Privacy-Preserving and Traceable Data Aggregation in Energy Internet

Yue Zhang and Zhitao Guan(✉)

North China Electric Power University, Beijing 102206, China
guan@ncepu.edu.cn

Abstract. Energy Internet 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. Edge nodes like *GWs* (gateway) are used for data aggregation to improve the efficiency of the system. Due to a large number of *GWs* are widely distributed and difficult to be managed, which brings potential security threats for the Energy Internet. Existing data aggregation schemes fails in preventing the adversary from controlling or destroying *GWs*. In this paper, we propose an IBE-based Device Traceable Privacy-Preserving Aggregation Scheme, named IBE-DTPPA. Increasing the *RA* (Residential Area) users' data aggregation integrity verification by BGN Cryptosystem; using IBE Cryptosystem to encrypt aggregation data, calculating ciphertext based on *GW's* dynamic ID, realizing the target *GW* traceability; choosing *CC* (Control Center) dynamic identity information as public key to realize *CC* authentication, preventing adversary from using *CC*'s identity fraudulently. Through extensive analysis, we demonstrate that IBE-DTPPA resists various security threats, and can trace target *GW* efficiently.

Keywords: Device tracking · Authentication · Data aggregation
Energy Internet

1 Introduction

Energy Internet as a pluralistic energy network [1], as the issues of environmental pollution and energy crisis are becoming increasingly serious, Energy Internet supports the large-scale use of renewable energy sources, which has been given broad intensive attention. Energy Internet can be divided into energy network and information network. Energy generated from various users turns into electricity and interacts with the power plant through the energy transfer network and information network, as shown in Fig. 1. Compared with smart grid, the Energy Internet can make full use of the various types of distributed energy [2], so the energy management and real-time data analysis are important in Energy Internet [3, 4]. In the Energy Internet, the scope of system data collection will be expanded greatly, SMs and a variety of smart appliances will be used as collection devices to upload nearly real-time periodically, however, frequently electricity usage data collection may bring user sensitive information leakage and other issues, which threaten user privacy [5], and calculation cost and communication overhead bring much pressure to the system. Using data aggregation [6, 7] not only

L. Zhu and S. Zhong (Eds.): MSN 2017, CCIS 747, pp. 435–449, 2018.
https://doi.org/10.1007/978-981-10-8890-2_33

reduces communication overhead but also protects individual data privacy. Most of the existing aggregation schemes use homomorphic encryption to encrypt users' data, device like SM (smart meter) encrypts data and aggregates in edge nodes in communication network without decryption, which can reduce the communication overhead and calculation cost for other entities, improving system efficiency, as show in Fig. 1.

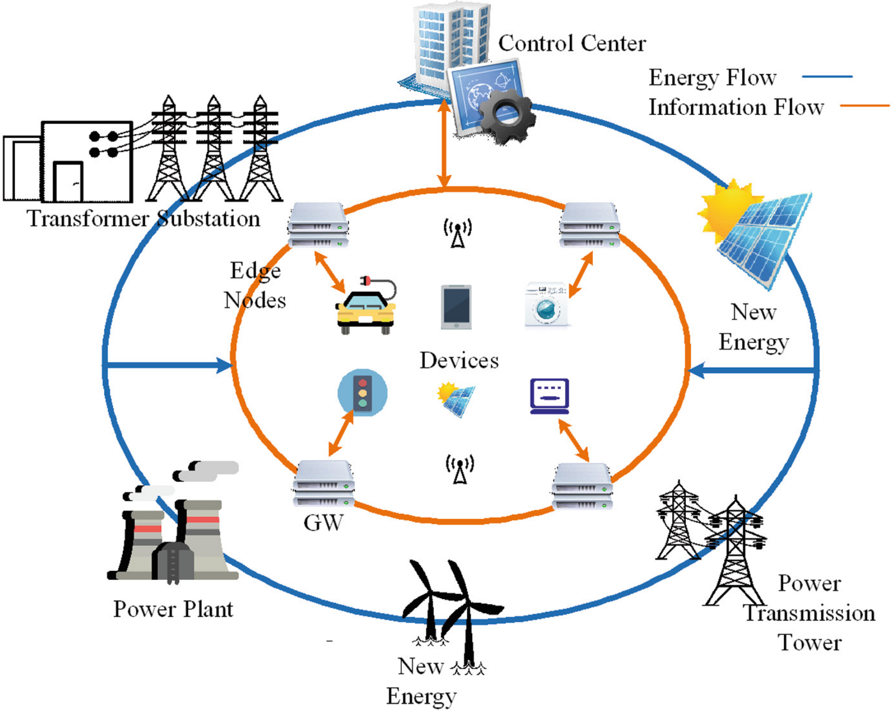

Fig. 1. Energy Internet system architecture

Due to a large number of GWs are widely distributed in the RA, it is difficult to manage, vulnerable to be destroyed or controlled by the adversary, resulting in the error aggregation data will be transmitted to CC, improper power generation plan or dynamic price will reduce system reliability, as shown in Fig. 2, then how to trace the target gateway in time to ensure Energy Internet reliability, which is still a problem. In addition, CC's identity is vulnerable to be used fraudulently by the adversary, which may cause user privacy disclosure. To solve above problems, in this paper, we propose IBE-based Device Traceable Privacy-Preserving Aggregation Scheme based on IBE (IBE-DTPPA).

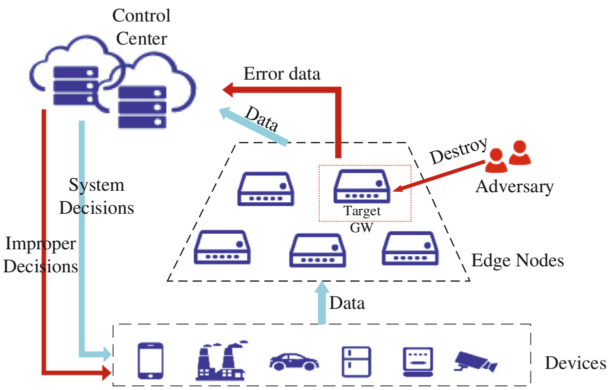

Fig. 2. Energy Internet edge equipment threat

A security-efficient, supporting target GW traceability. The main contributions of this paper are divided into three parts as follows:

(1) We add a random number in aggregation in BGN Cryptosystem to realize verification of aggregation data integrity. Choose CC's dynamic ID as public key IBE encryption, ID updates aperiodically in short period of time, ensuring the authenticity of CC's identity.
(2) We encrypt the RA aggregation data by IBE Cryptosystem, calculating ciphertext based on GW's dynamic ID, realizing the target GW traceability.
(3) We prove the security of our scheme, analyze the relevant parameters through detailed analysis, proving our scheme is secure against different attacks and can realize device traceability efficiently.

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

2 Related Work

Existing data aggregation schemes have a common concern, individual user's privacy-sensitive data should not be exposed. The common solutions to realize data aggregation contain homomorphic encryption [8] and data obfuscation [9]. However, the selection of parameters in data obfuscation is a difficult task. Therefore, homomorphic encryption has been widely used. Existing schemes use a homomorphic encryption to encrypt user's privacy-sensitive data and the edge nodes like gateway in the Energy Internet can aggregate all user's data without decryption, Przydatek et al. propose a specific framework for secure data aggregation in distributed energy environment, although Przydatek et al.'s framework could provide efficient data aggregation, the data privacy still needs to be improved. To address the individual user privacy issue in data aggregation, Shi et al. [10] propose a scheme to aggregate time-series data, which allows a group of collection devices upload the encrypted user's data to the aggregator periodically, and aggregate the data without disclosing any information. Homomorphic hash function [11] has been used to authenticate SM and CC. In [12], Lu et al. proposes an efficient and privacy-preserving aggregation scheme by homomorphic multidimensional data encryption schemes (EPPA), which can realize the multidimensional data aggregation. On this basis Chen et al. [13] try to use third parties to achieve fault tolerance of data aggregation, but the obvious disadvantages is that third party security is difficult to guarantee. Shi et al. [14] proposes the DG-APED scheme, which can resolve the problems caused by malfunctioning SMs. it will aggregate the data by grouping, and drop the group which contains the damaged SM. However, because of error rate is not ideal and extra computational cost in searching the damaged member also needs to spend. Works [15] are committed to achieve the efficient data aggregation, but the cost of realizing fault tolerance is still too high, and there is still room for improvement. Works [16] are proposed to realize the differential privacy in aggregation schemes. Wang *et al.* [17] proposes an electric vehicle in the smart grid traceability of privacy protection and precision incentive scheme, using a restrictive partially blind signature technique and pseudonym in V2G (vehicle-to-grid) networks to achieve traceability of malicious users. Several other papers (e.g., [18–24]) have studied related security and network issues.

3 Preliminaries

3.1 Bilinear Maps

Let G_0 and G_1 be two multiplicative cyclic groups of prime order p and g be the generator of G_0. The bilinear map e is, $e : G_0 \times G_0 \rightarrow G_1$, for all $a, b \in \mathbb{Z}_p$:

Bilinearity: $\forall u, v \in G_1, e(u^a, v^b) = e(u, v)^{ab}$
Non-degeneracy: $e(g, g) \neq 1$
Symmetric: $e(g^a, g^b) = e(g, g)^{ab} = e(g^b, g^a)$

3.2 Elliptic Curve Cryptography (ECC)

Elliptic curve encryption (ECC) algorithm [25, 26], proposed by Koblitz and Miller in 1985, Define an elliptic curve E and a field $GF(q)$.Consider x, y Abel with a form of rational number $E(q)$, Elliptic curve equation E defined as

$$y^2 + a_1xy + a_2y = x^3 + a_3x^2 + a_4x + a_6$$

The point $E(K)$ on the elliptic curve that satisfies the equation plus the set of infinity points is expressed:

$$E(K) = \{(x, y) \in k^2 | y_2 + a_1xy + a_2y = x^3 + a_3x^2 + a_4x + a_6\} \cup \{0\}$$

3.3 Complexity Assumptions

Definition 1. ECC is based on the problem of finding elliptic curve discrete pairs (ECDLP) is difficult.

That is, for a base point on the elliptic curve, it is easy to give an integer test, but it is very difficult to derive the integer from the point and point, that is, there is no algorithm to solve the polynomial time, which is elliptic curve discrete Logarithmic problem, to provide security for ECC-based encryption algorithms.

Definition 2 Bilinear Diffie-Hellman (BDH) Problem. The Bilinear Diffie-Hellman (BDH) problem in G is as follows: Given $(P, aP, bP, cP)(a, b, c \in Z_q^*)$, calculate, $\omega = e(P, P)^{abc} \in G_2$, e is a bilinear mapping, P is the generator of G_1, G_1, G_2 is the order of prime numbers q of the two groups, Set the algorithm A to solve the BDH problem, The advantage of an adversary τ is defined as $\Pr|A(P, aP, bP, cP) = e(P, P)^{abc}| \geq \tau$.

There is no valid algorithm to solve the BDH problem, so it can be assumed that the BDH problem is a difficult problem.

3.4 Based on BDH IBE (Identity-Based Cryptosystem)

IBE [25] algorithm consists of four steps:

Step 1 System initialization:

Let $k \in Z^+$ be a safety parameter, run the BDH parameter generation algorithm g, Output prime number q, group orders of q, G_1, G_2, a bilinear mapping $e : G_1 \times G_1 \rightarrow G_2$. Select a random generator $P \in G_1$, random selection $s \in Z_q^*$, calculating $P_{pub} = sP$. Select a hash function $H_1 : \{0,1\}^* \rightarrow G_1^*$, for n, Select another hash function $H_2 : G_2 \rightarrow \{0,1\}^n$, the message space is $\mathrm{M} = \{0,1\}^n$ ciphertext space is $\mathrm{C} = G_1^* \times \{0,1\}^n$, System parameters are public: params $= <q, G_1, G_2, e, n, P, P_{pub}, H_1, H_2>$, s is the master key, is confidential.

Step 2 Encryption:

The identity ID of the recipient is encrypted as a public key, $M \in \mathrm{M}$, calculate $Q_{ID} = H_1(ID) \in G_1^*$, choosing random number $r \in Z_q^*$, generating ciphertext:

$$C = <rP, M \oplus H_2(\mathrm{g}_{ID}^r)> , \ \mathrm{g}_{ID} = e(Q_{ID}, P_{\mathrm{pub}}) \in G_2^* \tag{1}$$

Step 3 Key generation:

For a given bit string $\mathrm{ID} = \{0,1\}^*$, calculate $Q_{ID} = H_1(ID) \in G_1^*$, then calculate secret key $d_{ID} = sQ_{ID}$, master key is s.

Step 4 Decryption:

Set ciphertext is $C = <U, V> \in$ C, then use d_{ID} calculate

$$V \oplus H_2(e(d_{ID}, U)) = M,\ \text{Get the plaintext}\ M \tag{2}$$

3.5 BGN (Boneh-Goh-Nissim) Cryptosystem

Given the security parameter g, composite bilinear parameters $(p, q, \mathbb{G}, \mathbb{G}_1, e)$ are generated by $\varsigma(\kappa)$, where $n = pq$ and p, q are two k-bit prime numbers $g \in \mathbb{G}$ is a generator of order n. Set $h = g^q$, then h is a random generator of the subgroup of $\mathbb{G}$ order p. The public key is $PK = (N, \mathbb{G}, \mathbb{G}_1, e, g, h)$, and the corresponding private key is $SK = p$.

Step 2 Encryption:

We assume the message space consists of integers in the set $m = \{0, 1, \ldots\ldots W\}$ with $W \ll q$. To encrypt a message m, we choose a random number $r \in \mathbb{Z}_N$ and compute the ciphertext:

$$c = E(m, r) = g^m \cdot h^r \in \mathbb{G} \tag{3}$$

Step 3 Decryption:

Given the ciphertext $c = E(m, r) = g^m h^r \in \mathbb{G}$, the corresponding message can be recovered by the private key $SK = p$,

$$c^p = (g^m \cdot h^r)^p = (g^p)^m. \tag{4}$$

Let $g^* = g^p$, To recover m, it suffices to compute the discrete log of c^p base g^*. Since $0 \le m \le T$, the expected time is around $O(\sqrt{T})$ when using the Pollard's lambda method [26].

4 Models and Goals

4.1 System Model

In this section, we propose an IBE-based Device Traceable Privacy-Preserving Aggregation Scheme in the Energy Internet. The system model as Fig. 3 shows, mainly composed of CC, TCA (Trusted Third Party), edge nodes like GWs, and a varied of Users in the RA.

User: 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 SMs. As the real-time data is related to user privacy, the data must be encrypted by the SM before sending to the CC.

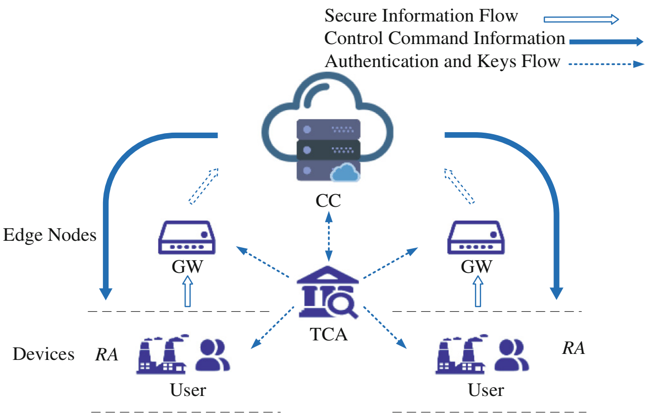

Fig. 3. System model

GW (gateway): is responsible for collecting the encrypted data sent by SMs in RA, calculating the aggregation of real-time data by running the homomorphic algorithm and uploading the sum to the control center. Responsible for data aggregation integrity verification and encryption of aggregated data by IBE. In order to improve the efficiency of the system, the user selects the nearest available GW in RA.

TCA (Trusted Third Party): responsible for the SM, GW and CC initialization to generate keys and system parameters, generating dynamic IDs for the GW in RA and CC, and CC authentication.

CC (Control Center): Can acquire the summary of real–time data from GW with these data, CC can get the trend of power consumption and create the power generation plan or dynamic price immediately. In order to improve efficiency of the Energy Internet, different regions set up different CCs.

4.2 IBE-DTPPA Scheme Procedure

The procedure of IBE-DTPPA Scheme has the following four steps:

Step 1 User data request and encryption:

(1) When the CC Sends a data request in RA, or Users' data is collected periodically (15 min), the TCA is initialized to generate the encryption parameters for SM and GW. (2) SM encrypts current data by BGN, and transfers to the nearest available GW in RA.

Step 2 Data aggregation and aggregation integrity verification:

(1) When GW receives encrypted data from users in RA, then GW aggregates data and user random numbers. (2) The aggregation integrity of the user data in RA is verified by the random number aggregation.

Step 3 Secondary Encryption:

If the data is successfully aggregated, the aggregation is re-encrypted by IBE encryption based on the dynamic ID of CC in GW, choosing CC's ID as public key, calculating ciphertext based on GW's ID. To realize the CC real-time authentication and traceability of malicious GW. The ciphertext is forwarded to CC.

Step 4 Decryption and GW traceability:

If the authentication of CC is successful, CC gets decrypt permission, getting the aggregation data in RA, if CC doubts the authenticity of the aggregation data, and wants to trace the source, then the GW which responsible for the data aggregation will be traced. If find the GW is destroyed or controlled by the adversary, the malicious GW will be isolated and replaced by other available GWs in RA in time.

4.3 Adversary Model

We assume that SM installed on the user side is a trusted device. The communication channel is not secure and adversary may eavesdrop on the channel. The GW is vulnerable to be controlled or destroyed by the adversary. CC is not fully credible, will not take the initiative to disclose user information, but the adversary will use CC's identity fraudulently to steal user's data, which will bring the privacy and security threats to users.

4.4 Design Goals

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

(1) Privacy-preserving: users' data in RA is inaccessible to any other users. The outside adversary, GW or CC should not acquire the real-time data of users even if they try to conspire with each other.

(2) Target GW traceable: The aggregation data encrypted by IBE Cryptosystem, calculating the ciphertext by GW's dynamic ID. When CC wants to trace the source of the aggregation data, tracing the target GW efficiently.

(3) CC authentication and aggregation integrity verification: preventing the adversary from fraudulently using CC's identity, using CC's dynamic ID as IBE public key to realize real-time authentication of CC. In order to ensure the accuracy of data collection of RA, random number aggregation is used to verify the integrity of user data aggregation in RA by BGN Cryptosystem.

5 IBE-DTPPA Scheme

5.1 System Initialization

(1) Device dynamic identity generation

In order to achieve CC real-time authentication, preventing the adversary tracing the data owner based on the fixed ID of GW, in our scheme, updating the dynamic ID of

GW and CC ID_{G_i} ID_{C_i} in a short period, updated ID_{G_i}, ID_{C_i} by TCA. The update period is bounded by the times of calculations of RA data collection. For example, the number of GW calculations $\mathrm{Times}_{\mathrm{GW}}(\mathrm{Times}_{\mathrm{GW}} \leq 50)$ and CC $\mathrm{Times}_{\mathrm{CC}}(\mathrm{Times}_{\mathrm{CC}} \leq 100)$, and the TCA updates the ID for the device when the threshold is reached.

(2) System parameter generation

Step 1. TCA runs $\mathrm{Gen}_1(k)$, generating the parameters used for DBH-based IBE Cryptosystem: Given the security parameter $k \in Z^+$, calculating a prime number q_{IBE}, groups G_1, G_2, $G_1 \times G_1 \rightarrow G_2$ of order q_{IBE}. Select the random generate $P \in G_1$, selecting random number $s \in Z_q^*$, calculating $\mathrm{PK}_{IBE} = sP$, selecting Hash Function $H_1 : \{0,1\}^* \rightarrow G_1^*$ $H_2 : G_2 \rightarrow \{0,1\}^n$. Public parameter is $\mathrm{par}_{IBE} = <q_{IBE}, G_1, G_2, e, n, P, P_{pub}, H_1, H_2>$.

Step 2. Run $\mathrm{Gen}_2(k)$, generating the required parameters for BGN Cryptosystem, (p, q, G), p, q are two prime numbers, selecting random numbers $g \in G$, $x \in G$, calculating $h = x^q$, $\mathrm{PK}_{BGN} = (N, G, g, h)$, $\mathrm{SK}_{BGN} = p$.

Step 3. In order to achieve aggregation integrity verification, when RA users $U = \{U_1, U_2, \ldots, U_n\}$ data encrypted by BGN (assigned to the same GW), TCA will generates a system random number r_s for the RA users, calculating the random number of each user based on the system random number:

$$(r_1 + r_2 + \ldots + r_n) = r_s \bmod p \tag{5}$$

Send the different random number r_i for each user to the user in RA for encryption. Parameter generation process as Fig. 4 shows.

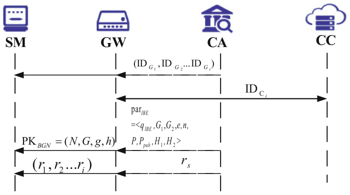
SM
GW
CA
CC
$(\mathrm{ID}_{G_1}, \mathrm{ID}_{G_2} \ldots \mathrm{ID}_{G_i})$
ID_{C_i}
$\mathrm{par}_{IBE} = <q_{IBE}, G_1, G_2, e, n, P, P_{pub}, H_1, H_2>$
$\mathrm{PK}_{BGN} = (N, G, g, h)$
$(r_1, r_2 \ldots r_i)$
r_s

Fig. 4. System initialization

5.2 User Data Encryption

(1) SM (Smart Meter)

User U_i in RA_j collects user's data d_i periodically (15 min) by SM, encrypting d_i by BGN Cryptosystem, and based on the user's random number r_i, according to the formula (3), calculating $C_{BGNi} = g^{d_i} h^{r_i}$.

After the encryption process, in order to prevent the attacker from listening at the target GW, and increase the efficiency of the system. TCA choose the nearest available GW in RA_j for the users to aggregate data randomly. And then SM forwards C_{BGNi} to the chosen GW.

(2) GW (Gateway)

Upon receiving all the encrypted data from SMs, GW_a aggregates all the data by:

$$\begin{aligned} \mathrm{C}_{Ua_j} &= \prod\nolimits_{i=1}^{n} C_{BGNi} \\ &= g^{d_1}h^{r_1} \cdot g^{d_2}h^{r_2} \cdots g^{d_n}h^{r_n} \\ &= \left(g^{\sum_{i=1}^{n} d_i} h^{\sum_{i=1}^{n} r_i}\right) \\ &= g^{\sum_{i=1}^{n} d_i} h^{r_s{}'} \end{aligned} \tag{6}$$

After aggregating data in GW_a, and then aggregates user random number r_i, compared with system random number r_s, $\sum_{i=1}^{n} r^i \overset{?}{=} r_s$ if it does hold, proved aggregation is successful, otherwise, directly abandon the data, sending a data request to CC again, which will increase system strategy reliability and reduce overhead of error aggregation data for the system.

5.3 Secondary Encryption

In order to achieve the traceability of the GW device and increase the security of the CC, we encrypts aggregation data by IBE Cryptosystem in IBE-DTPPA scheme. We use CC's dynamic ID as the public key, calculating ciphertext based on GW's dynamic ID as random number. Secondary aggregation data encryption, increasing the data security, CC real-time identity authentication to ensure that CC is not be used fraudulently and trace target GW efficiently. The process as:

When the GW requests the secondary encryption of the aggregated data, TCA generates the public parameters par_{IBE} for the IBE encryption, sending the GW dynamic ID, ID_{g_a} CC dynamic ID, $\mathrm{ID}_{\mathrm{C}_i}$ and the public parameters par_{IBE} to the target GW. TCA calculates the public key based on $\mathrm{ID}_{\mathrm{C}_i}$, calculating ciphertext C' by IBE Encryption. The current time stamp. TS_t is set, in order to prevent replay attack. And in order to ensure the integrity of the message, we select the hash function $H_2 : G_2 \rightarrow \{0,1\}^n$, generating a message digest δ, and GW sends it with TS_t, C' to the CC.

Step 1. Calculate $Q_{\mathrm{ID}_{\mathrm{C}_i}} = H_1(\mathrm{ID}_{\mathrm{C}_i}) \in G_1^*$;
Step 2. The GW dynamic identity information $\mathrm{ID}_{g_i} \in \mathrm{Z}_q^*$ is taken as a random number.
Step 3. According to the formula (1), calculating the ciphertext:

$$C' = <\mathrm{ID}_{ga}P,\ g^{\sum_{i=1}^{n} d_i} h^{r'} \oplus H_2(g_{\mathrm{ID}_{ga}}^{\mathrm{ID}_{ga}}) > \tag{7}$$

Step 4. Calculate $\delta = H_2(C')$, sending $\{\delta = H_2(C'), C', TS_t\}$ to the target CC.

5.4 Data Decryption and Devices Traceability

(1) CC Authentication

After receiving $\{\delta = H_2(C'), C', TS_t\}$, and CC verifies whether $H_2(C') \overset{?}{=} \delta$, If it does hold, the message has not been tampered, otherwise the data request is sent again to the user in RA_j. Then verifies whether the aggregated data is available by checking TS_t then CC sends a decrypted data request to the TCA, following as:

Step 1 TCA authenticates the CC's current identity and generates the key: $Q_{\mathrm{ID}_{\mathrm{C}_i}} = H_1(\mathrm{ID}_{\mathrm{C}_i}) \in G_1^*$, $SK_{IBE} = d_{ID_{\mathrm{C}_i}} = sQ_{ID_{\mathrm{C}_i}}$, when the verification is successful, sending $sQ_{ID_{\mathrm{C}_i}}$ to CC.

Step 2 Decrypt $C' = <\mathrm{ID}_{g_i}P,\ g^{\sum_{i=1}^{n} d_i} h^{r'} \oplus H_2(g_{\mathrm{ID}_{g_i}}^{\mathrm{ID}_{g_i}}) >$ by IBE Cryptosystem. According to the formula (2), as:

$$\begin{aligned} \mathrm{C}_{Ua_j} &= g^{\sum_{i=1}^{n} d_i} h^{r'} \oplus H_2\left(g_{\mathrm{ID}_{g_i}}^{\mathrm{ID}_{g_i}}\right) \\ &\quad \oplus H_2(e(sQ_{\mathrm{ID}}, \mathrm{ID}_{g_i}P)) \\ &= g^{\sum_{i=1}^{n} d_i} h^{r'} \end{aligned} \tag{8}$$

Generate $ID_{G_i}P$ and C_{Ua_j} by IBE Cryptosystem encryption.

Step 3 Decrypt C_{Ua_j} according to secret key $\mathrm{SK}_{BGN} = p$ by BGN Cryptosystem, as:

$$\begin{aligned} C^{SK_{BGN}} &= \left(g^{\sum_{i=1}^{n} d_i} h^{r'}\right)^p = g^{\sum_{i=1}^{n} d_i p} x^{nr'} \\ &= g^{\sum_{i=1}^{n} d_i p} e^{r'} \\ &= (g^p)^{\sum_{i=1}^{n} d_i} \end{aligned} \tag{9}$$

To recover $\sum_{i=1}^{n} d_i$, which suffices to compute the discrete log of c^p base g^*. Since $0 \leq d \leq T$, CC can get the sum of users' data $\sum_{i=1}^{n} d_i$ in expected time $O(\sqrt{nT})$ using the Pollard's lambda method [26].

(2) Target GW Device Traceability

If CC doubts the authenticity of the aggregation data, and wants to trace the source, then the GW which responsible for the data aggregation will be traced, sending $ID_{G_i}P$ in the ciphertext C' by IBE Cryptosystem to TCA, CC send $ID_{G_i}P$ to TCA, calculating the target GW's dynamic ID, ID_{G_i} based on public parameter P, TCA trace the target GW by ID_{G_i}. If TCA finds the GW is destroyed or controlled by the adversary, the malicious GW will be isolated and replaced by other available GWs in RA in time.

6 Security Analysis

In this section, we analyze the security properties of the proposed IBE-DTPPA scheme. In particular, following the security requirements discussed earlier, our analysis will focus on how IBE-DTPPA scheme can achieve the privacy of individual user data in RA, the authentication of CC and the verification of data aggregation, and the suspicious GW traced efficiently.

(1) **The individual user's data is privacy-preserving in the proposed IBE-DTPPA scheme**

In the propose IBE-DTPPA scheme, user U_i's data in RA, $(d_1, d_2, \ldots, d_i)$ sensed by SMs are encrypted as $C_{BGNi} = g^{d_i}h^{r_i}$ by BGN cryptosystem. Since BGN cryptosystem is provably secure against chosen plaintext attack based on the subgroup decision assumption, the data $(d_1, d_2, \ldots, d_i)$ in C_{BGNi} is also semantic secure and privacy-preserving. Therefore, even though the adversary $\mathcal{A}$ eavesdrops C_{BGNi}, he still cannot identify the corresponding contents. After collecting all reports $(C_{BGN1}, C_{BGN2}, \ldots, C_{BGNi})$ from the RA, the GW will not recover each user's data, instead, it just computes $\mathrm{C}_{Ua_j} = \prod_{i=1}^{n} C_{BGNi}$ to perform report aggregation. Therefore, even if the adversary $\mathcal{A}$ intrudes in the GW's database, he cannot get the individual report $(d_1, d_2, \ldots, d_i)$ either. Finally, after receiving $\mathrm{C}_{Ua_j} = \prod_{i=1}^{n} C_{BGNi}$ from GW, the CC recovers C_{Ua_j} as $D_j = \sum_{i=1}^{n} d_i$. However, since D_j is an aggregated result, even if the adversary $\mathcal{A}$ steals the data, he still cannot get the individual user U_i's data $(d_1, d_2, \ldots, d_i)$ Therefore, from the above three aspects, the individual user's report is privacy-preserving in the proposed IBE-DTPPA scheme.

(2) **The authentication of CC and the security of aggregation data can be guaranteed in IBE-DTPPA scheme**

(1) In the propose IBE-DTPPA scheme, each individual user's data is encrypted by BGN cryptosystem and the aggregated report are encrypted by IBE Cryptosystem, choosing CC's dynamic ID, ID_{C_i} as public key and encrypt the aggregation data generate C'_{Ua_j} by IBE Cryptosystem, the CC's identity authentication can be realized. Since ID_{C_i} updates aperiodically by TCA, the adversary $\mathcal{A}$ cannot get CC's current ID, preventing the adversary $\mathcal{A}$ from using CC identity fraudulently.

(2) GW sends message $M = \{\delta, C', TS_t\}$, $\delta = H_2(C')$ to the CC, δ is the digest of hash function $H_2 : G_2 \rightarrow \{0,1\}^n$ in random oracle model, C' and is a valid ciphertext

of IBE Cryptosystem. Since in IBE-DTPPA scheme, IBE Cryptosystem is based ECC (Elliptic curve cryptography) algorithm, which is under the assumption that ECDLP problem is hard, IBE is semantic secure against the chosen plaintext attack under the assumption that BDH problem is hard. Therefore, $M=\{\delta, C', TS_t\}$ is semantic secure against chosen-plaintext attack based on IBE Cryptosystem and random oracle model. As a result, the authentication of CC's identity can be realized, adversary $\mathcal{A}$ in the Energy Internet cannot fraudulently use CC identity to steal the user's data, the security of ciphertext encrypted by IBE Cryptosystem can be guaranteed in IBE-DTPPA scheme.

(3) **Target GW in the Energy Internet can be traced efficiently in IBE-DTPPA scheme**

After the CC's authentication is successful, CC recovers the aggregated data D_j in RA_i from C'_{Ua_j}. If CC doubts the authenticity of the aggregation data, and wants to trace the source, CC will send $ID_{\mathrm{G}_i}P$ in C'_{Ua_j} to TCA, to trace the target GW which responsible for the data aggregation, TCA calculates GW's dynamic ID, ID_{G_i} based on public parameter P, GW will be traced efficiently. If find the GW is destroyed or controlled by the adversary, the malicious GW will be isolated and replaced by other available GWs in RA in time. As a result, the adversary $\mathcal{A}$ in the Energy Internet cannot control any GW to transmit error aggregated data, thus improving the CA's system strategic-making reliability.

7 Conclusion

This paper, we proposed IBE-DTPPA scheme, IBE-based Device Traceable Privacy-Preserving Aggregation Scheme. It can realize: (1) the traceability of target GW device; (2) CC real-time authentication to preventing the adversary from using CC identity fraudulently; (3) increase data aggregation integrity verification, to ensure the accuracy of system decision-making while creating the power generation plan or dynamic price immediately. We also provide security analysis to demonstrate its security.. For future work, we will work on resolving the fault-tolerant in GW, deepen the IBE-DTPPA scheme.

Acknowledgement. This work is partially supported by Natural Science Foundation of China under grant 61402171, the Fundamental Research Funds for the Central Universities under grant 2016MS29.

References

1. Wang, K., Yu, J., Yu, Y., et al.: A survey on Energy Internet: architecture, approach, and emerging technologies. IEEE Syst. J. **PP**(99), 1–14 (2017)
2. Guan, Z., Li, J., Zhu, L., Zhang, Z., Du, X., Guizani, M.: Towards delay-tolerant flexible data access control for smart grid with renewable energy resources. IEEE Trans. Ind. Inform. **13**(6), 3216–3225 (2017)

3. Wang, K., Ouyang, Z., Krishnan, R., et al.: A game theory-based energy management system using price elasticity for smart grids. IEEE Trans. Ind. Inform. **11**(6), 1607–1616 (2015)
4. Guan, Z., Li, J., Wu, L., Zhang, Y., Wu, J., Du, X.: Achieving efficient and secure data acquisition for cloud-supported Internet of Things in smart grid. IEEE Internet Things J. **4** (6), 1934–1944 (2017)
5. Davies, S.: Internet of energy [smart grid security]. Eng. Technol. **5**(1), 1–2 (2010)
6. Efthymiou, C., Kalogridis, G.: Smart grid privacy via anonymization of smart metering data In: First IEEE International Conference on Smart Grid Communications. IEEE, pp. 238–243 (2010)
7. Tan, X., Zheng, J., Zou, C., et al.: Pseudonym-based privacy-preserving scheme for data collection in smart grid. Int. J. Ad Hoc Ubiquitous Comput. **22**(2), 120 (2016)
8. Guan, Z., Si, G., Wu, J., et al.: Utility-privacy tradeoff based on random data obfuscation in internet of energy. IEEE Access **5**, 3250–3262 (2017)
9. Beussink, A., Akkaya, K., Senturk, I.F., Mahmoud, M.M.E.A.: Preserving consumer privacy on IEEE 802.11s-based smart grid AMI networks using data obfuscation. In: Proceedings of the IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp. 658–663, April 2014
10. Shi, E., Chan, T.-H.H., Rieffel, E.G., Chow, R., Song, D.: Privacy-preserving aggregation of time-series data. In: NDSS, vol. 2, p. 4 (2011)
11. Kim, Y.S., Heo, J.: Device authentication protocol for smart grid systems using homomorphic hash. J. Commun. Netw. **14**(6), 606–613 (2012)
12. Lu, R., Liang, X., Li, X., et al.: EPPA: an efficient and privacy-preserving aggregation scheme for secure smart grid communications. IEEE Trans. Parallel Distrib. Syst. **23**(9), 1621–1631 (2012)
13. Chen, L., Lu, R., Cao, Z.: PDAFT: a privacy-preserving data aggregation scheme with fault tolerance for smart grid communications. PeerPeer Netw. Appl. **8**(6), 1122–1132 (2015)
14. Shi, Z., Sun, R., Lu, R., Chen, L., Chen, J., Shen, X.S.: Diverse grouping-based aggregation protocol with error detection for smart grid communications. IEEE Trans. Smart Grid **6**(6), 2856–2868 (2015)
15. Han, S., Zhao, S., Li, Q., Ju, C.-H., Zhou, W.: PPM-HDA: privacy-preserving and multifunctional health data aggregation with fault tolerance. IEEE Trans. Inf. Forensics Secur. **11**(9), 1940–1955 (2015)
16. Hua, J., Tang, A., Fang, Y., Shen, Z., Zhong, S.: Privacy-preserving utility verification of the data published by non-interactive differentially private mechanisms. IEEE Trans. Inf. Forensics Secur. **11**(10), 2298–2311 (2016)
17. Wang, H., Qin, B., Wu, Q., et al.: TPP: traceable privacy-preserving communication and precise reward for vehicle-to-grid networks in smart grids. IEEE Trans. Inf. Forensics Secur. **10**(11), 2340–2351 (2015)
18. Xiao, Y., Du, X., Zhang, J., Guizani, S.: Internet Protocol Television (IPTV): the killer application for the next generation internet. IEEE Commun. Mag. **45**(11), 126–134 (2007)
19. Du, X., Chen, H.H.: Security in wireless sensor networks. IEEE Wirel. Commun. Mag. **15** (4), 60–66 (2008)
20. Xiao, Y., Rayi, V., Sun, B., Du, X., Hu, F., Galloway, M.: A survey of key management schemes in wireless sensor networks. J. Comput. Commun. **30**(11–12), 2314–2341 (2007)
21. Du, X., Xiao, Y., Guizani, M., Chen, H.H.: An effective key management scheme for heterogeneous sensor networks. Ad Hoc Netw. **5**(1), 24–34 (2007)
22. Du, X., Guizani, M., Xiao, Y., Chen, H.H.: A routing-driven elliptic curve cryptography based key management scheme for heterogeneous sensor networks. IEEE Trans. Wirel. Commun. **8**(3), 1223–1229 (2009)

23. Du, X., Guizani, M., Xiao, Y., Chen, H.H.: Secure and efficient time synchronization in heterogeneous sensor networks. IEEE Trans. Veh. Technol. **57**(4), 2387–2394 (2008)
24. Du, X., Xiao, Y., Chen, H.H., Wu, Q.: Secure cell relay routing protocol for sensor networks. Wirel. Commun. Mob. Comput. **6**(3), 375–391 (2006)
25. Shamir, A.: Identity-based cryptosystems and signature schemes. In: Blakley, G.R., Chaum, D. (eds.) CRYPTO 1984. LNCS, vol. 196, pp. 47–53. Springer, Heidelberg (1985). https://doi.org/10.1007/3-540-39568-7_5
26. Gallant, R., Lambert, R., Vanstone, S.: Improving the parallelized pollard lambda search on anomalous binary curves. Math. Comput. Am. Math. Soc. **69**(232), 1699–1705 (2000)