Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2015, Article ID 915795, 12 pages
http://dx.doi.org/10.1155/2015/915795

Research Article

PAS: An Efficient Privacy-Preserving Multidimensional Aggregation Scheme for Smart Grid

Hui Zhu, Fen Liu, Rong Yan, and Hui Li

State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China

Correspondence should be addressed to Hui Zhu; zhuhui@xidian.edu.cn

Received 8 June 2015; Revised 24 August 2015; Accepted 2 September 2015

Academic Editor: Qinghua Li

As a convergence of traditional power system engineering and information technology, smart grid, which can provide convenient environment of operation and management for the power provider, has attracted considerable interest recently. However, the flourish of smart grid is still facing many challenges in data security and privacy preservation. In this paper, we propose an efficient privacy-preserving multidimensional aggregation scheme for smart grid, called PAS. Without disclosing the privacy-sensitive information (e.g., identity and power consumption) of users, the operation center can obtain the number of users and power consumption at each step in different dimensions. Based on an improved Paillier cryptosystem, the operation center can acquire more valid information to regulate the generated energy, and an efficient anonymous authentication scheme is employed to protect the privacy of user's identity from the regional center. Detailed security analysis shows the security and privacy-preserving ability of PAS. In addition, performance evaluations via extensive simulations demonstrate that PAS is implemented with great efficiency for smart grid in terms of computation and communication overhead.

1. Introduction

The 22nd world energy congress under the theme of "Securing Tomorrow's Energy Today" was held on October 13, 2013, in Daegu, South Korea [1], which offered a unique opportunity for participants to get a better understanding of energy issues and solutions from a global perspective. Simultaneously, the optimization of energy structure and the improvement of energy utilization were taken into consideration inevitably. Especially in the electric area, the lack of reasonable electrical structure to meet growing demands has stretched the power grid to its limits. Furthermore, in the past 100 years, there has been no dramatic change in the basic structure of electrical power grid. Experience has shown that the hierarchical, centrally controlled power grid cannot meet the growing demands of the 21st century [2]. Therefore, researching for a new power grid, which can support the new power system and enhance the efficiency of power consumption, is currently an emergency for most countries.

Recently the appearance of smart grid has attracted plenty of countries' interest and has been regarded as the next generation of power grid [3–8]. Compared with traditional grid, smart grid supports centralized two-way transmission and efficiency-driven response. Besides, smart grid relies on cyber-physical systems or the Internet of Things to provide intelligent scheduling for power transmission and distribution [9]. For instance, smart meter equipped with network interfaces (e.g., wireless sensors) reports power consumption to the operation center via the advanced meter controlled by the regional center, as shown in Figure 1. The operation center spans different geographic regions and can transmit the result of data analysis to the power supplier in a distributed way, whose responsibility is to adjust power supply dynamically to meet demands and detect and respond to the weaknesses or failures of power system in real time.

Smart meter (SM) which is two-way communication device and deployed at users premise is indispensable component in the smart grid, since it can record real-time power consumption periodically. With SM, smart grid is capable of collecting real-time information about grid operations and status at the operation center. SM will send the encrypted power consumption to the regional center after being authorized. Through the regional center, smart grid

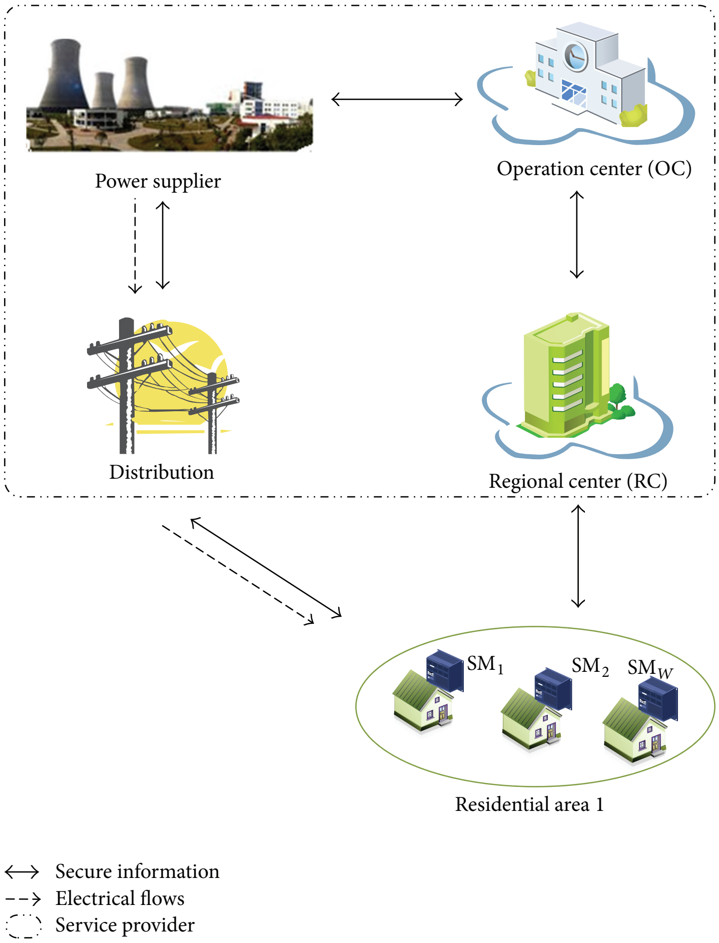


FIGURE 1: The conceptual architecture of smart grid.

can transmit the user's detailed electricity information to the operation center, which may then change electricity price accordingly or even adjust user's usage. However, if there is an adversary in the regional area, he/she may infer user's physical activities based on the power consumption of a family. For instance, unusual low daily power consumption of a household indicates that the home owners are probably away from home. Therefore, such privacy-sensitive information must be protected from unauthorized access [10, 11].

However, the majority of existing schemes [12–15] can offer total power consumption to the power supplier; this user-centric information is far from enough for the service provider (e.g., the electric company) to allocate power resource. For example, many service providers adopt the multistep electricity pricing policy [16] where the price of electricity is based on the user's power consumption. In order to utilize and dispatch different kinds of step electricity for service providers, more information related to users (e.g., the number of users and power consumption at each step, etc.) is required. With regard to the power supplier, not only can it classify users depending on the power consumption, but also it works out a practical and reasonable business strategy.

In this paper, from the perspective of power supplier, under the promise of protecting the confidentiality of user's identity and power consumption, we propose an efficient privacy-preserving multidimensional aggregation scheme for smart grid, called PAS. The characteristics of proposed PAS are offering more valid user information to the power supplier and achieving user's identity privacy and data confidentiality. The main contributions of this paper are threefold.

(i) First, PAS provides more valid user information without disclosing user's data and identity privacy. Upon the employment of an improved Paillier cryptosystem [17] technique, the operation center can acquire both the number of users and power consumption at each step by decrypting the preliminary aggregations from the regional center. Simultaneously, compared with the existing aggregation schemes [18–20], PAS provides more effective information to the power supplier to make more planed electricity strategy.

(ii) Second, PAS can guarantee fine-grained privacy preserving. Although the aggregation scheme [12] can guarantee the privacy of user's information, the authenticator still knows the identity which corresponds to the signature. However, in our scheme, the improved group signature [21] scheme makes user's

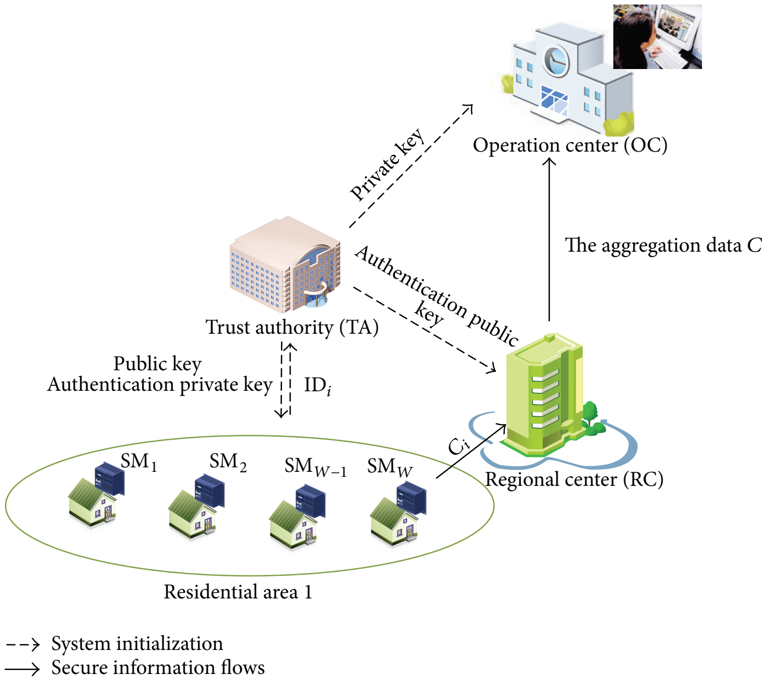


Figure 2: System model under consideration.

identity unrecognized by the regional center. Meanwhile, data privacy preservation requirements can be achieved through the improved Paillier cryptosystem.

(iii) Third, PAS achieves the traceable anonymity of smart grid. We employ a specific group signature scheme between the smart grid and the regional center, which can obtain the authentication and traceability of message. The regional center checks out the message and signature without knowing the smart grid's identity. But if the smart grid's signature is proved to be invalid, the identity of malicious smart grid could be ferreted out with the help of TA.

The remainder of this paper is organized as follows. In Section 2, we formalize the system model and security requirements and identify our design goal. Then, we recall the discrete logarithm and Paillier cryptosystem as the preliminaries in Section 3. Then, we present our PAS scheme in Section 4, followed by the security analysis and performance evaluation in Sections 5 and 6, respectively. We also discuss some related works in Section 7. Finally, we draw our conclusions and identify some future work in Section 8.

2. System Model, Security Requirements, and Design Goal

In this section, we formalize the system model and security requirements and identify our design goal.

2.1. System Model. In our system model, we mainly focus on how to report user's privacy-preserving power consumption information to the operation center. Specifically, the system consists of four parts: trusted authority (TA), residential area (RA), operation center (OC), and regional center (RC), as shown in Figure 2.

TA: the trusted authority is fully trustable in the whole system, whose duty is to bootstrap the system initialization and distribute secret keys to the SM, RC, and OC (some parameters must be kept private, while others are public). In addition to system initialization, TA is offline in other phases. Only in the case that errors exist in the process of authentication, TA can respond to it in time; for example, when errors arise in the authentication, TA can recover the malicious identity of SM based on the group signature scheme. Therefore, TA is an indispensable part in this system.

RA: we consider a typical residential area, which mainly consists of a connection with RC and a great number of users that are equipped with smart meters $\{\mathrm{SM}_1, \mathrm{SM}_2, \ldots, \mathrm{SM}_n\}$. SM_i records the real-time data about power consumption, and the data will be reported to RC in a certain period.

RC: RC is a powerful workshop, which mainly performs three functions: authentication, aggregation, and relaying. Some authentication operations must be performed to guarantee the privacy of user's identity before SM transmits user's encrypted data to RC. The aggregation component is responsible for aggregating user's multidimensional power consumption into a compressed one. The relaying component transmits aggregation data to OC.

OC: the responsibility of OC is to analyze the aggregated regional data. It can supply power supplier with valid user information which includes the number of users and power consumption at each step, so as to carry out the dynamic step tariff and locate the trouble spot.

2.2. Security Requirements. In our security model, we consider RA (e.g., a typical residential area) mainly consists of a connection with RC (e.g., the property management) and a great number of users that are equipped with SM. OC (e.g., the electric company) spans different RC and analyzes the aggregation regional data. Specifically, we assume that RC and OC are not in collusion with each other and both of them are honest but curious. That is, RC and OC aggregate the encrypted data from SM exactly, but it is curious to SM's real-time data and identity. In addition, the confidentiality and the integrity of data need be guaranteed. Therefore, in order to prevent the user's data in RC and OC, PAS must satisfy the following three security requirements.

(i) *Confidentiality.* The scheme should protect user's data from being analyzed by RC, and guarantee the user's identity privacy. Though RC has all the data from SM_i, it cannot identify the contents of reports, neither can it distinguish any user's data. In such a way, our system model can meet the privacy-preserving requirements of user's data.

(ii) *Authentication.* The scheme should guarantee that an encrypted report at RC is really sent by a legitimate residential user and has not been transformed during the transmission to RC; that is, if the adversary forges or modifies a report, the malicious operations should be detected at RC. Thus, only the correct user information can be received by RC.

(iii) *Anonymity.* The identity of user should be unknown in the scheme. On one hand, when RC receives the messages from SM_i, the signature could be checked without disclosing the identity of SM_i. On the other hand, OC can acquire the number of users at each step without knowing the users' identities. Under this circumstance, the identity of user is kept secret from RC and OC.

2.3. Design Goal. According to the aforementioned system model and security requirements, our design goal is to develop an efficient privacy-preserving multidimensional aggregation scheme for smart grid. Specifically, the following three objectives should be achieved.

(i) The scheme should provide more effective information to the power supplier and satisfy a classified management. The power supplier should obtain the total power usage information that is the number of users and power consumption at each step. With these valid data, not only can it develop a reasonable step tariff strategy, but also the power consumption in one RA can realize load balancing. Through the implementation of step tariff strategy, we can get energy conservation to some extent. Therefore, compared with electric energy production, system engineers can discover the trouble spot in the process of power transport.

(ii) The confidentiality of user's identity and power consumption should be guaranteed. If the smart grid does not take the security of user information into account, the residential user's privacy information would be revealed, and the real-time power consumption data could be modified. Therefore, the proposed scheme should achieve identity and data privacy preservation requirements simultaneously.

(iii) The proposed PAS scheme should be effective and practical to the smart grid. Although the performance of smart grid is continuously improved today, it still cannot afford the high computational overhead. The proposed scheme should also consider the effectiveness in terms of computation and communication to be acceptable to smart grid.

3. Preliminaries

In this section, we review the discrete logarithm problem and the Paillier cryptosystem [17], which support the security of signature and can serve as the basis of proposed PAS scheme.

3.1. Discrete Logarithm Problem. The difficulty of computing discrete logarithms [22] is the basic problem underlying the results of this thesis.

Definition 1. Let G be a finite cyclic group and let $g' \in G$ be a generator of G. Given an element $a \in G$, the discrete logarithm to the base g' of a in the group G is the unique integer b $(0 \le b \le |G|)$ such that $a = g'^{b}$.

Usually, the discrete logarithm is also called index of an element a. If g' is not a generator, the notion of discrete logarithm of a to the base g' is extended to be the smallest integer b, such that $a = g'^{b}$, if it exists.

Definition 2. The discrete logarithm problem (DLP) is stated as follows: given a finite cyclic group G, a generator g', and an element a of G, to find the integer b $(0 \le b \le |G| - 1)$ such that $a = g'^{b}$.

Definition 3 (computational Diffie-Hellman (CDH)). Let p and q be primes such that $q \mid (p-1)$. Suppose g is an element selected from Z_q^* with order q. Let $\mathscr{B}$ be an attacker. $\mathscr{B}$ tries to solve the following problem: *given* (g, g^a, g^b) *for uniformly chosen* $a, b \in Z_q^*$ *at random, compute* g^{ab}. We define $\mathscr{B}$'s advantage in solving the CDH problem by $\mathrm{Adv}(\mathscr{B}) = \Pr[\mathscr{B}(g, g^a, g^b) = g^{ab}]$.

3.2. Paillier Cryptosystem. The Paillier cryptosystem is widely used in privacy-preserving applications [23, 24], which can achieve the privacy-preserving properties. Specifically, the Paillier cryptosystem is comprised of three parts: key generation, encryption, and decryption.

(i) Key generation ($\mathrm{Gen}(k_1)$): given a security parameter k_1, then system selects two large primes p_1, p_2 $(|p_1| = |p_2| = k_1)$. Then, the RSA modulus $n = p_1 \cdot p_2$ and $\lambda = \mathrm{lcm}(p_1 - 1, p_2 - 1)$ are achieved. Define a function $L(u) = (u-1)/n$, and after choosing a generator $g \in Z_{n^2}^*$, $\mu = (L(g^{\lambda} \bmod n^2))^{-1}$ will be calculated. Finally, the public key is $pk = (n, g)$ and the corresponding private key is $sk = (\lambda, \mu)$.

(ii) Encryption: given a user's data $m \in Z_n$, choose a random number $r \in z_n^*$, and the ciphertext can be expressed as $c = E(m) = g^m r^n \bmod n^2$.

(iii) Decryption: given the ciphertext $c \in Z_{n^2}^*$, the corresponding data can be recovered as $m = D(c) = L(c^{\lambda \bmod n^2}) \cdot \mu \bmod n$.

4. Proposed Schemes

In this section, we propose the efficient privacy-preserving multidimensional aggregation scheme for smart grid, called PAS, which mainly consists of the following four parts: system initialization, user data acquisition, privacy-preserving regional aggregation, and data analysis. For easier expression, we give the description of notations used in PAS in the following notations.

Notations Used in PAS

$\mathscr{A}$: the large prime added to the last step of consumption,

p_1, p_2, f, p_1', p_2': five big primes, and $p_1 = 2fp_1' + 1$; $p_2 = 2fp_2' + 1$,

$(n = p_1 \cdot p_2, g)$: the public key of holomorphic encryption,

g': $g' \in Z_n^*$, $\langle g' \rangle$ generates a circle group with the order f,

g: $g \in Z_{n^2}^*$, a generator in the Paillier cryptosystem,

$h()$: $h()$ is a secure cryptographic hash function,

x_{TA}: $x_{\mathrm{TA}} \in Z_n^*$, the private key of TA,

y_{TA}: $y_{\mathrm{TA}} = g'^{x_{\mathrm{TA}}}$, the authentication public key of TA,

$\mathrm{ID}_{\mathrm{TA}}$: $\mathrm{ID}_{\mathrm{TA}} \in Z_n^*$, TA's identity number,

(e, d): $d \cdot e = 1 \bmod \phi(n)$,

k: a security parameter k for Gen(),

k_i: $k_i \in Z_n^*$, the private key of SM,

r_i: a random number $r_i \in Z_n^*$ chosen by SM_i,

x_{RC}: $x_{\mathrm{RC}} \in Z_n^*$, the private key of RC,

(n, g', y_{RC}): the public key of RC, $y_{\mathrm{RC}} = g'^{x_{\mathrm{RC}}}$ (mod n),

(λ, μ): the private key in the Paillier cryptosystem,

ω: the maximum number of households in a residential area,

l: the number of step electricity,

$(T_1, T_2, \dots, T_l)$: the power consumption at each step, and T_1 is the basic step,

d_2: the value of each type T_i is less than a constant d_2,

$\vec{a} = (a_1, a_2, \dots, a_l)$: $(a_2, \dots, a_l)$ are big prime numbers,

$\vec{g} = (g_1, g_2, \dots, g_l)$: $g_i = g^{a_i}$ (for $i = 1, 2, \dots, l$).

4.1. System Initialization. For a single authority smart grid system under consideration, a trusted TA bootstraps the whole system which is responsible for distributing secret keys to SM, RC, and OC. In particular, TA first chooses a security parameter k to obtain the Paillier cryptosystem's key $(g, \mu, \lambda, p_1, p_2, n = p_1 \cdot p_2)$ by running Gen(k), where $pk = (n, g)$ and $sk = (\lambda, \mu)$ are the public key and the corresponding private key, respectively, and then selects $g' \in Z_n^*$, where $\langle g' \rangle$ generates a circle group with the order f. Afterwards, TA chooses a secure cryptographic hash function $h()$, where $h : \{0,1\}^* \rightarrow Z_n^*$.

Assuming that the maximum number of SM in RC is less than ω, the power consumption at each step is $(T_1, T_2, \dots, T_l)$ in which T_1 is the basic step and the value of each type T_i is less than a constant d_2. A big prime vector $\vec{a} = (a_1 = 1, a_2, \dots, a_l)$ satisfies that the length $|a_i| \geq k$, $\sum_{j=1}^{i-1} a_j \cdot \omega \cdot d_2 < a_i$ $(i = 2, \dots, l)$, and $\sum_{i=1}^{l} a_j \cdot \omega \cdot d_2 < n$, and then the vector $\vec{g} = (g_1, g_2, \dots, g_l)$ is calculated as $g_i = g^{a_i}$ (for $i = 1, 2, \dots, l$). TA chooses an integer number e which satisfies $\gcd(e, \phi(n)) = 1$ and calculates the corresponding d according to $d \cdot e = 1 \bmod \phi(n)$. At the same time, TA selects a large prime $\mathscr{A}$ whose length is 30 bits less than q and $\mathscr{A}$ is much larger than the total power consumption of users in 15 minutes. In addition, TA chooses a random number $x_{\mathrm{TA}} \in Z_n^*$ as its private key and calculates the corresponding public key $y_{\mathrm{TA}} = g'^{x_{\mathrm{TA}}}$. RC chooses a random number $x_{\mathrm{RC}} \in Z_n^*$ as its private key and calculates the corresponding public key $y_{\mathrm{RC}} = g'^{x_{\mathrm{RC}}}$.

When a local SM_i in the residential area registers in the system, firstly SM_i chooses a random number k_i $(k_i \in Z_n^*)$ as its private key and calculates $\mathrm{ID}_i = g'^{k_i} \bmod n$ and then transmits ID_i to TA. Once getting the ID_i of SM, TA chooses a random number $\alpha \in Z_n^*$ and calculates $r = g'^{\alpha} \bmod n$, $s = \alpha + r x_{\mathrm{TA}} h(\mathrm{ID}_i) \bmod f$, and $w = (\mathrm{ID}_{\mathrm{TA}} r y_{\mathrm{TA}}^{rh(\mathrm{ID}_i)} \mathrm{ID}_i)^{-d} \bmod n$. TA stores the qualification certificate of $\mathrm{SM}_i(r, s, w, \mathrm{ID}_i)$ in the regional database.

Finally, TA sends $\langle r, s, w \rangle$ and Paillier cryptosystem's key $\langle \lambda, \mu \rangle$ to the registered SM_i and OC, respectively, by a secure channel, keeps $\langle x_{\mathrm{TA}}, q_1, q_2 \rangle$ secretly, and publishes the system parameters $\langle n, g, g', G, h(), \vec{a}, \vec{g}, \mathscr{A}, y_{\mathrm{TA}} \rangle$.

4.2. User Data Acquisition. In order to achieve the nearly real-time user's power consumption at every 15 minutes, each user in the regional area equips a SM_i to collect l types of data $(d_{i1}, d_{i2}, \dots, d_{il})$ which is the power consumption at each step $(T_1, T_2, \dots, T_l)$. And then SM_i calculates the encrypted data c_i and signature of SM_i as follows.

Step 1. SM_i gets the power consumption m_i, which is located at two-step interval $[k, k+1]$, and computes the real power consumption $(dt_{i1}, dt_{i2}, \dots, dt_{il})$ at each step where $dt_{i1} = T_1$, $dt_{ij} = T_j - T_{j-1}$ (for $2 \leq j \leq k$), $dt_{i(k+1)} = m_i - T_k$, and $dt_{ij} = 0$ (for $k + 1 < j \leq l$). Then SM_i obtains $(d_{i1}, d_{i2}, \dots, d_{il})$, where $d_{im} = dt_{im}$ $(m = 1, 2 , \dots, k, k+2 , \dots, l)$ and $d_{i(k+1)} = dt_{i(k+1)} + \mathscr{A}$.

Step 2. SM_i chooses a random number $r_i \in Z_n^*$ and then calculates the encrypted power consumption $c_i = \prod_{j=1}^{l} g_j^{d_{ij}} r_i^n \bmod n^2$.

Step 3. SM_i selects two random numbers $q_1, q_2 \in Z_n^*$ and uses the private key k_i to make the signature $\mathrm{Sig}_{\mathrm{SM}_i} = (u, v_1, v_2, \mathrm{TS})$ as

$$\begin{aligned} z &= q_2^e g'^{q_1} \bmod n, \\ u &= h(z, c_i, \mathrm{TS}), \\ v_1 &= q_1 + (s + k_i) u, \\ v_2 &= q_2 w^u \bmod n \end{aligned} \tag{1}$$

in which the TS is the current time stamp.

Step 4. SM_i sends the encrypted power consumption and the signature $\langle c_i, \mathrm{Sig}_{\mathrm{SM}_i}\rangle$ to RC.

4.3. Privacy-Preserving Regional Aggregation. While receiving $\langle c_i, \mathrm{Sig}_{\mathrm{SM}_i}\rangle$ $(i = 1, 2, \ldots, \omega)$ at time TS', RC first checks the validity of time stamp TS in order to resist the replaying attack. If $|\mathrm{TS}' - \mathrm{TS}| \le \Delta T$, where ΔT denotes the excepted valid time interval for transmission delay, the TS is accepted. Then, RC computes $z' = \mathrm{ID}_{\mathrm{TA}}^u g'^{v_1} v_2^e \bmod n$, $u' = h(z', c_i, \mathrm{TS})$ and checks whether $u = u'$. If it does hold, the signature is accepted, since

$$\begin{aligned} z' &= \mathrm{ID}_{\mathrm{TA}}^u g'^{v_1} v_2^e \bmod n = \mathrm{ID}_{\mathrm{TA}}^u g'^{q_1+(s+k_i)} \left(q_2 w^u\right)^e \\ &\quad \cdot \bmod n = \mathrm{ID}_{\mathrm{TA}}^u g'^{q_1} \left(r y_{\mathrm{TA}}^{rh(\mathrm{ID}_i)}\right)^u \\ &\quad \cdot \mathrm{ID}_i^u q_2^e \left(\mathrm{ID}_{\mathrm{TA}} r y_{\mathrm{TA}}^{rh(\mathrm{ID}_i)} \mathrm{ID}_i\right)^{-u} \bmod n = q_2^e g'^{q_1} \bmod n \\ &\quad = z \\ u' &= h\left(z', c_i, \mathrm{TS}\right). \end{aligned} \tag{2}$$

If the validity is checked, RC performs the following steps to achieve the privacy-preserving data aggregation.

Step 1. RC computes the aggregated and encrypted data C as $C = \prod_{i=1}^{\omega} c_i \bmod n^2$.

Step 2. RC chooses a random number $t \in Z_n^*$, which satisfies $\gcd(t, n-1) = 1$, and then uses private key x_{RC} to make a signature $\mathrm{Sig}_{\mathrm{RC}} = (\mathrm{TS}_{\mathrm{RC}}, R_{\mathrm{RC}}, S_{\mathrm{RC}})$ as $H_{\mathrm{RC}} = h(C, \mathrm{TS}, \mathrm{RC}_i)$, $R_{\mathrm{RC}} = g'^t (\bmod n)$, and $S_{\mathrm{RC}} = (H_{\mathrm{RC}} - x_{\mathrm{RC}} R_{\mathrm{RC}}) t^{-1} (\bmod (n-1))$, where $\mathrm{TS}_{\mathrm{RC}}$ is the current time stamp.

Step 3. RC sends the preliminary aggregation data and the signature $\langle C, \mathrm{Sig}_{\mathrm{RC}}\rangle$ to OC.

If the above verification fails, RC will verify the signature with the help of TA and announce the malicious ID_i. In this case, TA uses the unaccepted signature $\langle c_i, \mathrm{Sig}_{\mathrm{SM}_i}\rangle$ to compute $\eta = u^{-1} \bmod \phi(n)$ and $\delta = \mathrm{ID}_{\mathrm{TA}}^u g'^{v_1} \bmod n$ and then contrasts the stored (g'^s, w, ID_i) in the regional database to reveal the malicious identity ID_i that sent the unaccepted message as $\mathrm{ID}_i = (g'^{v_1} (\delta g'^{s \cdot u})^{-1})^{\eta} w^{-e} \bmod n = \mathrm{ID}_{\mathrm{TA}}^{-1} g'^{-s} (\mathrm{ID}_{\mathrm{TA}} g^{\alpha + x_{\mathrm{TA}} rh(\mathrm{ID}_i)} \mathrm{ID}_i)$. By doing this, RC can achieve the tracking of malicious operations.

4.4. Data Analysis. After receiving the message $\langle C, \mathrm{Sig}_{\mathrm{RC}}\rangle$ from RC, OC first checks the validity of time stamp $\mathrm{TS}_{\mathrm{RC}}$ with the receiving time $\mathrm{TS}'_{\mathrm{RC}}$. If $|\mathrm{TS}'_{\mathrm{RC}} - \mathrm{TS}_{\mathrm{RC}}| \le \Delta T$, the $\mathrm{TS}_{\mathrm{RC}}$ is accepted. Then RC uses the public key of RC y_{RC} to verify the signature $\mathrm{Sig}_{\mathrm{RC}}$ by checking whether $y_{\mathrm{RC}}^{R_{\mathrm{RC}}} R_{\mathrm{RC}}^{S_{\mathrm{RC}}} = g'^{h(\mathrm{C},\mathrm{TS},\mathrm{RC}_i)} \bmod n$. If it does hold, OC will decrypt the aggregated cipher report C as follows:

$$\begin{aligned} C &= \prod_{i=1}^{\omega} c_i \bmod n^2 = \prod_{i=1}^{\omega} g_1^{d_{i1}} g_2^{d_{i2}} \cdots g_l^{d_{il}} \cdot r_i^n \bmod n^2 \\ &= g_1^{\sum_{i=1}^{\omega} d_{i1}} g_2^{\sum_{i=1}^{\omega} d_{i2}} \cdots g_l^{\sum_{i=1}^{\omega} d_{il}} \prod_{i=1}^{\omega} r_i^n \bmod n^2 \\ &= g^{a_1 \sum_{i=1}^{\omega} d_{i1}} g^{a_2 \sum_{i=1}^{\omega} d_{i2}} \cdots g^{a_l \sum_{i=1}^{\omega} d_{il}} \prod_{i=1}^{\omega} r_i^n \bmod n^2 \\ &= g^{a_1 \sum_{i=1}^{\omega} d_{i1} + a_2 \sum_{i=1}^{\omega} d_{i2} + \cdots + a_l \sum_{i=1}^{\omega} d_{il}} \prod_{i=1}^{\omega} r_i^n \bmod n^2. \end{aligned} \tag{3}$$

Step 1. Let M and R be $a_1 \sum_{i=1}^{\omega} d_{i1} + a_2 \sum_{i=1}^{\omega} d_{i2} + \cdots + a_l \sum_{i=1}^{\omega} d_{il} \bmod n$ and $\prod_{i=1}^{\omega} r_i$, respectively, and $C = g^M \cdot R^n \bmod n^2$. According to $M = L(C^{\lambda \bmod n^2}) \cdot \mu \bmod n$, OC recovers M by the master key (λ, μ).

Step 2. By executing Algorithm 1, OC can recover the aggregated data $(D_1, D_2, \ldots, D_l)$, where $D_j = \sum_{i=1}^{\omega} d_{ij}$ is the power information at the jth step.

Step 3. After OC achieving power information at each step, OC invokes Algorithm 2 to obtain the real power consumption DT_j and the number of SM n_j $(j = 1, 2, \ldots, l)$ at each step, respectively.

Since $D_j = \sum_{i=1}^{\omega} d_{ij} = n_j \mathscr{A} + \sum_{i=1}^{\omega} dt_{ij}$ ($\sum_{i=1}^{\omega} dt_{ij}$ is real power consumption at jth step), $DT_j = D_j \bmod \mathscr{A} = \sum_{i=1}^{\omega} dt_{ij}$, and $n_j = (D_j - DT_j)/\mathscr{A}$, the correctness of Algorithm 2 is shown.

Finally, OC can obtain the valid information, which contains the true power consumption $(DT_1, DT_2, \ldots, DT_l)$ and the number of SM $(n_1, n_2, \ldots, n_l)$ at each step without knowing any private information of SM, and simultaneously OC accomplishes data interaction with power supplier.

5. Security Analysis

In this section, we analyze the security properties of PAS scheme. According to the security requirements discussed earlier, our analysis will focus on how PAS can achieve the privacy preservation of user's data and identity.

(i) The Confidentiality of User's Data and the Aggregated Data Are Obtained. In the proposed PAS scheme, each SM_i's power

```
Procedure: RECOVER THE AGGREGATED DATA
Input: $\vec{a} = (a_1 = 1, a_2, \ldots, a_l)$ and $M$
Output: $(D_1, D_2, \ldots, D_l)$
    set $X_{l+1} = M$
    for $j = 1$ to $l$ do
        $X_j = X_{j+1} \bmod a_j$
        $D_j = X_{j+1} - X_j / a_j \bmod q$
    end for
    return $(D_1, D_2, \ldots, D_l)$
end procedure
```

Algorithm 1: Recover the aggregated data.

```
Procedure: RECOVER THE POWER CONSUMPTION
Input: $(D_1, D_2, \ldots, D_l)$
Output: $(n_1, n_2, \ldots, n_l)$ and $(DT_1, DT_2, \ldots, DT_l)$
    set $(DT_1, DT_2, \ldots, DT_l) = (D_1, D_2, \ldots, D_l)$
    for $j = 1$ to $l$ do
        $DT_j = D_j \bmod \mathscr{A}$
        $n_j = (D_j - DT_j)/\mathscr{A}$
    end for
    return $(n_1, n_2, \ldots, n_l)$ and $(DT_1, DT_2, \ldots, DT_l)$
end procedure
```

Algorithm 2: Recover the power consumption.

consumption data $(d_{i1}, d_{i2}, \ldots, d_{il})$ are formed as a Paillier cryptosystem [17] ciphertext $c_i = g_1^{d_{i1}} \cdot g_2^{d_{i2}} \cdots g_l^{d_{il}} \cdot r_i^n \bmod n^2$, which can be implicitly expressed as $c_i = g^{a_1 d_{i1} + a_2 d_{i2} + \cdots + a_l d_{il}} \cdot r_i^n \bmod n^2$. Let m_i be $a_1 d_{i1} + a_2 d_{i2} + \cdots + a_l d_{il}$, and then the data $(d_{i1}, d_{i2}, \ldots, d_{il})$ in $c_i = g^{m_i} \cdot r_i^n \bmod n^2$ is also semantically secure and privacy preserving. Therefore, although an adversary $\mathscr{M}$ eavesdrops c_i, he/she still cannot recognize the unencrypted contents. After collecting all reports $(c_1, c_2, \ldots, c_\omega)$ from the residential users, RC does not recover each report; instead, it just computes $C = \prod_{i=1}^{\omega} c_i \bmod n^2$ to perform report aggregation. Therefore, even if the adversary $\mathscr{M}$ intrudes in RC's database, he/she cannot obtain the individual report $(d_{i1}, d_{i2}, \ldots, d_{il})$ either. Finally, after receiving $C = \prod_{i=1}^{\omega} c_i \bmod n^2$ from RC, OC recovers C as $(D_1, D_2, \ldots, D_l)$. However, since each $D_j = \sum_{j=1}^{\omega} d_i = n_j \mathscr{A} + \sum_{i=1}^{\omega} dt_{ij}$ is an aggregated result, OC still cannot obtain the user SM_i's data $(d_{i1}, d_{i2}, \ldots, d_{il})$. Therefore, from the above analysis, the user's data is privacy preserving in the proposed PAS scheme.

(ii) The Authentication of User's Data and the Aggregated Data Are Achieved. In the proposed PAS scheme, each user's data and the aggregated data are signed by an improved group signature algorithm.

Definition 4 (the improved efficient group signature scheme). The improved efficient group signature scheme is defined by the following algorithms.

Setup. Providing some security parameter k as input, the key generation center (KGC) runs this algorithm to create a master key x_{TA} and a list of public parameters *params*.

SetSecretValue. Taking *params* as input, this algorithm picks $k_i \in_R Z_n^*$ and outputs the secret value k_i.

SetPublicKey. Providing *params* and k_i as input, the user (the owner of k_i) runs this algorithm to set a public key ID_i.

PartialPrivateKeyExtract. Providing *params* and ID_i as input, KGC runs this algorithm to set (r, s, w) and sends to user by a secure channel.

Sign. Providing *params*, (r, s, w, ID_i), k_i, and a message M as input, any entity runs this algorithm to create a signature $\mathrm{Sig} = (u, v_1, v_2, M)$.

Verify. Providing *params*, x_{TA}, Sig, and ID_i, the receiver run this algorithm to verify the signature.

Definition 5 (security of the signature scheme). A signature scheme is said to achieve unforgeability against a chosen message attack (UF-CMA secure) if no polynomials bounded adversary $\mathscr{M}$ has a nonnegligible advantage in the following game.

Setup. The challenger runs setup by taking a security parameter k as input to generate a master key x_{TA} and a list of public parameters *params*. It gives *params* to adversary $\mathscr{M}$ and keeps x_{TA} secret.

SecretValue Request. The challenger runs SetSecretValue to generate the secret value k_i and return it to adversary $\mathscr{M}$.

PartialPrivateKey Request. The challenger runs PartialPrivateKeyExtract to generate the private key (r, s, w) and return it to adversary $\mathscr{M}$.

PublicKeyReplacement. The adversary $\mathscr{M}$ can repeatedly replace the public key for any entity with any value of its choice. The current value of an entity's public key is used by the challenger in any computation or response to the adversary's requests.

Verify Queries. The challenger runs Verify with x_{TA} and Sig as input and returns the verification result to adversary $\mathscr{M}$.

Sign Queries. The challenger runs Sign with (r, s, w, ID_i), k_i, and a message M as input to generate the Sig and then return it to adversary $\mathscr{M}$.

Challenge Phase. Once $\mathscr{M}$ decides that above is over, it outputs the challenge identity ID^* and two equal-length messages M_0, M_1. Note that ID^* has not been queried to extract a partial private key nor a private key at any time. The challenger picks $\beta \in_R \{0,1\}$ and creates a target signature Sig^* of M_β under the current public key $\mathrm{PK}_{\mathrm{ID}^*}$. The challenger returns Sig^* to $\mathscr{M}$.

Guess. $\mathscr{M}$ outputs its guess $\beta' \in \{0,1\}$ for β.

We define $\mathscr{M}$'s advantage in the above game by $\mathrm{Adv}(\mathscr{M}) = 2(\Pr[\beta' = \beta] - 1/2)$. A scheme is said to be UF-CMA secure if no probabilistic polynomial-time adversary has nonnegligible advantage in the above games.

Then, we analyze the security of the improved group signature.

Theorem 6. *One's improved efficient signature scheme is EUF-CMA secure in the random oracle model, assuming that the CDH problem is intractable.*

The above theorem is obtained by combining Lemma 7.

Lemma 7. *Suppose H_1, H_2 are random oracles and there exists a UF-CMA adversary $\mathscr{M}$ against the signature scheme with advantage ε when making q_{pv} PrivateValue requests, q_{ppk} PartialPrivateKey queries, q_{pkr} PublicKeyReplacement queries, q_s Sign queries, q_v Verify queries, and q_i random oracle queries to H_i ($i = 1$ or 2). Then, for any $0 \le \nu \le \varepsilon$ there exists either an algorithm $\mathscr{B}$ to solve the CDH problem with advantage $\varepsilon' \ge (\varepsilon - \nu)/q_2$ or an attacker that breaks the UF-CMA security of the RSA signature with advantage ν.*

Proof. To prove the lemma, we first assume that the RSA signature scheme is UF-CMA secure with advantage ν $(0 \le \nu \le \varepsilon)$ within time t'.

Let $\mathscr{M}$ be an adversary against the signature scheme. We show how to construct an algorithm $\mathscr{B}$ to solve the CDH problem.

$\mathscr{B}$ is given a random instance (g, g^a, g^b) of the CDH problem. $\mathscr{B}$ sets $y_{\mathrm{TA}} = g^a$ and simulates the *Setup* algorithm of the signature scheme by supplying $\mathscr{M}$ with $\langle n, e, g, H_1(), H_2(), y_{\mathrm{TA}} \rangle$ as public parameters, where H_1, H_2 are random oracles controlled by $\mathscr{B}$. $\mathscr{B}$ chooses a random I, $1 \le I \le n$.

$\mathscr{M}$ may make queries to random oracles H_i ($i = 1$ or 2) at any time during its attack and $\mathscr{B}$ responds as follows.

H_1 Queries. $\mathscr{B}$ maintains a H_1 list of tuples $\langle (i, \mathrm{ID}_i), e_i \rangle$. On receiving such a query on i, $\mathscr{B}$ does the following.

If there is a tuple $\langle (i, \mathrm{ID}_i), e_i \rangle$ on the H_1 list, $\mathscr{B}$ returns e_i as answer.

Otherwise, $\mathscr{B}$ chooses $e_i \in_R Z_q^*$, adds $\langle i, \mathrm{ID}_i, e_i \rangle$ to the H_1 list, and returns e_i as answer.

H_2 Queries. $\mathscr{B}$ maintains a H_2 list of tuples $\langle (z_i, M_i), u_i \rangle$. On receiving such a query on (z_i, M_i), if there is a tuple $\langle (z_i, M_i), u_i \rangle$ on the H_2 list, $\mathscr{B}$ returns u as answer. Otherwise, $\mathscr{B}$ picks $u_i \in_R Z_q^*$, adds $\langle (z_i, M_i), u_i \rangle$ to the H_2 list, and returns u_i as answer.

SecretValue Request. $\mathscr{B}$ maintains a SecretValue list of tuples $\langle i, \mathrm{ID}_i, k_i \rangle$. On receiving such a query on i, $\mathscr{B}$ responds as follows.

If $i \ne I$, chooses $k_i \in_R Z_n^*$ and computes public key $\mathrm{ID}_i = g^{k_i}$. Then $\mathscr{B}$ adds $\langle i, \mathrm{ID}_i, k_i \rangle$ to the public key list. Return k_i as answer.

Otherwise $(i = I)$, $\mathscr{B}$ aborts.

PartialPrivateKey Request. $\mathscr{B}$ maintains a PartialPrivateKey list of tuples $\langle i, \mathrm{ID}_i, (r, s, w) \rangle$. On receiving such a query on ID_i, $\mathscr{B}$ responds as follows.

If $i \ne I$, $\mathscr{B}$ searches the H_1 list for $\langle i, \mathrm{ID}_i, k_i \rangle$, chooses a random number $\alpha \in Z_n^*$, and calculates $r = g'^{\alpha} \bmod n$, $s = \alpha + r x_{\mathrm{TA}} h(\mathrm{ID}_i) \bmod f$, and $w = (\mathrm{ID}_{\mathrm{TA}} r y_{\mathrm{TA}}^{r h(\mathrm{ID}_i)} \mathrm{ID}_i)^{-d} \bmod n$. Then $\mathscr{B}$ adds (r, s, w) to the PartialPrivateKey list and returns it as answer.

Otherwise $(i = I)$, $\mathscr{B}$ aborts.

Public Key Replacement. $\mathscr{M}$ may replace the public key ID_i for any entity i with any value ID_i' of its choice. $\mathscr{B}$ records the change.

Verify Queries. Suppose the request is to verify signature $\mathrm{Sig} = (u, v_1, v_2, M)$ for ID_i. $\mathscr{B}$ searches the public key for ID_G. Then $\mathscr{B}$ does the following.

(1) If the public key has not been replaced and $i \ne I$, $\mathscr{B}$ searches the PartialPrivateKey list for a tuple $\langle i, \mathrm{ID}_i, (r, s, w) \rangle$ and computes $z = \mathrm{ID}_{\mathrm{TA}}^u g^{v_1} v_2^e \bmod n$. If $u = h(z_i, M_i, \mathrm{TS})$, return Y. Otherwise, output $\perp$.

(2) Otherwise, search the H_2 list for a tuple $\langle (z_i, M_i), u_i \rangle$ satisfying $u = u_i$, $z_i = \mathrm{ID}_{\mathrm{TA}}^u g^{v_1} v_2^e \bmod n$. Return Y if such a tuple exists. Otherwise, output $\perp$.

Signature Queries. $\mathscr{M}$ outputs ID^* and two messages M_0, M_1 on which it wishes to be challenged. Upon receiving ID^*, $\mathscr{B}$ searches the public key list for the tuple $\langle i, \mathrm{ID}_i, k_i \rangle$ and then conducts the following:

(1) picks $q_1^*, q_2^* \in_R Z_n^*$, $\beta \in_R \{0,1\}$, $\mathrm{ID}_i \in_R$,

(2) sets $v_1^* = q_1^* + (s^* + k_i^*) b$ and $s^* = H_1(\mathrm{ID}^*)$, $v_2^* = q_2^* {w^*}^{b} \bmod n$,

(3) defines $b = H_2(z^*, M_\beta)$ and $z^* = {q_2^*}^{e^*} g^{q_1^*} \bmod n$,

(4) computes $\mathrm{Sig}^* = (u^*, v_1^*, v_2^*, M_\beta^*)$ as the challenge signature.

$\mathscr{B}$ continues to respond to $\mathscr{M}$'s requests in the same way as it did above. Note that $\mathscr{M}$ cannot make a private key extraction query on ID^*.

Guess. Eventually, $\mathscr{M}$ outputs its guess β'. If $\beta' = \beta$, $\mathscr{B}$ chooses a random pair $\langle \mathrm{ID}_i, e_i \rangle$ from H_1 list and outputs $(g^{v_1^* - q_1^*} / \mathrm{ID}^* r^*)^{1/b e^* r^*}$ as the solution to the CDH problem. Otherwise, $\mathscr{B}$ cannot solve the CDH problem.

Analysis. We first evaluate the simulations of the random oracles. From the construction of H_1, it is clear that the simulation of H_1 is perfect. And as long as $\mathscr{M}$ does not query (z^*, M_β) to H_2, the simulation of H_2 is perfect. Let $\mathrm{Ask}H_2^*$ denote the event that (z^*, M_β) has been queried to H_2.

Let ¬Abort denote the event that $\mathscr{B}$ do not abort. Let $E = \text{Ask}H_2^* \mid \neg\text{Abort}$. It is clear that if E does not happen during the simulation, $\mathscr{B}$ will not gain any advantage greater than 1/2 to guess β. Namely, $\Pr[\beta' = \beta \mid E] \le 1/2$. We obtain $\Pr[\beta' = \beta] = \Pr[\beta' = \beta \mid E]\Pr[E] + \Pr[\beta' = \beta \mid \neg E]\Pr[\neg E]$.

By definition of ε, we have $\varepsilon - \nu \le (\varepsilon - \nu)/(1-\nu) \le \Pr[E] \le \Pr[\text{Ask}H_2^*]/\Pr[\neg\text{Abort}]$, $\Pr[\neg\text{Abort}] \ge 1/q_2$. Hence, it is not difficult for us to reach $\Pr[\text{Ask}H_2^*] \ge (\varepsilon - \nu)\Pr[\neg\text{Abort}] \ge (\varepsilon - \nu)/q_2$. The advantage for $\mathscr{B}$ to solve the CDH problem is $\varepsilon' \ge (\varepsilon - \nu)/q_2$.

The running time of the CDH bounded by $t' \le t + (q_1 + q_2)O(1) + q_{\text{ppk}}(2T_{\text{ex}} + O(1)) + (q_s + q_v)(3T_{\text{ex}} + O(1))$, where T_{ex} donates the time for computing exponentiation in Z_n^*. □

(iii) The Anonymity of Users and Traceability Can Be Guaranteed. In the proposed PAS scheme, each user's data is signed by an improved group signature.

Theorem 8. *Our improved efficient signature scheme satisfies the anonymity and traceability.*

Proof. Assume that the H_1 and H_2 are random functions. Given a valid signature $\text{Sig}_{\text{SM}_i} = (u, v_1, v_2, \text{TS})$, identifying the actual signer is computationally hard for everyone but the group manager (KGC). Deciding whether some group member with certificate originated requires computing the discrete logarithms $\log_g y_{\text{TA}}$. This is assumed to be infeasible under DLP and hence anonymity is guaranteed.

The group manager (KGC) is able to open any valid group signature and provably identify the actual signer: assuming that the signature is valid, this implies that u, v_1, v_2, are of the required form and so ID_i can be uniquely recovered as $\text{ID}_i = (g'^{v_1}(\delta g'^{s\cdot u})^{-1})^{\eta} w^{-e} \bmod n$.

According to the anonymity, it is impossible for the regional center to recognize user's identity. And if there are any errors in the process of verification, TA can recover the malicious SM's identity. Therefore, the identify privacy of user can be guaranteed.

From the above analysis, we can see that the proposed PAS scheme is secure and privacy preserving and achieves our security design goal. □

6. Performance Evaluation

In this section, we evaluate the performance of proposed PAS scheme in terms of the computation complexity and overhead of the SM, RC, and OC using a simulator built in Java. We run it on a workstation with 2.0 GHz 6-core processor, 64 GB RAM. Concretely, with the parameter setting $|n| = 1024$, we give that each RC includes 100 users and the number of RC is 10.

6.1. Computation Complexity. We evaluate the proposed PAS scheme in the computation complexity of SM, RC, and OC. Above all, we denote the computation overhead of an exponentiation operation in Z_{n^2}, a multiplication operation in Z_{n^2}, an exponentiation operation in Z_n, a multiplication operation in Z_n, and a multiplication operation in Z_{n-1} by C_1^e, C_1^m, C_2^e, C_2^m, and C_3^m, respectively.

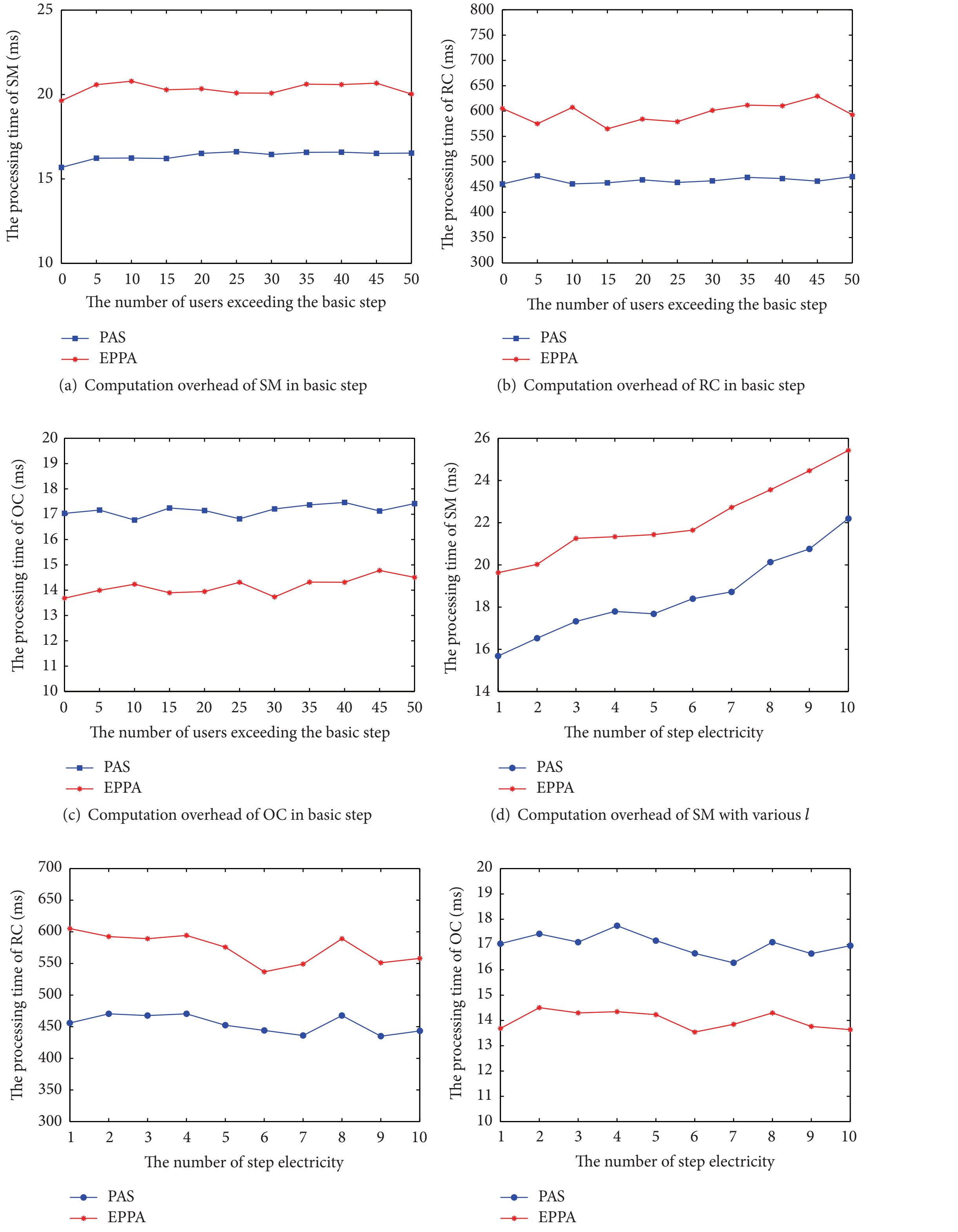
In the proposed PAS scheme, SM_i generates the encrypted power consumption c_i and the signature Sig_{SM_i} as shown in Section 4.2, which includes $(l+1) * C_1^e$ and $3 * C_2^m$. As shown in Section 4.2, after receiving the ciphertext from ω users, RC first verifies the received data from different users, then aggregates it, and makes a signature Sig_{RC}. The computation complexity of RC is $(3\omega+1) * C_2^e + 3 * C_2^m + (\omega - 1) * C_1^m + C_3^m$. With regard to OC, it verifies the aggregated data from RC and obtains valid user information by the Paillier decryption, which includes $3C_2^e$ and C_1^e.

Comparing with the EPPA [12], the proposed PAS scheme enables acquiring more effective information and guarantees the anonymity of user's identify simultaneously and has similar computation overhead as shown in Table 1, while C_m, C_{et}, C_p represent a multiplication operation in G, an exponentiation operation in G_T, and a pairing operation, respectively.

6.2. Simulation Results. According to the algorithms in our scheme, there are two factors that may affect the computation overhead. One is the number of users whose power consumption exceeds the basic step in RC which is denoted by r and the other is the number of step electricity l. Therefore, r and l are chosen to illustrate the computation overhead of SM, RC, and OC, and we present the comparison between PAS and EPPA. Note that EPPA cannot disclose users' identify and cannot provide the detailed information (e.g., users' numbers and power consumptions at each step in different dimensions).

For the comparison with EPPA, we plot the computation overhead of SM, RC, and OC in basic step with 10 different r from 0 to 50 in Figures 3(a), 3(b), and 3(c). They have shown that both PAS's and EPPA's processing times increase negligibly with the increase of r. The average processing times of SM and RC in PAS are 16 ms and 460 ms, respectively, which are a little less than those in EPPA. Then, we consider the processing time of OC. The time used to recover M of OC in PAS approaches to 17 ms which is higher than in EPPA. From the figure, we can see that, by increasing r, the average processing times of SM, RC, and OC are barely growing, which means r has a little influence on the computation overhead of SM, RC, and OC.

As shown in Figures 3(d), 3(e), and 3(f), we plot the PAS and EPPA's computation overhead of SM, RC, and OC with the number of step consumption l selected from 1 to 10. We can see that, with the increase of l, the average processing times of SM also increase in PAS and EPPA. The reason is that SM's power consumption may be in a higher step with l becoming larger and the SM will take more calculation on average. Afterwards, the average processing times of RC and OC are given, and by increasing l, the average operation times of RC and OC are barely growing. Thus, the number of step consumption l has a great influence on SM, but little on RC and OC. It is shown that the average processing times of SM and RC in PAS are a little less than that in EPPA, but the average processing time of OC in PAS is higher than that of EPPA.



(a) Computation overhead of SM in basic step

(b) Computation overhead of RC in basic step

(c) Computation overhead of OC in basic step

(d) Computation overhead of SM with various *l*

(e) Computation overhead of RC with various *l*

(f) Computation overhead of OC with various *l*

Figure 3: The computation overhead of SM, RC, and OC.

Table 1: Comparisons between PAS and EPPA scheme.

	PAS	EPPA
Smart meter	$(l+1) * C_1^e + 3 * C_2^m$	$(l+1) * C_1^e + C_m + 4 * C_p$
Regional center	$(3\omega+1) * C_2^e + 3 * C_2^m + (\omega-1) * C_1^m + C_3^m$	$(\omega+3) * C_p + C_m$
Operation center	$3C_2^e + C_1^e$	$2 * C_p + C_1^e + 4 * C_m + C_{\text{et}}$

From the evaluation results above, the PAS scheme has similar computation overhead with the EPPA scheme. Furthermore, the processing time of users is less than 20 ms which is acceptable to SM. And for RC and OC, they only need 400 ms and 18 ms, respectively, which is very efficient. The results of experiment show that our proposed PAS scheme is not only privacy preserving but also efficient and has stable computation performance.

7. Related Works

The study of privacy-preserving multidimensional aggregation for smart grid has gained great interest from the research community recently, and we briefly review some of them closely related to ours.

Part of the existing schemes needs to decrypt the received data before aggregating them and then encrypt the aggregated result before forwarding it. This process is fairly expensive and risky when aggregator is not trusted. Castelluccia et al. [25] designed a symmetric-key aggregation scheme that blends inexpensive encryption techniques with simple aggregation methods to achieve very efficient aggregation of encrypted data. Although the aggregator is under attack, the adversary cannot acquire any privacy-sensitive information about residents.

Then, differential privacy techniques are introduced to achieve multidimensional aggregation in smart grid without decrypting data, and the middle aggregator can be untrusted [26, 27]. Shi et al. [26] proposed a construction that could allow a group of participants to periodically upload encrypted values to a data aggregator. First, it utilizes applied cryptographic techniques to allow the aggregator to decrypt the sum from multiple ciphertexts encrypted under different users' keys. Second, it describes a distributed data randomization procedure that guarantees the differential privacy of outcome statistic, even when a subset of participants might be compromised. Chan et al. [27] offered a protocol with an untrusted aggregator as well. Specifically, the protocol builds a binary interval tree over n users and allows the aggregator to estimate the sum of contiguous intervals of users represented by nodes in the interval tree. In order to achieve multidimensional privacy-preserving data aggregation, Kursawe et al. [14] designed a protocol that can be employed to compute aggregate meter measurements over defined sets of meters, allowing for fraud and leakage detection as well as further statistical processing of meter measurements, without revealing any additional information about the individual meter readings. Lin et al. [28] proposed a scheme to achieve privacy-preserving aggregation that can integrate the superincreasing sequence and perturbation techniques into compressed data aggregation and has the ability to combine more than one aggregated piece of data into one. Jia et al. [29] proposed a novel privacy-preserving data aggregation scheme which could support efficient data aggregation for time-series metering data without leaking the individual value. It could also thwart HAD attack by introducing some randomness to the aggregation result without affecting the aggregation utility. Lu et al. [12] thought the scheme in [28] is not practical to deploy and manage thousands of keys for users and RC and proposed the EPPA based on public cryptograph system.

However, almost all the existing aggregation schemes focus on the privacy of users' information, cannot afford more valid information to the service provider (e.g., the power supplier) for allocating resources, and neglect the anonymity of user's identify. In this paper, we propose an efficient privacy-preserving multidimensional aggregation scheme, which can achieve more information from the aggregation data without disclosing user's identity. Specifically, the employment of superincreasing sequence techniques and the importation of a big prime in the Paillier cryptosystem ensure that the scheme provides more effective information, and a group signature algorithm is adopted to protect the user's identify from RC.

8. Conclusions

In this paper, we propose an efficient privacy-preserving multidimensional aggregation scheme for smart grid, named PAS. Without disclosing the privacy of user, we can acquire more valid information about the number of users and power consumption at each step after the detailed analysis of preliminary aggregated data by the operation center. Specifically, we employ an improved Paillier cryptosystem technique to obtain more information at the operation center and an efficient anonymous authentication scheme to protect the privacy of user's identity from the regional center. Detailed security analysis shows the security strength and privacy-preserving ability of PAS. Moreover, with reasonable computation overhead, PAS can satisfy the high-frequency multidimensional data collection requirements in smart grid. For the future work, we will focus on reducing the computation overhead of smart meter and achieving a more efficient security scheme for smart grid.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research was supported by National Natural Science Foundation of China (nos. 61303218 and 61272457), Huawei Innovation Research Program, and the 111 Project (B08038).

References

[1] The 22th World Energy Congress, 2015, http://www.worldenergy.org/.

[2] V. C. Güngör, D. Sahin, T. Kocak et al., "Smart grid technologies: communication technologies and standards," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 4, pp. 529–539, 2011.

[3] F. Li, W. Qiao, H. Sun et al., "Smart transmission grid: vision and framework," *IEEE Transactions on Smart Grid*, vol. 1, no. 2, pp. 168–177, 2010.

[4] K. Moslehi and R. Kumar, "A reliability perspective of the smart grid," *IEEE Transactions on Smart Grid*, vol. 1, no. 1, pp. 57–64, 2010.

[5] D. Niyato, L. Xiao, and P. Wang, "Machine-to-machine communications for home energy management system in smart grid," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 53–59, 2011.

[6] Z. M. Fadlullah, M. M. Fouda, N. Kato, A. Takeuchi, N. Iwasaki, and Y. Nozaki, "Toward intelligent machine-to-machine communications in smart grid," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 60–65, 2011.

[7] H. Liang, B. J. Choi, W. Zhuang, and X. Shen, "Towards optimal energy store-carry-and-deliver for PHEVs via V2G system," in *Proceedings of the IEEE Conference on Computer Communications (INFOCOM '12)*, pp. 1674–1682, Orlando, Fla, USA, March 2012.

[8] V. C. Gungor, D. Sahin, T. Kocak et al., "A survey on smart grid potential applications and communication requirements," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 1, pp. 28–42, 2013.

[9] S. Goguri, J. Hall, R. Mudumbai, and S. Dasgupta, "A distributed, real-time and non-parametric approach to demand response in the smart grid," in *Proceedings of the 49th Annual Conference on Information Sciences and Systems (CISS '15)*, pp. 1–5, Baltimore, Md, USA, March 2015.

[10] H. Zhu, T. Liu, G. Wei, and H. Li, "PPAS: privacy protection authentication scheme for VANET," *Cluster Computing*, vol. 16, no. 4, pp. 873–886, 2013.

[11] R. Lu, X. Lin, Z. Shi, and X. S. Shen, "A lightweight conditional privacy-preservation protocol for vehicular traffic monitoring systems," *IEEE Intelligent Systems*, vol. 28, no. 3, pp. 62–65, 2013.

[12] R. Lu, X. Liang, X. Li, X. Lin, and X. Shen, "EPPA: an efficient and privacy-preserving aggregation scheme for secure smart grid communications ," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 9, pp. 1621–1632, 2012.

[13] F. Li, B. Luo, and P. Liu, "Secure information aggregation for smart grids using homomorphic encryption," in *Proceedings of the 1st IEEE International Conference on Smart Grid Communications (SmartGridComm '10)*, pp. 327–332, IEEE, Gaithersburg, Md, USA, October 2010.

[14] K. Kursawe, G. Danezis, and M. Kohlweiss, "Privacy-friendly aggregation for the smart-grid," in *Privacy Enhancing Technologies*, vol. 6794 of *Lecture Notes in Computer Science*, pp. 175–191, Springer, Berlin, Germany, 2011.

[15] Y. Yan, Y. Qian, and H. Sharif, "A secure data aggregation and dispatch scheme for home area networks in smart grid," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '11)*, pp. 1–6, Houston, Tex, USA, December 2011.

[16] K. Ma, G. Hu, and C. J. Spanos, "Distributed energy consumption control via real-time pricing feedback in smart grid," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 5, pp. 1907–1914, 2014.

[17] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Proceedings of the 17th International Conference on Theory and Application of Cryptographic Techniques (EUROCRYPT '99)*, pp. 223–238, Prague, Czech Republic, May 1999.

[18] C. Castelluccia, A. C.-F. Chan, E. Mykletun, and G. Tsudik, "Efficient and provably secure aggregation of encrypted data in wireless sensor networks," *ACM Transactions on Sensor Networks*, vol. 5, no. 3, 2009.

[19] D. Westhoff, J. Girao, and M. Acharya, "Concealed data aggregation for reverse multicast traffic in sensor networks: encryption, key distribution, and routing adaptation," *IEEE Transactions on Mobile Computing*, vol. 5, no. 10, pp. 1417–1431, 2006.

[20] J. Shi, R. Zhang, Y. Liu, and Y. Zhang, "PriSense: privacy-preserving data aggregation in people-centric urban sensing systems," in *Proceedings of the IEEE (INFOCOM '10)*, pp. 758–766, San Diego, Calif, USA, March 2010.

[21] G.-D. Si, Y.-P. Li, and G.-Z. Xiao, "An improvement on group signature," *Journal of Xidian University*, vol. 34, no. 1, pp. 106–121, 2007.

[22] J. Camenisch and J. Groth, "Group signatures: better efficiency and new theoretical aspects," in *Security in Communication Networks (SCN '04)*, pp. 120–133, Springer, 2005.

[23] Y. Sang, H. Shen, and H. Tian, "Privacy-preserving tuple matching in distributed databases," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 12, pp. 1767–1782, 2009.

[24] S. Zhong, "Privacy-preserving algorithms for distributed mining of frequent itemsets," *Information Sciences*, vol. 177, no. 2, pp. 490–503, 2007.

[25] C. Castelluccia, E. Mykletun, and G. Tsudik, "Efficient aggregation of encrypted data in wireless sensor networks," in *Proceedings of the 2nd Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (MobiQuitous '05)*, pp. 109–117, July 2005.

[26] E. Shi, T. Chan, E. G. Rieffel, R. Chow, and D. Song, "Privacy-preserving aggregation of time-series data," in *Proceedings of the Network and Distributed System Security Symposium (NDSS '11)*, vol. 2, p. 4, San Diego, Calif, USA, February 2011.

[27] T.-H. H. Chan, E. Shi, and D. Song, "Privacy-preserving stream aggregation with fault tolerance," in *Financial Cryptography and Data Security*, pp. 200–214, Springer, Berlin, Germany, 2012.

[28] X. Lin, R. Lu, and X. S. Shen, "MDPA: multidimensional privacy-preserving aggregation scheme for wireless sensor networks," *Wireless Communications and Mobile Computing*, vol. 10, no. 6, pp. 843–856, 2010.

[29] W. Jia, H. Zhu, Z. Cao, X. Dong, and C. Xiao, "Human-factor-aware privacy-preserving aggregation in smart grid," *IEEE Systems Journal*, vol. 8, no. 2, pp. 598–607, 2014.

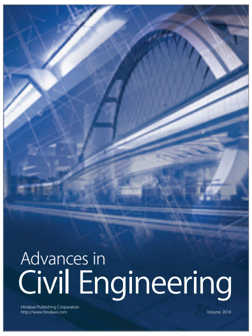
Journal of
Engineering

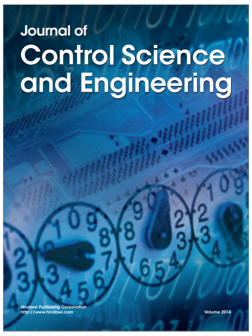
The Scientific World Journal

International Journal of
Rotating Machinery

Journal of
Sensors

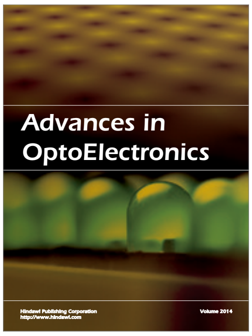
International Journal of
Distributed Sensor Networks


Advances in
Civil Engineering


Journal of
Control Science and Engineering

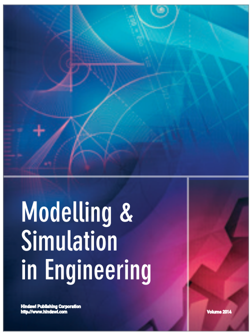
Journal of
Robotics

Journal of
Electrical and Computer Engineering

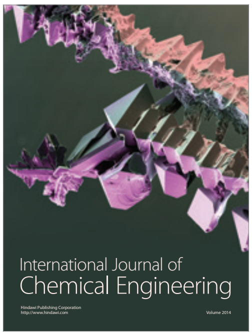

Advances in
OptoElectronics

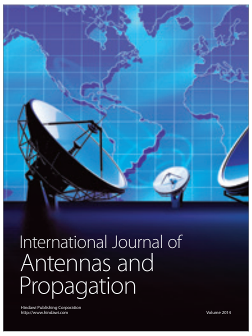
VLSI Design

International Journal of
Navigation and Observation

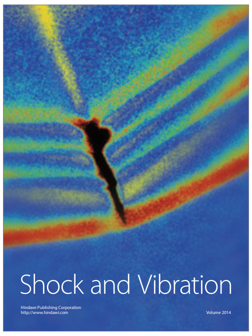

Modelling & Simulation in Engineering

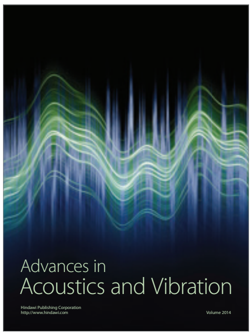
International Journal of
Aerospace Engineering


International Journal of
Chemical Engineering


International Journal of
Antennas and Propagation

Active and Passive Electronic Components


Shock and Vibration


Advances in
Acoustics and Vibration