42–45, 2010.

[8] P. Samarati and L. Sweeney, "Protecting privacy when disclosing information: *k*-anonymity and its enforcement through generalization and suppression," SRI Int., Tech. Rep., 1998.

[9] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "*L*-diversity: Privacy beyond *k*-anonymity," *ACM Trans. Knowl. Discovery Data*, vol. 1, no. 1, Mar. 2007, Art. no. 3.

[10] K. Wang, Y. Shao, L. Shu, C. Zhu, and Y. Zhang, "Mobile big data fault-tolerant processing for ehealth networks," *IEEE Netw.*, vol. 30, no. 1, pp. 36–42, Jan./Feb. 2016.

[11] G. Kalogridis, C. Efthymiou, S. Z. Denic, and T. A. Lewis, and R. Cepeda, "Privacy for smart meters: Towards undetectable appliance load signatures," in *Proc. 1st IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Oct. 2010, pp. 232–237.

[12] G. Kalogridis, Z. Fan, and S. Basutkar, "Affordable privacy for home smart meters," in *Proc. 9th IEEE Int. Symp. Parallel Distrib. Process. Appl. Workshops (ISPAW)*, May 2011, pp. 77–84.

[13] O. Tan, D. Gunduz, and H. V. Poor, "Increasing smart meter privacy through energy harvesting and storage devices," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1331–1341, Jul. 2013.

[14] J. Yao and P. Venkitasubramaniam, "The privacy analysis of battery control mechanisms in demand response: Revealing state approach and rate distortion bounds," *IEEE Trans. Smart Grid*, vol. 6, no. 5, pp. 2417–2425, Sep. 2015.

[15] W. Zhong, R. Yu, X. Shengli, Y. Zhang, and D. K. Y. Yau, "On stability and robustness of demand response in V2G mobile energy networks," *IEEE Trans. Smart Grid*, to be published.

[16] C. Efthymiou and G. Kalogridis, "Smart grid privacy via anonymization of smart metering data," in *Proc. 1st IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Oct. 2010, pp. 238–243.

[17] X. Tan, J. Zheng, C. Zou, Y. Niu, "Pseudonym-based privacy-preserving scheme for data collection in smart grid," *Int. J. Ad Hoc Ubiquitous Comput.*, vol. 22, no. 2, pp. 120–127, 2016.

[18] J. Long, A. Liu, M. Dong, and Z. Li, "An energy-efficient and sink-location privacy enhanced scheme for WSNs through ring based routing," *J. Parallel Distrib. Comput.*, vols. 81–82, pp. 47–65, Jul. 2015.

[19] F. G. Mármol, C. Sorge, O. Ugus, and G. M. Pérez, "Do not snoop my habits: Preserving privacy in the smart grid," *IEEE Commun. Mag.*, vol. 50, no. 5, pp. 166–172, May 2012.

[20] F. Borges and M. Muhlhauser, "EPPP4SMS: Efficient privacy-preserving protocol for smart metering systems and its simulation using real-world data," *IEEE Trans. Smart Grid*, vol. 5, no. 6, pp. 2701–2708, Nov. 2014.

[21] A. Beussink, K. Akkaya, I. F. Senturk, and M. M. E. A. Mahmoud, "Preserving consumer privacy on IEEE 802.11s-based smart grid AMI networks using data obfuscation," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr. 2014, pp. 658–663.

[22] 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, Nov. 2015.

[23] 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 Trans. Smart Grid*, vol. 6, no. 6, pp. 2856–2868, Nov. 2015.

[24] 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 Trans. Inf. Forensics Security*, vol. 11, no. 9, pp. 1940–1955, Sep. 2015.

[25] 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 Trans. Netw.*, vol. 24, no. 3, pp. 2804–2812, Jun. 2016.

[26] J. He, M. Dong, K. Ota, M. Fan, and G. Wang, "NetSecCC: A scalable and fault-tolerant architecture for cloud computing security," *Peer-to-Peer Netw. Appl.*, vol. 9, no. 1, pp. 67–81, Jan. 2016.

[27] J. Hua, A. Tang, Y. Fang, Z. Shen, and S. Zhong, "Privacy-preserving utility verification of the data published by non-interactive differentially private mechanisms," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 10, pp. 2298–2311, Oct. 2016.

[28] Y. Hong, J. Vaidya, H. Lu, P. Karras, and S. Goel, "Collaborative search log sanitization: Toward differential privacy and boosted utility," *IEEE Trans. Dependable Secure Comput.*, vol. 12, no. 5, pp. 504–518, Sep./Oct. 2015.

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

[30] H. Bao and L. Chen, "A lightweight privacy-preserving scheme with data integrity for smart grid communications," *Concurrency Comput. Pract. Exper.*, vol. 28, no. 4, pp. 1094–1110, Mar. 2016.

[31] H. J. Jo, I. S. Kim, and D. H. Lee, "Efficient and privacy-preserving metering protocols for smart grid systems," *IEEE Trans. Smart Grid*, vol. 7, no. 3, pp. 1732–1742, May 2015.

[32] L. Gu, Y. Pan, M. Dong, and K. Ota, "Noncommutative lightweight signcryption for wireless sensor networks," *Int. J. Distrib. Sensor Netw.*, vol. 9, no. 3, pp. 1–10, 2013.

[33] A. Abdallah and X. Shen, "Lightweight security and privacy preserving scheme for smart grid customer-side networks," *IEEE Trans. Smart Grid*, to be published.

[34] T. W. Chim, S.-M. Yiu, V. O. K. Li, L. C. K. Hui, and J. Zhong, "PRGA: Privacy-preserving recording & gateway-assisted authentication of power usage information for smart grid," *IEEE Trans. Depend. Sec. Comput.*, vol. 12, no. 1, pp. 85–97, Jan./Feb. 2015.

[35] F. Borges, D. Demirel, L. Böck, J. Buchmann, and M. Mühlhäuser, "A privacy-enhancing protocol that provides in-network data aggregation and verifiable smart meter billing," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Jun. 2014, pp. 1–6.

[36] X. Liang, X. Li, R. Lu, X. Lin, and X. Shen, "UDP: Usage-based dynamic pricing with privacy preservation for smart grid," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 141–150, Mar. 2013.

[37] L. Yang, X. Chen, J. Zhang, and H. V. Poor, "Cost-effective and privacy-preserving energy management for smart meters," *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 486–495, Jan. 2015.

[38] F. Knirsch, D. Engel, M. Frincu, and V. Prasanna, "Model-based assessment for balancing privacy requirements and operational capabilities in the smart grid," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, Feb. 2015, pp. 1–5.

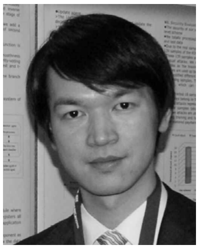
[39] M. Savi, C. Rottondi, and G. Verticale, "Evaluation of the precision-privacy tradeoff of data perturbation for smart metering," *IEEE Trans. Smart Grid*, vol. 6, no. 5, pp. 2409–2416, Sep. 2015.

[40] L. Sankar, S. R. Rajagopalan, S. Mohajer, and H. V. Poor, "Smart meter privacy: A theoretical framework," *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 837–846, Jun. 2013.

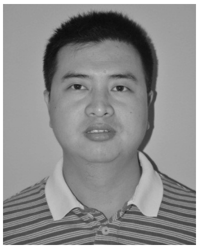
[41] J. Long, M. Dong, K. Ota, A. Liu, and S. Hai, "Reliability guaranteed efficient data gathering in wireless sensor networks," *IEEE Access*, vol. 3, pp. 430–444, May 2015.

ZHITAO GUAN (M'13) received the B.Eng. and Ph.D. degrees in computer application from the Beijing Institute of Technology, China, in 2002 and 2008, respectively. He is currently an Associate Professor with the School of Control and Computer Engineering, North China Electric Power University. His current research interests include smart grid security, wireless security, and cloud security. He has authored over 20 peer-reviewed journal and conference papers in these areas.

GUANLIN SI received the B.Eng. degree from Hebei Agricultual University in 2015. He is currently pursuing the master's degree with the School of Control and Computer Engineering, North China Electric Power University. His current research focuses on smart grid security.



JUN WU (M'08) received the Ph.D. degree from the Global Information and Telecommunication Studies, Waseda University, Japan. He was a Post-Doctoral Researcher from the Research Institute for Secure Systems, National Institute of Advanced Industrial Science and Technology, Japan, from 2011 to 2012. He was a Researcher from the Global Information and Telecommunication Institute, Waseda University, from 2011 to 2013. He is currently an Associate Professor of Electronic Information and Electrical Engineering with Shanghai Jiao Tong University, China. His research interests include the advanced computation and communications techniques of smart sensors, wireless communication systems, industrial control systems, wireless sensor networks, and smart grids. He has hosted and participated in several research projects for the National Natural Science Foundation of China, National 863 Plan, and 973 Plan projects. He has been a Guest Editor of the IEEE Sensors Journal and a TPC Member of several international conferences, including WINCON 2011 and GLOBECOM 2015.



LIEHUANG ZHU is currently a Professor with the Department of Computer Science, Beijing Institute of Technology. He is selected into the Program for New Century Excellent Talents in University from the Ministry of Education, China. His current research interests include Internet of Things, cloud computing security, and Internet and mobile security.

ZIJIAN ZHANG was a Visiting Scholar with the Computer Science and Engineering Department, State University of New York at Buffalo, in 2015. He is currently an Assistant Professor with the Department of Computer Science, Beijing Institute of Technology. His current research interests include smart grid, data privacy, and mobile security.

YINGLONG MA received the Ph.D. degree in computer science from the Institute of Software Chinese Academy of Sciences in 2006. He was a Visiting Scholar with the School of Computer Science, Georgia Institute of Technology, in 2010. He is currently an Associate Professor with the School of Control and Computer Engineering, North China Electric Power University. His current research interests include knowledge engineering and electric power information processing.