Hindawi
Security and Communication Networks
Volume 2022, Article ID 5895176, 18 pages
https://doi.org/10.1155/2022/5895176

WILEY | Hindawi

Research Article

Efficient Privacy-Preserving Data Aggregation Scheme with Fault Tolerance in Smart Grid

Yang Ming, Yabin Li, Yi Zhao, and Pengfei Yang

School of Information Engineering, Chang'an University, Xi'an 710064, China

Correspondence should be addressed to Yang Ming; yangming@chd.edu.cn

Received 8 October 2021; Revised 25 November 2021; Accepted 15 December 2021; Published 25 January 2022

Academic Editor: Yu Yao

As the traditional grid produces a large amount of greenhouse gas and cannot adapt to such new demands as dynamic electricity prices, data analysis, and early warning, smart grid with high efficiency and reliability is increasingly valued. It plays a key role in achieving carbon neutrality. Nonetheless, smart grid requires the collection of real-time power data, and personal privacy may be leaked through the frequent electricity measurement reports. With the requirements of data analysis and prediction while preserving users' personal privacy, data aggregation schemes have emerged. However, existing schemes cannot resolve all the troubles well. Some schemes do not consider the failures for smart meters, and most of the schemes have expensive computation cost. In view of this, an efficient privacy-preserving data aggregation scheme with fault tolerance in smart grid is put forward in this paper. To be specific, the proposed scheme is lightweight due to the application of the symmetric homomorphic encryption technology and the elliptic curve cryptography. Even if some smart meters are destroyed, the proposed scheme can still successfully obtain aggregated data. Moreover, the proposed data aggregation scheme is proved to be secure, and all security requirements can be satisfied. Performance evaluation illustrates the relatively low computation cost and communication overhead of the proposed scheme compared to other related schemes.

1. Introduction

In recent years, the negative effects of global warming have become increasingly significant, as can be observed from the rising sea levels and the destruction of biodiversity. All countries are looking for ways to achieve carbon neutrality [1]. The application of smart grid (SG) can effectively accelerate the realization of this goal, and SG is included in long-term development plans [2–6]. Compared with traditional grid, SG has the advantages of conforming to low-carbon sustainable development, adopting a two-way communication mode, and allowing for diversified gradient electricity prices and early warning based on status analysis [7–9]. These features compensate for various shortcomings of traditional power grid; therefore, SG is considered an excellent next-generation power system. Figure 1 illustrates the framework of SG, which consists of the markets, control center, service provider, energy generation, transmission, distribution, and customers [10].

For the information communication in SG, a large number of sensors are employed, especially smart meters (SM), which need to collect real-time household power measurement data every 10–15 minutes and send them to the control center (CC) for electricity data analysis and dispatch [11]. It is very time-consuming for a large amount of data transmission; at the same time, real-time data transmission also raises people's privacy concern. According to the survey [12, 13], individual real-time electricity consumption data will expose sensitive information of users; for example, the lifestyle and living habits of family members might be exploited by malicious adversary.

In order to deal with the above contradiction simultaneously, data aggregation technology has been proposed, where SM will use homomorphic encryption to protect real-time power data and upload ciphertext to gateway (GW); then, data ciphertext is aggregated by GW and sent to CC. Finally, CC can take advantage its private key to decrypt the aggregated ciphertext, but it is unable to gain a single power

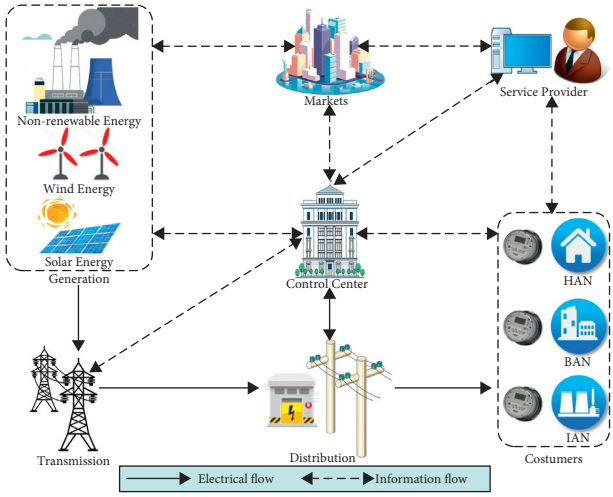

Figure 1: Framework of the smart grid.

measurement data of a SM. In this way, the users' personal privacy information can be protected, and in the meanwhile, CC can analyze power measurement data and allocate and adjust power supply in a timely and reasonable manner.

Although there have been many data aggregation schemes [14–37], many issues are still worthy of further improvement.

First, SM exposed to an environment without any protection may malfunction and cannot send reports; the lack of partial power data reports may make the system fail to recover the aggregated power data. The feature of fault tolerance enables the recovery of the aggregated data despite the SM malfunction. Some previously existing aggregation schemes [14, 16–20] do not support fault tolerance, so they cannot obtain normally aggregated data, and the entire system will be paralyzed. Although some other schemes [22, 24, 29] achieve fault tolerance, their trusted authority (TA) or CC needs to do some special operations, such as generating dummy ciphertext; this is not very practical because the number of GW and SM under the control of the TA is enormous, which will bring unbearable computation costs to the TA.

Second, since the GW and CC are semitrusted, they may launch collision attacks to obtain the private data of a single SM. In the existing schemes [22, 28, 30, 37], the data reports are encrypted directly with the public key of CC. If GW sends the ciphertext of a certain SM to CC, or CC accidentally obtains the single ciphertext, the individual report can be decrypted by CC through its private key.

Third, identity privacy is also a kind of secret information, which should be protected; meanwhile, when malicious SM appear, its real identity should be revealed by the TA. Some existing schemes [18, 23, 25, 26] fail to consider identity privacy, some other schemes [16, 17, 20, 27] achieve identity anonymity, but the way is that the data reports do not contain identity information, which makes it impossible to trace the identity of the malicious SM when it appears.

In order to settle the above problems and realize further optimization, an efficient data aggregation scheme that would support fault tolerance is proposed. The primary contributions are shown below:

(i) The proposed scheme applies lightweight symmetric homomorphic encryption technology and elliptic curve signature to accomplish efficiency, instead of commonly used time-consuming public key homomorphic encryption technologies such as Paillier [38] and BGN [39]. It is also characterized by the feature of fault tolerance, thus being able to run normally even if some SM fail to upload data reports.

(ii) The security analysis formally proves that the proposed aggregation scheme is secure based on $(\mathscr{L}, \widehat{p})$-based decision problem and elliptic curve discrete logarithm problem. Moreover, the proposed scheme would implement required security requirements, especially to resist collusion attacks.

(iii) Performance evaluation carries out quantitative analysis, and the result displays; the proposed scheme involves less computation cost and communication overhead compared with other related data aggregation schemes.

The structure of the rest of this paper is allocated as follows. Section 2 displays the related works of data aggregation schemes. The background and preliminaries are given in Sections 3 and 4, respectively. In Section 5, the proposed scheme is introduced in detail. Sections 6 and 7, respectively, illustrate the security analysis and the performance evaluation. Ultimately, the conclusion is described in Section 8.

2. Related Work

With the long-term research of data aggregation technology, many problems are considered to satisfy the security requirements, for instance, collision attack, fault tolerance, and identity privacy protection.

The related aggregation schemes are vulnerable to various attacks, especially collision attacks. Compared with external adversaries, internal attackers are more likely to damage the SG system because they have more private information. Fan et al. [14] first considered collision attacks and successfully resisted them by virtue of blinding factors assigned by a trusted third party. Regrettably, Bao and Lu [15] illustrated the integrity drawback of the scheme [14], which lies in that the private key of the user was easily recovered so that data pollution would be caused. He et al. [16] created the certificateless data aggregation scheme by the mechanism of elliptic curve cryptography which could speed up the process and withstand the collision attacks. He et al. [17] improved the BGN scheme to realize data aggregation scheme against collision attacks. Zhang et al. [18] considered the false data injection attacks and prevented them with the blinding factors. Li et al. [19] applied the BGN encryption and blinding factors to complete data aggregation scheme that can prevent collision attacks. Shen et al.

[20] put forward the aggregation scheme that can counteract new malicious data mining attacks and internal attacks with BLS short signature. Based on elliptic curve cryptography, a scalable data aggregation scheme was designed by Chen et al. [21], where the encryption key, instead of the public key of CC, was generated independently, even if CC cannot decrypt the single ciphertext.

Once SM malfunction appears and hinders the normal submission of electricity reports, most of the above schemes [14, 16–20] would be paralyzed. Therefore, fault tolerance needs to be taken into consideration. In the scheme [22] of Chen et al., a trusted third party additionally generated the dummy ciphertext for the damaged SM to ensure the smooth running of the agreement as the trusted third party held the private keys of all users. Bao and Lu [23] advanced the differentially private aggregation scheme with fault tolerance, where CC was still able to receive the aggregated data from the remaining reports. Pan et al. [24] combined with Lagrangian interpolation technology to propose a two-dimensional and fault-tolerable privacy-preserving aggregation scheme. Ge et al. [25] put forward a fine-grained data analysis scheme which could still run even if the meter was failed, and this scheme could obtain a variety of statistics. Xue et al. [26] proposed the privacy-preserving service outsourcing scheme, which supported fault tolerance mechanism and flexible electricity price. Guan et al. [27] utilized the Shamir sharing method and RSA signature to implement the fault tolerant aggregation protocol, yet this protocol cannot be decrypted correctly. In [28], Ding et al. raised an identity-based secure data aggregation scheme, which supported fault tolerance due to their particular ciphertext structure. Wang et al. [29] skillfully improved the Paillier encryption to get the fault tolerant data aggregation scheme through collaborating between users; unfortunately, the integrity was not considered here. In general, when facing SM malfunction, TA will perform additional operations to achieve fault tolerance, but it will be overloaded because it manages many GW and many SM under GW.

In addition, user identity privacy is an important security problem. Liu et al. [30] utilized the blind signature technology to realize an anonymous data aggregation scheme, where the token was unlinkable to any valid signature. Combined with the ring signature technology, Badra and Zeadally [31] blocked the connection between the content of the report and the identity of the SM. Tan et al. [32] designed a privacy-preserving pseudonym-based collection scheme, where the SM adopted the group key to generate a pseudoidentity so that the adversary was unable to get its real identity. Gong et al. [33] satisfied anonymity by separating data reports and identity of the SM.

Although the above schemes achieve different functions and features, most of them are time-consuming, which can make SM with limited calculation resources embarrassing. He et al. [34] applied batch verification to accelerate the execution of the aggregation scheme. Combining the elliptic curves cryptography and super-increasing sequence technology, Ming et al. [35] came up with an efficient privacy-preserving multidimensional data aggregation scheme, which can classify power measurement data and achieve fine-grained analysis. In scheme [36] of Shen et al., XOR operation of pseudorandom function was employed to encrypt power data and realize confusion so that the adversary could not identify the source of the reports. Zhang et al. [37] adopted online and offline signature technology to create a lightweight aggregation scheme, which would help to speed up the signature verification process.

3. Background

The background of the proposed scheme is described, mainly including system model, security requirements, and design goal in this section.

3.1. System Model. In the proposed scheme, the system model is divided into four entities: trusted authority (TA), control center (CC), gateway (GW), and n smart meters (SM), as shown in Figure 2. For the ease of description, considering only one GW, we link $n\,(n > 1)$ smart meters in the model.

(1) TA: it is a fully trusted entity, who produces the blinding factors and the secret value for SM. TA would recover the exact identity of the SM with malicious behavior.

(2) CC: it is a semitrusted entity, who generates the system parameters. CC is also in charge of the registration of SM and GW. In addition, after receiving aggregated encrypted data from GW, CC will decrypt and analyze them.

(3) GW: it is a semitrusted entity. GW collects and aggregates the encrypted electricity report from each domain header; then, GW transmits it to CC.

(4) SM: according to geographical proximity principle, entire n SM are divided into ω domains $D_1, D_2, \ldots, D_\omega$, and each domain $D_i\,(i = 1, 2, \ldots, \omega)$ contains t members, that is, $\omega \cdot t = n$. In the ith domain D_i, a random member is selected as the domain header. Without loss of generality, assuming that the first member SM_{i1} is appointed as the domain header SM_{i1}^H by GW, obviously, the domain header SM_{i1}^H itself is also a member in D_i. Here, the domain member $SM_{ij}\,(j = 1, 2, \ldots, t)$ is charge of collecting electricity measurement data of each user's household and sending it to SM_{i1}^H. The domain header SM_{i1}^H is responsible for preaggregating the data report in the domain D_i and then uploading it to GW. Besides, SM_{ij} is not allowed to send electricity report directly to GW. It is worth noting that all SM cannot collude with GW or CC.

3.2. Security Requirements. The security requirements that the proposed scheme should satisfy are as follows:

(1) Confidentiality: the electricity data are closely related to users' privacy information. Therefore, only useless knowledge can be obtained even if the adversary gets the transmitted ciphertext.

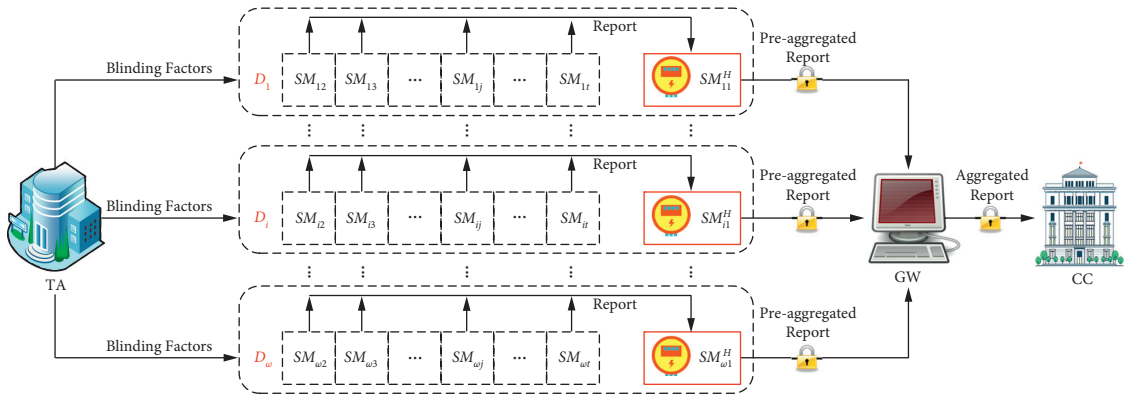

Figure 2: System model.

(2) Authentication: it is necessary to realize the authentication because the report transmission between any two entities must verify each other to ensure legal identity.

(3) Integrity: the report transmitted in the open channel may be tampered, and wrong message may be conveyed. So, the proposed scheme would detect whether the report has been altered.

(4) Anonymity and traceability: no entity other than TA can determine or distinguish identity by analyzing transmitted reports. From another aspect, when a malicious SM uploads fake data, its true identity should be revealed by TA to supervise the behavior.

(5) Resistance against common attacks: the proposed scheme should guarantee that many common types of attacks would be rejected, including but not limited to collision attack, modification attack, and replay attack.

3.3. Design Goal. The proposed scheme satisfies the following the objectives.

(1) Privacy-preserving: the actual data and identity of a single SM are prohibited from being obtained by anyone. CC is only allowed to decrypt aggregated data ciphertext. In addition, the abovementioned attacks should be resisted.

(2) Fault tolerance: it is unbearable that the aggregated data cannot be recovered when few SM are damaged. Therefore, even if some SM fail to submit reports, the system should continue to run normally.

(3) High efficiency: on the premise of fulfilling the above security requirements, the proposed scheme tries to reduce the computation cost and communication overhead. For practical smart grid, an efficient scheme is more suitable for SM with limited resources.

4. Preliminaries

Two preknowledge are briefly stated in this section, including elliptic curve cryptography and symmetric homomorphic encryption.

4.1. Elliptic Curve Cryptography. The definition of elliptic curve cryptography (ECC) comes from Millier [40]. Let $\mathbb{G}$ be an additive cyclic group of prime order q; the generator is P. The security problem and assumption are described as follows.

Elliptic curve discrete logarithm problem (ECDL problem) [41]: given two elements $P, Q \in \mathbb{G}$ in the elliptic curve E as input, output an integer $a \in \mathbb{Z}_q^*$ where $Q = aP$.

Elliptic curve discrete logarithm assumption (ECDL assumption) [41]: it is difficult for the probabilistic polynomial time algorithm to solve the ECDL problem with a nonnegligible advantage.

4.2. Symmetric Homomorphic Encryption. Mahdikhani et al. [42] designed a new symmetric homomorphic encryption; the algorithm is described as follows.

KeyGen(τ): on inputting the security parameter (k_0, k_1, k_2) satisfying $k_1 \ll k_2 < (k_0/2)$, the probabilistic key generation algorithm outputs the symmetric homomorphic encryption key $K = \langle \widehat{p}, \widehat{q}, \mathscr{L} \rangle$, where the two large prime numbers $\widehat{p}, \widehat{q}$ satisfy $|\widehat{p}| = |\widehat{q}| = k_0$ and $\mathscr{L}$ is randomly selected from $\{0,1\}^{k_2}$. Next, the algorithm calculates $N = \widehat{p}\widehat{q}$ and publishes the system parameters $PP = \langle k_0, k_1, k_2, N \rangle$.

Enc(K, m, r, r'): on inputting the symmetric homomorphic encryption key K and the plaintext $m \in \mathscr{M}$, where the message space $\mathscr{M}$ is $\{0,1\}^{k_1}$, the encryption algorithm selects two random numbers $r \in \{0,1\}^{k_2}$ and $r' \in \{0,1\}^{k_0}$ and encrypts the plaintext:

$$c = (r\mathscr{L} + m)(1 + r'\widehat{p}) \bmod N. \tag{1}$$

Dec(K, c): on inputting the ciphertext c and the symmetric homomorphic encryption key K, the decryption algorithm decrypts the ciphertext:

$$m = (c \bmod \widehat{p}) \bmod \mathscr{L}. \tag{2}$$

The security of symmetric homomorphic encryption [42] is based on the following security assumption.

$(\mathscr{L}, \widehat{p})$-based decision problem [43]: given (k_0, k_2, N), the $(\mathscr{L}, \widehat{p})$-based decision problem is to determine whether an integer $x \in \mathbb{Z}_N$ belongs to $\mathbb{S}$ or $\overline{\mathbb{S}}$ without $(\widehat{p}, \widehat{q}, \mathscr{L})$, where

$$\begin{cases} \mathbb{S}\text{: } \{x = (\alpha\mathscr{L} + \beta\widehat{p}) \bmod N | \alpha, \beta \in \mathbb{Z}_N, \alpha\mathscr{L} < \widehat{p}\} \\ \overline{\mathbb{S}} = \mathbb{Z}_N/\mathbb{S}\text{: } \{x = (\alpha\mathscr{L} + \beta\widehat{p}) \bmod N | \alpha, \beta \in \mathbb{Z}_N, \alpha\mathscr{L} \geq \widehat{p}\} \end{cases} \tag{3}$$

$(\mathscr{L}, \widehat{p})$-based decision assumption [43]: it is difficult for the probabilistic polynomial time algorithm to solve the $(\mathscr{L}, \widehat{p})$-based decision problem with a nonnegligible advantage in k_0 and k_2.

5. The Proposed Scheme

In this section, a detailed privacy-preserving data aggregation scheme that supports fault tolerant in the smart grid is proposed, consisting of six phases: initialization phase, registration phase, report generation phase, report aggregation phase, report reading phase, and fault tolerant phase. All the notations used in this paper are described in Table 1. The general picture of the proposed scheme is depicted in Figure 3.

5.1. Initialization. In this section, CC would produce the system parameters, TA would generate the blinding factors and the secret value for smart meters.

5.1.1. Control Center

(1) Given the security parameter $(k_0, k_1, k_2, \mathscr{K})$, CC produces an additive cyclic group $\mathbb{G}$ of the prime order q satisfying $|q| = \mathscr{K}$; $\mathbb{G}$ is based on a nonsingular elliptic curve E which is defined over a finite field F_p, satisfying $p > q$. CC chooses the generator P of $\mathbb{G}$.

(2) CC randomly chooses two large prime numbers $\widehat{p}, \widehat{q}$ satisfying $|\widehat{p}| = |\widehat{q}| = k_0$ and computes the public parameter $N = \widehat{p}\widehat{q}$. CC chooses arbitrary $\mathscr{L} \in \{0,1\}^{k_2}$ and computes $E_{ij}(0) = (r_{ij}\mathscr{L} + 0)(1 + r_{ij}'\widehat{p}) \bmod N$ and $E_{ij}(1) = (r_{ij}\mathscr{L} + 1)(1 + r_{ij}'\widehat{p}) \bmod N$ for SM_{ij} $(i = 1, 2, \ldots, \omega; j = 1, 2, \ldots, t)$, where $r_{ij} \in \{0,1\}^{k_2}$ and $r_{ij}' \in \{0,1\}^{k_0}$ are two random numbers. CC secretly transmits the key $K = \langle \widehat{p}, \widehat{q}, \mathscr{L} \rangle$ to TA.

(3) CC chooses five secure hash functions $H\text{: } \{0,1\}^* \longrightarrow \{0,1\}^{k_2}$, $H_1\text{: } \{0,1\}^* \longrightarrow \{0,1\}^l$, $H_i\text{: } \{0,1\}^* \longrightarrow \mathbb{Z}_q^*$ $(i = 2, 3, 4)$, where l is the length of the real identity RID_i.

(4) CC publishes the system parameters $\langle \mathbb{G}, p, q, P, k_0, k_1, k_2, \mathscr{K}, N, H, H_1, H_2, H_3, H_4 \rangle$.

5.1.2. Trusted Authority

(1) TA selects the random number $s_{TA} \in \mathbb{Z}_q^*$ as the master secret key and calculates the corresponding public key $P_{\text{pub}} = s_{TA}P$.

(2) TA selects n arbitrary blinding factors $\theta_{ij} \in \mathbb{Z}_{\mathscr{L}}$ satisfying $\sum_{i=1}^{\omega}\sum_{j=1}^{t}\theta_{ij} = 0 \bmod \mathscr{L}$.

(3) TA chooses a random secret value $k \in \mathbb{Z}_{\mathscr{L}}$.

(4) TA secretly transmits $\langle \theta_{ij}, k \rangle$ to SM_{ij} $(i = 1, 2, \ldots, \omega; j = 1, 2, \ldots, t)$.

5.2. Registration. SM_{ij} and GW would register with CC, respectively, in this section.

5.2.1. Smart Meters' Registration

(1) SM_{ij} randomly chooses $s_{ij}, u_{ij} \in \mathbb{Z}_q^*$ and calculates the public key $S_{ij} = s_{ij}P$ and knowledge signature:

$$\begin{aligned} U_{ij} &= u_{ij}P, \\ v_{ij} &= s_{ij}H_2\left(\text{RID}_{ij}, S_{ij}, U_{ij}\right) + u_{ij}. \end{aligned} \tag{4}$$

Then, SM_{ij} transmits $\langle \text{RID}_{ij}, S_{ij}, U_{ij}, v_{ij} \rangle$ to CC.

(2) After receiving $\langle \text{RID}_{ij}, S_{ij}, U_{ij}, v_{ij} \rangle$, CC verifies whether $v_{ij}P \stackrel{?}{=} S_{ij}H_2(\text{RID}_{ij}, S_{ij}, U_{ij}) + U_{ij}$. If it holds, CC randomly selects $\pi_{ij} \in \mathbb{Z}_q^*$ and calculates the pseudoidentity $ID_{ij} = \{ID_{ij1}, ID_{ij2}\}$ for SM_{ij}, in which $ID_{ij1} = \pi_{ij}P$ and $ID_{ij2} = \text{RID}_{ij} \oplus H_1(\pi_{ij}P_{\text{pub}})$.

(3) CC publishes $\langle ID_{ij}, S_{ij}, U_{ij}, v_{ij} \rangle$ and secretly transmits $\langle E_{ij}(1), E_{ij}(0) \rangle$ to SM_{ij}.

5.2.2. Gateway's Registration

(1) GW randomly selects $s_G, u_G \in \mathbb{Z}_q^*$ and computes the public key $S_G = s_GP$ and knowledge signature

$$\begin{aligned} U_G &= u_GP, \\ v_G &= s_GH_2\left(\text{RID}_G, S_G, U_G\right) + u_G. \end{aligned} \tag{5}$$

Then, GW sends $\langle \text{RID}_G, S_G, U_G, v_G \rangle$ to CC.

(2) CC verifies whether $v_GP \stackrel{?}{=} S_GH_2(\text{RID}_G, S_G, U_G) + U_G$. If it holds, CC publishes $\langle \text{RID}_G, S_G, U_G, v_G \rangle$.

5.3. Report Generation. In this section, SM_{ij} would collect and transmit electricity data to GW.

5.3.1. SM_{ij} Submits the Data Report to SM_{i1}^H

(1) SM_{ij} collects electricity measurement data m_{ij}, randomly selects $r_{ij}'' \in \{0,1\}^{k_2}$, and computes

$$C_{ij} = \left(m_{ij} + H(T,k)\theta_{ij}\right)E_{ij}(1) + r_{ij}''E_{ij}(0) \bmod N, \tag{6}$$

where T is the current timestamp.

(2) SM_{ij} randomly chooses $e_{ij} \in \mathbb{Z}_q^*$ and calculates

$$\begin{aligned} E_{ij} &= e_{ij}P, \\ \sigma_{ij} &= s_{ij}H_3\left(C_{ij}, S_{ij}, ID_{ij}, E_{ij}, T\right) + e_{ij}. \end{aligned} \tag{7}$$

Table 1: Notations.

Notations	Description
TA	The trusted authority
CC	The control center
GW	The gateway
SM_{ij}	The jth smart meter in the ith domain
SM_{i1}^{H}	The domain header in the ith domain
$\mathbb{G}$	The additive cyclic group
P	The generator of $\mathbb{G}$
(s_{TA}, P_{pub})	The private/public key pair of TA
H_i	The secure hash functions
θ_{ij}	The blinding factor of SM_{ij}
n	The number of smart meters
(s_{ij}, S_{ij})	The private/public key pair of SM_{ij}
(s_G, S_G)	The private/public key pair of GW
RID_{ij}	The real identity of SM_{ij}
ID_{ij}	The pseudoidentity of SM_{ij}
K	The key of symmetric homomorphic encryption
T	The current timestamp

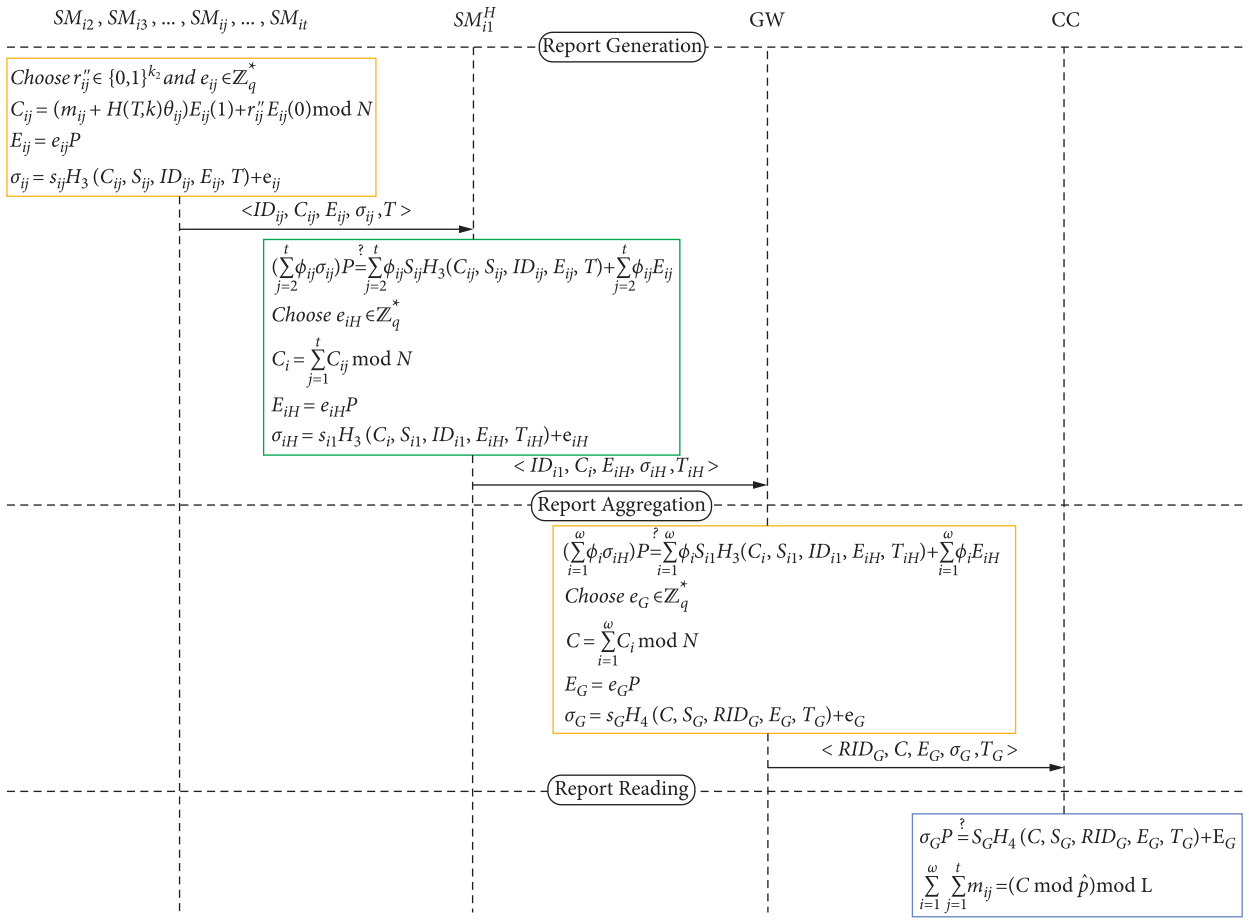

Figure 3: Overview of the proposed scheme.

(3) SM_{ij} submits the data report $\langle ID_{ij}, C_{ij}, E_{ij}, \sigma_{ij}, T\rangle$ to SM_{i1}^{H}.

5.3.2. SM_{i1}^{H} Uploads the Preaggregated Report to GW. (1) Given $\langle ID_{ij}, C_{ij}, E_{ij}, \sigma_{ij}, T\rangle$ from other $t-1 SM_{ij} \in D_i$, SM_{i1}^{H} examines the timestamp T and verifies whether

$$\sigma_{ij} P \stackrel{?}{=} S_{ij} H_3\left(C_{ij}, S_{ij}, ID_{ij}, E_{ij}, T\right) + E_{ij}. \tag{8}$$

In order to speed up the verification, SM_{i1}^{H} uses small exponent test technology [44] to achieve batch verification. SM_{i1}^{H} randomly selects a set of tiny numbers $\phi_{i2}, \phi_{i3}, \ldots, \phi_{it} \in [1, 2^t]$ and verifies whether

$$\left(\sum_{j=2}^{t} \phi_{ij}\sigma_{ij}\right) P \stackrel{?}{=} \sum_{j=2}^{t} \phi_{ij} S_{ij} H_3\left(C_{ij}, S_{ij}, ID_{ij}, E_{ij}, T\right) + \sum_{j=2}^{t} \phi_{ij} E_{ij}. \tag{9}$$

(2) Given $t-1$ data reports and his own data report C_{i1}, SM_{i1}^{H} randomly selects $e_{iH} \in \mathbb{Z}_q^*$ and calculates

$$\begin{aligned} C_i &= \sum_{j=1}^{t} C_{ij} \bmod N, \\ E_{iH} &= e_{iH} P, \\ \sigma_{iH} &= s_{i1} H_3\left(C_i, S_{i1}, ID_{i1}, E_{iH}, T_{iH}\right) + e_{iH}, \end{aligned} \tag{10}$$

where T_{iH} is the current timestamp.

(3) Finally, SM_{i1}^{H} uploads the preaggregated report $\langle ID_{i1}, C_i, E_{iH}, \sigma_{iH}, T_{iH}\rangle$ to GW.

5.4. Report Aggregation. In this section, GW would verify and aggregate the preaggregated reports from SM_{i1}^{H}. Afterward, GW would upload the aggregated report to CC.

(1) Given $\langle ID_{i1}, C_i, E_{iH}, \sigma_{iH}, T_{iH}\rangle$ from the domain headers $SM_{i1}^{H}(i = 1, 2, \ldots, \omega)$, GW examines the timestamp T_{iH}, randomly selects a group of tiny values $\phi_1, \phi_2, \ldots, \phi_\omega \in [1, 2^\omega]$, and verifies whether

$$\left(\sum_{i=1}^{\omega} \phi_i \sigma_{iH}\right) P \stackrel{?}{=} \sum_{i=1}^{\omega} \phi_i S_{i1} H_3\left(C_i, S_{i1}, ID_{i1}, E_{iH}, T_{iH}\right) + \sum_{i=1}^{\omega} \phi_i E_{iH}. \tag{11}$$

(2) GW randomly selects $e_G \in \mathbb{Z}_q^*$ and calculates

$$\begin{aligned} C &= \sum_{i=1}^{\omega} C_i \bmod N, \\ E_G &= e_G P, \\ \sigma_G &= s_G H_4\left(C, S_G, \mathrm{RID}_G, E_G, T_G\right) + e_G. \end{aligned} \tag{12}$$

where T_G is the current timestamp.

(3) Finally, GW uploads the aggregated report $\langle \mathrm{RID}_G, C, E_G, \sigma_G, T_G\rangle$ to CC.

5.5. Report Reading. CC would verify and decrypt the aggregated report from GW in this section.

(1) Receiving $\langle \mathrm{RID}_G, C, E_G, \sigma_G, T_G\rangle$ from GW, CC checks the timestamp T_G and verifies whether

$$\sigma_G P \stackrel{?}{=} S_G H_4\left(C, S_G, \mathrm{RID}_G, E_G, T_G\right) + E_G. \tag{13}$$

(2) CC decrypts aggregated electricity measurement data:

$$\sum_{i=1}^{\omega} \sum_{j=1}^{t} m_{ij} = (C \bmod \widehat{p}) \bmod \mathscr{L}. \tag{14}$$

(3) CC analyzes and processes the aggregated data and makes optimal allocation.

Correctness:

$$
\begin{aligned}
&(C \bmod \widehat{p}) \bmod \mathscr{L}, \\
&= \left[\left(\sum_{i=1}^{\omega}\sum_{j=1}^{t} C_{ij} \bmod N\right) \bmod \widehat{p}\right] \bmod \mathscr{L} \\
&= \left[\left(\sum_{i=1}^{\omega}\sum_{j=1}^{t}\left(\left(m_{ij} + H(T,k)\theta_{ij}\right)E_{ij}(1) + r''_{ij}E_{ij}(0) \bmod N\right) \bmod \widehat{p}\right] \bmod \mathscr{L}\right. \\
&= \left[\left(\sum_{i=1}^{\omega}\sum_{j=1}^{t}\left(m_{ij} + H(T,k)\theta_{ij} + \alpha_{ij}\mathscr{L} + \beta_{ij}\widehat{p}\right) \bmod N\right) \bmod \widehat{p}\right] \bmod \mathscr{L} \\
&= \left[\sum_{i=1}^{\omega}\sum_{j=1}^{t}\left(m_{ij} + H(T,k)\theta_{ij} + \alpha_{ij}\mathscr{L}\right)\right] \bmod \mathscr{L} \\
&= \left[\sum_{i=1}^{\omega}\sum_{j=1}^{t} m_{ij} + H(T,k)\sum_{i=1}^{\omega}\sum_{j=1}^{t}\theta_{ij} + \mathscr{L}\sum_{i=1}^{\omega}\sum_{j=1}^{t}\alpha_{ij}\right] \bmod \mathscr{L} \\
&= \sum_{i=1}^{\omega}\sum_{j=1}^{t} m_{ij},
\end{aligned}
\tag{15}
$$

where $\alpha_{ij} = (m_{ij} + H(T,k)\theta_{ij})r_{ij} + r''_{ij}r_{ij}$ and $\beta_{ij} = r'_{ij}(m_{ij} + H(T,k)\theta_{ij})(1 + r_{ij}\mathscr{L}) + r''_{ij}r'_{ij}r_{ij}\mathscr{L}$.

5.6. Fault Tolerant. This section describes how to obtain the aggregated data when some smart meters fail to work normally.

(1) Assuming that only receiving $\lambda - 1\,(\lambda < t)$ reports, SM_{i1}^{H} performs the same operations as Section 5.3.2 except $C_i = \sum_{j=1}^{\lambda} C_{ij} \bmod N$. Then, SM_{i1}^{H} broadcasts $\langle ID_{i1}, C_i, E_{iH}, \sigma_{iH}, T_{iH}, \lambda\rangle$ into the domain D_i.

(2) Given $\langle ID_{i1}, C_i, E_{iH}, \sigma_{iH}, T_{iH}, \lambda\rangle$, SM_{ij} examines the timestamp T_{iH} and verifies the signature σ_{iH}. If it is valid, SM_{ij} randomly selects $\widehat{e}_{ij} \in \mathbb{Z}_q^*$, $\widehat{r}_{ij} \in \{0,1\}^{k_2}$ and computes

$$
\begin{aligned}
\widehat{C}_{ij} &= C_i - \lambda H(T,k)\theta_{ij}E_{ij}(1) + \widehat{r}_{ij}E_{ij}(0) \bmod N, \\
\widehat{E}_{ij} &= \widehat{e}_{ij}P, \\
\widehat{\sigma}_{ij} &= s_{ij}H_3\left(\widehat{C}_{ij}, S_{ij}, ID_{ij}, \widehat{E}_{ij}, \widehat{T}_{ij}\right) + \widehat{e}_{ij},
\end{aligned}
\tag{16}
$$

where $\widehat{T}_{ij}$ is the current timestamp. Then, SM_{ij} submits the report $\langle ID_{ij}, \widehat{C}_{ij}, \widehat{E}_{ij}, \widehat{\sigma}_{ij}, \widehat{T}_{ij}\rangle$ to SM_{i1}^{H}.

(3) After receiving $\lambda - 1$ data reports $\langle ID_{ij}, \widehat{C}_{ij}, \widehat{E}_{ij}, \widehat{\sigma}_{ij}, \widehat{T}_{ij}\rangle$ from $SM_{ij} \in D_i$, the domain header SM_{i1}^{H} examines $\lambda - 1$ timestamp $\widehat{T}_{ij}$ and verifies $\lambda - 1$ signature $\widehat{\sigma}_{ij}$. If batch verification is valid, SM_{i1}^{H} randomly selects $\widehat{e}_{iH} \in \mathbb{Z}_q^*$ and calculates

$$
\begin{aligned}
\widehat{C}_i &= \frac{1}{\lambda}\sum_{j=1}^{\lambda}\widehat{C}_{ij} \bmod N, \\
\widehat{E}_{iH} &= \widehat{e}_{iH}P, \\
\widehat{\sigma}_{iH} &= s_{i1}H_3\left(\widehat{C}_i, S_{i1}, ID_{i1}, \widehat{E}_{iH}, \widehat{T}_{iH}\right) + \widehat{e}_{iH},
\end{aligned}
\tag{17}
$$

where $\widehat{T}_{iH}$ is the current timestamp. Finally, SM_{i1}^{H} uploads the preaggregated report $\langle ID_{i1}, \widehat{C}_i, \widehat{E}_{iH}, \widehat{\sigma}_{iH}, \widehat{T}_{iH}\rangle$ to GW.

(4) GW and CC normally execute the protocol as Sections 5.4 and 5.5, respectively. Finally, CC gets

aggregated electricity data for no malfunctioning smart meters.

Correctness:

$$
\begin{aligned}
&(C \bmod \widehat{p}) \bmod \mathscr{L}, \\
&= \left[\left(\sum_{i=1}^{\omega} \widehat{C}_i \bmod N\right) \bmod \widehat{p}\right] \bmod \mathscr{L} \\
&= \left[\left(\frac{1}{\lambda}\sum_{i=1}^{\omega}\sum_{j=1}^{\lambda} \widehat{C}_{ij} \bmod N\right) \bmod \widehat{p}\right] \bmod \mathscr{L} \\
&= \left[\left(\frac{1}{\lambda}\sum_{i=1}^{\omega}\sum_{j=1}^{\lambda} \left(C_i - \lambda H(T,k)\theta_{ij}E_{ij}(1) + \widehat{r}_{ij}E_{ij}(0)\right) \bmod N\right) \bmod \widehat{p}\right] \bmod \mathscr{L} \\
&= \left[\left(\sum_{i=1}^{\omega}\sum_{j=1}^{\lambda} \left(C_{ij} - H(T,k)\theta_{ij}E_{ij}(1) + \frac{1}{\lambda}\widehat{r}_{ij}E_{ij}(0)\right) \bmod N\right) \bmod \widehat{p}\right] \bmod \mathscr{L} \\
&= \left[\left(\sum_{i=1}^{\omega}\sum_{j=1}^{\lambda} \left(m_{ij}E_{ij}(1) + r''_{ij}E_{ij}(0) + \frac{1}{\lambda}\widehat{r}_{ij}E_{ij}(0)\right) \bmod N\right) \bmod \widehat{p}\right] \bmod \mathscr{L} \\
&= \left[\left(\sum_{i=1}^{\omega}\sum_{j=1}^{\lambda} \left(m_{ij} + \widehat{\alpha}_{ij}\mathscr{L} + \widehat{\beta}_{ij}\widehat{p}\right) \bmod N\right) \bmod \widehat{p}\right] \bmod \mathscr{L} \\
&= \left[\sum_{i=1}^{\omega}\sum_{j=1}^{\lambda} \left(m_{ij} + \widehat{\alpha}_{ij}\mathscr{L}\right)\right] \bmod \mathscr{L} \\
&= \left[\sum_{i=1}^{\omega}\sum_{j=1}^{\lambda} m_{ij} + \mathscr{L}\sum_{i=1}^{\omega}\sum_{j=1}^{\lambda} \widehat{\alpha}_{ij}\right] \bmod \mathscr{L} \\
&= \sum_{i=1}^{\omega}\sum_{j=1}^{\lambda} m_{ij},
\end{aligned}
\tag{18}
$$

where $\widehat{\alpha}_{ij} = m_{ij}r_{ij}(1 + r'_{ij}\widehat{p}) + r_{ij}(r''_{ij} + 1/\lambda\widehat{r}_{ij})(1 + r'_{ij}\widehat{p})$ and $\widehat{\beta}_{ij} = m_{ij}r'_{ij}$.

6. Security Analysis

6.1. Indistinguishability. The proposed scheme is proved to be the indistinguishability under the chosen plaintext attack (IND-CPA). The adversary $\mathscr{A}$ can execute the below queries:

Hash query: given a hash query, output a random value

Encryption query: given an encryption query on the message m_{ij}, output the ciphertext C_{ij}

The security model can be defined by the interactive game played between the adversary $\mathscr{A}$ and the challenger $\mathscr{C}$.

Setup: $\mathscr{C}$ produces the system parameters and sends them to $\mathscr{A}$.

Phase 1: $\mathscr{A}$ adaptively executes the hash queries and the encryption queries for polynomial times.

Challenge: after completing phase 1, $\mathscr{A}$ randomly selects two messages $m_{ij}^0, m_{ij}^1 \in \mathscr{M}$ and submits two messages to $\mathscr{C}$. Next, $\mathscr{C}$ randomly chooses $b \in \{0,1\}$, calculates the ciphertext C_{ij}^b corresponding to m_{ij}^b, and replies it to $\mathscr{A}$.

Phase 2: $\mathscr{A}$ executes the same queries as Phase 1 apart from the encryption query on the message m_{ij}^0 and m_{ij}^1.

Guess: $\mathscr{A}$ outputs $b' \in \{0,1\}$ as the result of guess.

The advantage for the adversary $\mathscr{A}$ to win the game is defined as

$$\mathrm{Adv}_{\mathscr{A}}^{\mathrm{IND-CPA}} = \left|\Pr[b' = b] - \frac{1}{2}\right|. \tag{19}$$

Definition 1. The proposed scheme ensures IND-CPA secure if the advantage of an adversary in the above game is negligible.

Theorem 1. *The proposed scheme is IND-CPA secure under the $(\mathscr{L}, \widehat{p})$-based decision assumption.*

Proof. Assume that the adversary $\mathscr{A}$ wins the game in Definition 1 with a nonnegligible advantage ε; an algorithm $\mathscr{B}$ would be constituted for breaking the $(\mathscr{L}, \widehat{p})$-based decision problem with advantage ε'. Given an arbitrary bit $z \in \{0, 1\}$, an instance (k_0, k_2, N, x) is established, in which x is arbitrarily selected from $\mathbb{S}$ if $z = 0$ and x is arbitrarily selected from $\overline{\mathbb{S}}$ if $z = 1$. The ultimate target of $\mathscr{B}$ is to guess z.

Setup: $\mathscr{B}$ sets the security parameter (k_0, k_1, k_2) satisfying $k_1 \ll k_2 < (k_0/2)$ and chooses two large prime numbers $\widehat{p}, \widehat{q}$ which satisfy $|\widehat{p}| = |\widehat{q}| = k_0$. $\mathscr{C}$ calculates $N = \widehat{p}\widehat{q}$. Next, $\mathscr{B}$ randomly chooses $\mathscr{L} \in (0, 1)^{k_2}$ and sets up message space $\mathscr{M} = \{m | m \in (0, 1)^{k_1}\}$. $\mathscr{B}$ secretly keeps $(\widehat{p}, \widehat{q}, \mathscr{L})$ and returns (k_0, k_1, k_2, N) to $\mathscr{A}$.

For the purpose of continuous rapid response and consistency, $\mathscr{B}$ holds the below list.

(1) L_H: it consists of tuples (T, k, h_i).

Phase 1: $\mathscr{A}$ executes the following queries adaptively.

Hash H query: $\mathscr{A}$ makes a H query on (T, k) and $\mathscr{B}$ responds according to the following steps:

(1) If (T, k) is included in the list L_H, $\mathscr{B}$ replies the hash value $h_i = H(T, k)$ to $\mathscr{A}$.
(2) If (T, k) is not included in the list L_H, $\mathscr{B}$ randomly selects $h_i \in \{0, 1\}^{k_2}$, inserts (T, k, h_i) into the list L_H, and returns h_i to $\mathscr{A}$.

Encryption query: $\mathscr{A}$ makes an encryption query on the message m_{ij} and $\mathscr{B}$ randomly picks $k, \theta \in \mathbb{Z}_{\mathscr{L}}$ and calculates $C_{ij} = m_{ij} + H(T, k)\theta + x$. Finally, $\mathscr{B}$ returns C_{ij} to $\mathscr{A}$.

Challenge: two messages $m_{ij}^0, m_{ij}^1 \in \mathscr{M}$ are provided by $\mathscr{A}$ which submits them to $\mathscr{B}$. Next, $\mathscr{B}$ randomly picks $k, \theta \in \mathbb{Z}_{\mathscr{L}}$ and a bit $b \in \{0, 1\}$, calculates $C_{ij}^b = m_{ij}^b + H(T, k)\theta + x$, and replies it to $\mathscr{A}$.

Phase 2: the same queries are executed by $\mathscr{A}$ as Phase 1 apart from the encryption query on the messages m_{ij}^0 and m_{ij}^1.

Guess: $\mathscr{A}$ outputs guess b' and submits it to $\mathscr{B}$. If $b' = b$, $\mathscr{B}$ outputs the guess $z = 0$.

When $z = 0$, which means $x \in \mathbb{S}$ and $\alpha\mathscr{L} < \widehat{p}$, the ciphertext $C_{ij}^b = m_{ij}^b + H(T, k)\theta + x = (m_{ij}^b + H(T, k)\theta + \alpha\mathscr{L} + \beta\widehat{p}) \bmod N$ is a valid ciphertext. The probability of $\mathscr{A}$ correctly guessing b is $1/2 + \varepsilon$. Therefore, the probability that $\mathscr{B}$ can successfully guess is $\Pr[\text{Success of } \mathscr{B} | z = 0] = 1/2 + \varepsilon$.

When $z = 1$, which means $x \in \overline{\mathbb{S}}$ and $\alpha\mathscr{L} \geq \widehat{p}$, the ciphertext $C_{ij}^b = m_{ij}^b + H(T, k)\theta + x = (m_{ij}^b + H(T, k)\theta + \alpha\mathscr{L} + \beta\widehat{p}) \bmod N$ is an invalid ciphertext. The probability of $\mathscr{A}$ correctly guessing b is 1/2. Therefore, the probability that $\mathscr{B}$ can successfully guess is $\Pr[\text{Success of } \mathscr{B} | z = 1] = 1/2$.

Based on the above two cases, the probability that $\mathscr{B}$ would break the $(\mathscr{L}, \widehat{p})$-based decision problem is

$$\begin{aligned} \varepsilon' &= \Pr[\text{Success of } \mathscr{B}] \\ &= \Pr[z = 0]\Pr[\text{Success of } \mathscr{B} | z = 0] \\ &\quad + \Pr[z = 1]\Pr[\text{Success of } \mathscr{B} | z = 1] \\ &= \frac{1}{2}\left(\frac{1}{2} + \varepsilon\right) + \frac{1}{2} \cdot \frac{1}{2} \\ &= \frac{1}{2} + \frac{\varepsilon}{2}. \end{aligned} \tag{20}$$

Consequently, $\mathscr{B}$ can break the $(\mathscr{L}, \widehat{p})$-based decision problem with nonnegligible probability, $\varepsilon' = 1/2 + \varepsilon/2$. This generates a conflict with $(\mathscr{L}, \widehat{p})$-based decision assumption; therefore, the proposed scheme is IND-CPA secure.

6.2. Unforgeability. The security of the proposed scheme satisfies the existential unforgeability under the adaptively chosen message attack (EUF-CMA). The adversary $\mathscr{A}$ can execute the following queries:

Hash query: given the hash query, output a random value

Create user query: given a create user query on ID_{ij} of SM_{ij}, output the public key (S_{ij}, U_{ij})

Corrupt user query: given a corrupt user query on ID_{ij} of SM_{ij}, output the private key s_{ij}

Signature query: given a signature query on the ciphertext C_{ij} under ID_{ij} of SM_{ij}, output the signature σ_{ij}

The security model can be defined by the interactive game between the adversary $\mathscr{A}$ and the challenger $\mathscr{C}$.

Initialization: $\mathscr{A}$ chooses a challenging identity ID_{ij}^* and submits it to $\mathscr{C}$.

Setup: $\mathscr{C}$ produces the system parameters and sends them to $\mathscr{A}$.

Query: $\mathscr{A}$ adaptively executes the hash queries, the create user queries, the corrupt queries, and the signature queries for polynomial times except the corrupt user query on ID_{ij}^*.

Forgery: $\mathscr{A}$ produces a forged signature σ_{ij}^* on the ciphertext C_{ij}^* and the challenging identity ID_{ij}^*, such that

(1) σ_{ij}^* is a valid signature
(2) ID_{ij}^* has never been queried in the corrupt user queries

The advantage for the adversary $\mathscr{A}$ to win the game is defined as

$$\mathrm{Adv}_{\mathscr{A}}^{\mathrm{EUF-CMA}} = \Pr[\text{Success of } \mathscr{A}]. \quad (21)$$

Definition 2. The proposed scheme ensures EUF-CMA secure if the advantage of an adversary in the above game is negligible.

Theorem 2. *The proposed scheme is EUF-CMA secure under ECDL assumption.*

Proof. Assume that the adversary $\mathscr{A}$ wins the game in Definition 2 with a nonnegligible advantage ε; an algorithm $\mathscr{B}$ would be constituted for breaking ECDL problem with advantage ε'. An instance $(P, aP = Q)$ of ECDL assumption is established, the ultimate target of $\mathscr{B}$ is to discover $a \in \mathbb{Z}_q^*$.

Initialization: $\mathscr{A}$ selects a challenging identity ID_{ij}^* and submits it to $\mathscr{B}$.

Setup: $\mathscr{B}$ selects security parameter $(k_0, k_1, k_2, \mathscr{K})$ and the cyclic group $\mathbb{G}$. Then, $\mathscr{B}$ randomly selects five hash functions H, H_1, H_2, H_3, H_4 that are regarded as random oracles. Finally, $\mathscr{B}$ sends the system parameters $(\mathbb{G}, p, q, P, k_0, k_1, k_2, \mathscr{K}, N, H, H_1, H_2, H_3, H_4)$ to $\mathscr{A}$.

$\mathscr{B}$ maintains the following three lists:

(1) L_{H_2}: it consists of tuples $(\mathrm{RID}_{ij}, S_{ij}, U_{ij}, h_{2,ij})$

(2) L_{H_3}: it consists of tuples $(C_{ij}, S_{ij}, ID_{ij}, E_{ij}, T, h_{3,ij})$

(3) $L_{SM_{ij}}$: it consists of tuples $(ID_{ij}, s_{ij}, S_{ij}, v_{ij}, U_{ij})$

Query: $\mathscr{A}$ adaptively executes the polynomial times following queries.

Hash H_2 query: $\mathscr{A}$ makes a H_2 query on $(\mathrm{RID}_{ij}, S_{ij}, U_{ij})$ and $\mathscr{B}$ responds according to the following steps:

(1) If $(\mathrm{RID}_{ij}, S_{ij}, U_{ij})$ is included in the list L_{H_2}, $\mathscr{B}$ responds $h_{2,ij} = H_2(\mathrm{RID}_{ij}, S_{ij}, U_{ij})$ to $\mathscr{A}$

(2) If $(\mathrm{RID}_{ij}, S_{ij}, U_{ij})$ is not included in the list L_{H_2}, $\mathscr{B}$ randomly selects $h_{2,ij} \in \mathbb{Z}_q^*$, inserts $(\mathrm{RID}_{ij}, S_{ij}, U_{ij}, h_{2,ij})$ into the list L_{H_2} and responds $h_{2,ij}$ to $\mathscr{A}$

Hash H_3 query: $\mathscr{A}$ executes a H_3 query for $(C_{ij}, S_{ij}, ID_{ij}, E_{ij}, T)$ and $\mathscr{B}$ responds according to the following steps:

(1) If $(C_{ij}, S_{ij}, ID_{ij}, E_{ij}, T)$ is included in the list L_{H_3}, $\mathscr{B}$ responds $h_{3,ij} = H_3(C_{ij}, S_{ij}, ID_{ij}, E_{ij}, T)$ to $\mathscr{A}$

(2) If $(C_{ij}, S_{ij}, ID_{ij}, E_{ij}, T)$ is not included in the list L_{H_3}, $\mathscr{B}$ randomly selects $h_{3,ij} \in \mathbb{Z}_q^*$, inserts $(C_{ij}, S_{ij}, ID_{ij}, E_{ij}, T, h_{3,ij})$ into the list L_{H_3}, and responds $h_{3,ij}$ to $\mathscr{A}$

Create user query: this query is issued by $\mathscr{A}$ on the identity ID_{ij} of SM_{ij} and $\mathscr{B}$ responds according to the following steps:

(1) If $(ID_{ij}, s_{ij}, S_{ij}, v_{ij}, U_{ij})$ is included in the list $L_{SM_{ij}}$, $\mathscr{B}$ responds (S_{ij}, U_{ij}) to $\mathscr{A}$.

(2) If $(ID_{ij}, s_{ij}, S_{ij}, v_{ij}, U_{ij})$ is not included in the list $L_{SM_{ij}}$, $\mathscr{B}$ executes the following steps:

(i) If $ID_{ij} = ID_{ij}^*$, $\mathscr{B}$ randomly chooses $v_{ij}, h_{2,ij} \in \mathbb{Z}_q^*$ and sets $S_{ij} = aP = Q$ and $U_{ij} = v_{ij}P - h_{2,ij}S_{ij}$; if $h_{2,ij}$ already emerges in the list L_{H_2}, $\mathscr{B}$ randomly selects another $v_{ij} \in \mathbb{Z}_q^*$ and tries again. Then, $\mathscr{B}$ inserts $(\mathrm{RID}_{ij}, S_{ij}, U_{ij}, h_{2,ij})$ into the list L_{H_2} and inserts $(ID_{ij}, \perp, S_{ij}, v_{ij}, U_{ij})$ into the list $L_{SM_{ij}}$, respectively. Ultimately, $\mathscr{B}$ responds (S_{ij}, U_{ij}) to $\mathscr{A}$.

(ii) If $ID_{ij} \neq ID_{ij}^*$, $\mathscr{B}$ executes the smart meters' registration algorithm to produce (S_{ij}, U_{ij}) and responds them to $\mathscr{A}$.

Corrupt user query: this query is performed by $\mathscr{A}$ on the identity ID_{ij} of SM_{ij} and $\mathscr{B}$ responds according to the following steps:

(1) If $ID_{ij} = ID_{ij}^*$, $\mathscr{B}$ aborts the game.

(2) If $ID_{ij} \neq ID_{ij}^*$, $\mathscr{B}$ executes the following steps:

(i) If $(ID_{ij}, s_{ij}, S_{ij}, v_{ij}, U_{ij})$ is included in the list $L_{SM_{ij}}$, $\mathscr{B}$ responds (s_{ij}, v_{ij}) to $\mathscr{A}$.

(i) If $(ID_{ij}, s_{ij}, S_{ij}, v_{ij}, U_{ij})$ is not included in the list $L_{SM_{ij}}$, $\mathscr{B}$ executes the create user query on ID_{ij} and responds (s_{ij}, v_{ij}) to $\mathscr{A}$.

Signature query: after receiving a ciphertext C_{ij} and ID_{ij} for a signature query, $\mathscr{B}$ responds according to the following steps:

(1) If $ID_{ij} = ID_{ij}^*$, $\mathscr{B}$ randomly chooses $\sigma_{ij}, T, h_{3,ij} \in \mathbb{Z}_q^*$ and calculates $E_{ij} = \sigma_{ij}P - h_{3,ij}S_{ij}$. If $h_{3,ij}$ already emerges in the list L_{H_3}, $\mathscr{B}$ randomly chooses another $\sigma_{ij} \in \mathbb{Z}_q^*$ and tries again. Afterward, $\mathscr{B}$ inserts $(C_{ij}, S_{ij}, ID_{ij}, E_{ij}, T, h_{3,ij})$ into the list L_{H_3} and responds $(ID_{ij}, C_{ij}, E_{ij}, \sigma_{ij}, T)$ to $\mathscr{A}$.

(2) If $ID_{ij} \neq ID_{ij}^*$, $\mathscr{B}$ executes the report generation algorithm to produce $(ID_{ij}, C_{ij}, E_{ij}, \sigma_{ij}, T)$ and responds them to $\mathscr{A}$.

Forgery: $\mathscr{A}$ produces a forged signature σ_{ij} on the ciphertext C_{ij} under identity ID_{ij} of SM_{ij}, such that

(1) If $ID_{ij} \neq ID_{ij}^*$, $\mathscr{B}$ aborts the game.

(2) If $ID_{ij} = ID_{ij}^*$, $\mathscr{B}$ can produce an additional valid signature σ_{ij}' through different hash value H_3 according to forking lemma [45]. The following two equations can be obtained:

$$\begin{aligned} \sigma_{ij}P &= h_{3,ij}S_{ij} + E_{ij}, \\ \sigma_{ij}'P &= h_{3,ij}'S_{ij} + E_{ij}. \end{aligned} \quad (22)$$

We can calculate

$$\left(\sigma_{ij} - \sigma_{ij}'\right)P = \left(h_{3,ij} - h_{3,ij}'\right)S_{ij} = \left(h_{3,ij} - h_{3,ij}'\right)aP. \quad (23)$$

ECDL problem's solution is obtained by $\mathscr{B}$:

$$a = \left(\sigma_{ij} - \sigma_{ij}'\right)\left(h_{3,ij} - h_{3,ij}'\right)^{-1}. \quad (24)$$

Probability analysis: considering that $\mathscr{A}$ is allowed to execute at most q_{H_2} times H_2 query, q_{H_3} times H_3 query, q_{cre}

times create user query, q_{cor} times corrupt user query, and q_s times signature query. The situation that $\mathscr{B}$ breaks the ECDL problem defined three events as follows:

(1) E_1: $\mathscr{B}$ never aborts the game for the corrupt user queries

(2) E_2: $\mathscr{B}$ can produce a valid signature

(3) E_3: $ID_{ij} = ID_{ij}^*$

According to the above simulation, there are $\Pr[E_1] \geq (1 - q_{H_2}/q)^{q_{cre}} \cdot (1 - 1/q_{H_2})^{q_{cor}} \cdot (1 - q_{H_3}/q)^{q_s}$, $\Pr[E_2|E_1] \geq \varepsilon$, and $\Pr[E_3|E_1 \wedge E_2] \geq 1/q_{H_2}$. Therefore, the probability that $\mathscr{B}$ can solve the ECDL problem is

$$\begin{aligned}\varepsilon' &= \Pr[E_1 \wedge E_2 \wedge E_3] \\ &= \Pr[E_1] \cdot \Pr[E_2|E_1] \cdot \Pr[E_3|E_1 \wedge E_2] \\ &\geq \frac{1}{q_{H_2}} \cdot \left(1 - \frac{q_{H_2}}{q}\right)^{q_{cre}} \cdot \left(1 - \frac{1}{q_{H_2}}\right)^{q_{cor}} \cdot \left(1 - \frac{q_{H_3}}{q}\right)^{q_s} \cdot \varepsilon.\end{aligned} \tag{25}$$

Thus, $\mathscr{B}$ can break the ECDL problem with non-negligible advantage $\varepsilon' \geq 1/q_{H_2} \cdot (1 - q_{H_2}/q)^{q_{cre}} \cdot (1 - 1/q_{H_2})^{q_{cor}} \cdot (1 - q_{H_3}/q)^{q_s} \cdot \varepsilon$. This produces a contradiction with ECDL assumption; consequently, the proposed scheme satisfies unforgeability security.

6.3. Analysis of Security Requirement. The security requirements are analyzed comprehensively in this section.

6.3.1. Confidentiality. On the basis of Theorem 1, the adversary cannot decrypts the ciphertext C_{ij}, C_i, and C to collect electricity data without the key of symmetric homomorphic encryption K. Consequently, confidentiality can be satisfied.

6.3.2. Authentication. Legal smart meter SM_{ij} will register its identity information with CC in advance. After receiving the reports of SM_{ij}, SM_{i1}^H will verify whether $\sigma_{ij}P \stackrel{?}{=} S_{ij}H_3(C_{ij}, S_{ij}, ID_{ij}, E_{ij}, T) + E_{ij}$ holds. Based on Theorem 2, the adversary cannot create a valid authentication without the private key s_{ij}. Obviously, authentication can be met.

6.3.3. Integrity. The ciphertext C_{ij} is signed to generate the signature (σ_{ij}, E_{ij}). On the basis of Theorem 2, the adversary cannot generate the legal signature without the private key s_{ij}, and only valid reports can be accepted. So, this means integrity can be achieved.

6.3.4. Anonymity. Every SM_{ij} is set as a pseudoidentity $ID_{ij} = \{ID_{ij,1}, ID_{ij,2}\}$ in the registration phase corresponding to the real identity RID_{ij}, where $ID_{ij,1} = \pi_{ij}P$ and $ID_{ij,2} = \mathrm{RID}_{ij} \oplus H_1(\pi_{ij}P_{\mathrm{pub}})$. The adversary cannot get real identity RID_{ij} without π_{ij} or s_{TA}. Thus, anonymity is guaranteed in the proposed scheme.

6.3.5. Traceability. When SM_{ij} has malicious behavior, only TA can calculate $\mathrm{RID}_{ij} = ID_{ij,2} \oplus H_1(s_{TA}ID_{ij,1})$ by using private key s_{TA} to uncover the true identity RID_{ij}. In this way, the proposed scheme realizes traceability.

6.3.6. Resistance against Collision Attack. GW can disclose extra ciphertext C_i to CC. Next, CC can obtain $\sum_{j=1}^{t}(m_{ij} + H(T,k)\theta_{ij})$ by calculating $(C_i \mathrm{mod} \widehat{p}) \mathrm{mod} \mathscr{L}$. However, they cannot gain the plaintext $\sum_{j=1}^{t} m_{ij}$ without θ_{ij} and k, even if CC obtains ciphertext C_i by accident. Similarly, CC is still unable to obtain the real electricity data. Hence, the proposed scheme would withstand the collision attack.

6.3.7. Resistance against Modification Attack. According to the guarantee of Theorem 2, any modification of the data report by the polynomial adversary will be detected. Hence, the proposed scheme could resist the modification attack.

6.3.8. Resistance against Replay Attack. Since the reports $\langle ID_{ij}, C_{ij}, E_{ij}, \sigma_{ij}, T\rangle$, $\langle ID_{i1}, C_i, E_{iH}, \sigma_{iH}, T_{iH}\rangle$, and $\langle \mathrm{RID}_G, C, E_G, \sigma_G, T_G\rangle$ contain the timestamp, the receiver could check the freshness of timestamp. Therefore, the replay attacks can be withstood.

6.4. Functionality Comparison. The functionality comparison with the related schemes [19–21, 27, 28] is shown in Table 2. Confidentiality, authentication, integrity, and fault tolerance are denoted by F1, F2, F3, and F4, respectively. In addition, anonymity, traceability, and resistance against collision attack, modification attack, and replay attack are represented by F5, F6, F7, F8, and F9, respectively.

The schemes [19, 20] do not support fault tolerance. Furthermore, the schemes [19, 21, 28] do not protect users' identity privacy, and the schemes [20, 27] cannot trace malicious behaviors. Besides, the schemes [27, 28] may be subjected to collision attacks, and the scheme [27] may be subjected to replay attacks. It is clear that the other related schemes fail to meet several requirements, yet the proposed scheme simultaneously fulfill all the security requirements.

7. Performance Evaluation

In this section, the computation cost and the communication overhead are compared and analyzed in a quantitative way between the related schemes [19–21, 27, 28] and the proposed scheme.

7.1. Computation Cost. In order to ensure fairness comparison, the proposed scheme should be compared with other existing data aggregation schemes [19–21, 27, 28] based on the same 80 bits security level. With respect to the schemes [20, 27] based on Paillier encryption, two large prime numbers u, v are selected as 512 bits, and $N = uv$ is

Table 2: Functionality comparison.

Schemes	F1	F2	F3	F4	F5	F6	F7	F8	F9
Li et al.'s scheme [19]	✓	✓	✓	×	×	✓	✓	✓	✓
Shen et al.'s scheme [20]	✓	✓	✓	×	✓	×	✓	✓	✓
Chen et al.'s scheme [21]	✓	✓	✓	✓	×	✓	✓	✓	✓
Guan et al.'s scheme [27]	✓	✓	✓	✓	✓	×	×	✓	×
Ding et al.'s scheme [28]	✓	✓	✓	✓	×	✓	×	✓	✓
The proposed scheme	✓	✓	✓	✓	✓	✓	✓	✓	✓

1024 bits. Considering the schemes [19–21, 27, 28] based on bilinear pairing, the symmetric bilinear pairing e: $\mathbb{G}_1 \times \mathbb{G}_1 \longrightarrow \mathbb{G}_T$ is exploited, where $\mathbb{G}_1$ is an additive group with generator P of the order q, that is defined on the super singular elliptic curve E: $y^2 = x^3 + x \bmod p$ with embedding degree 2, q is 160-bit Solinas prime number, and p is 512-bit primer number satisfying $q \cdot 12 \cdot r = p + 1$. In terms of the ECC-based schemes [20, 21] and the proposed scheme, an additive group $\mathbb{G}$ with prime order q is established by nonsingular elliptic curve E: $y^2 = x^3 + ax + b \mod p$, in which p, q are both 160 bits prime numbers and $a = -3$ and b is a random 160-bits prime number. With regard to the symmetric homomorphic encryption in this paper, $\widehat{p}$ and $\widehat{q}$ are two 512 bits prime numbers, and the length of $\mathscr{L}$ is 160 bits.

For making more accurate comparison, the running time of each cryptographic operation is estimated by the MIRACL Crypto SDK [46]. The hardware equipment is a PC with 2.90 GHz, whose CPU is i5-10400, memory is 16 GB, and the operating system is 64 bit Windows 10 system. Table 3 indicates the mean consumed time of 10000 executions corresponding to different cryptographic operations.

Considering simplicity, some lightweight operations have been ignored, such as general hash function and point addition. The specific details are described in Table 4, in which n represents the number of smart meters. Assume $t = 10$ in the proposed scheme, that is, $\omega = 0.1n$. The computation cost is divided into three phases, including report generation phase, report aggregation phase, and report reading phase.

First of all, the computation cost of the report generation phase is considered.

Li et al.'s scheme [19] employs $3n$ exponentiation operations in $\mathbb{Z}_N$, $2n$ multiplication operations in $\mathbb{Z}_N$, n exponentiation operations in $\mathbb{G}_1$, and n map-to-point hash operations. As a result, the running time is $3nT_{e-N} + 2nT_{m-N} + nT_{e-\mathbb{G}_1} + nT_{\text{mtp}} = 2.7585n\,ms$.

Shen et al.'s scheme [20] applies n Paillier public key encryption operations, $2n$ exponentiation operations in $\mathbb{Z}_{\overline{q}}$, n scale multiplication operations in ECC, and n map-to-point hash operations. In this way, the running time is $nT_{\text{Enc}-P} + 2nT_{e-\overline{q}} + nT_{m-\text{ECC}} + nT_{\text{mtp}} = 2.6266n\,ms$.

Chen et al.'s scheme [21] utilizes n bilinear pairing operations, $2n$ map-to-point hash operations, n exponentiation operations in $\mathbb{G}_T$, and n scale multiplication operations in ECC. Consequently, the running time is $nT_{bp} + 2nT_{\text{mtp}} + nT_{e-\mathbb{G}_T} + nT_{m-\text{ECC}} = 5.0429n\,ms$.

Guan et al.'s scheme [27] demands n exponentiation operations in $\mathbb{Z}_{N^2}$, n Paillier public key encryption operations, n exponentiation operations in $\mathbb{G}_1$, and n exponentiation operations in $\mathbb{Z}_N$. As a consequence, the running time is $nT_{e-N^2} + nT_{\text{Enc}-P} + nT_{e-\mathbb{G}_1} + nT_{e-N} = 2.237n\,ms$.

Ding et al.'s scheme [28] needs $2n$ exponentiation operations in $\mathbb{G}_T$ and n exponentiation operations in $\mathbb{G}_1$. Hence, the running time is $2nT_{e-\mathbb{G}_T} + nT_{e-\mathbb{G}_1} = 0.9108n\,ms$.

In the report generation phase of the proposed aggregation scheme, SM_{ij} executes n scale multiplication operations in ECC and $3n$ multiplication operations in $\mathbb{Z}_N$. SM_{i1}^H executes $1.2n$ scale multiplication operations in ECC. As a matter of fact, the running time is $2.2nT_{m-\text{ECC}} + 3nT_{m-N} = 0.3198n\,ms$.

Afterward, the computation cost of the report aggregation phase is analyzed.

Li et al.'s scheme [19] employs $n + 1$ bilinear pairing operations, $n + 1$ map-to-point hash operations, n multiplication operations in $\mathbb{Z}_N$, one exponentiation operation in $\mathbb{G}_1$, and one exponentiation operation in $\mathbb{Z}_N$. As a result, the running time is $(n+1)T_{bp} + (n+1)T_{\text{mtp}} + nT_{m-N} + T_{e-\mathbb{G}_1} + T_{e-N} = 3.2532n + 4.0169\,ms$.

Shen et al.'s scheme [20] applies $n + 2$ bilinear pairing operations, $n + 1$ map-to-point hash operations, and one scale multiplication operation in ECC. In this way, the running time is $(n+2)T_{bp} + (n+1)T_{\text{mtp}} + T_{m-\text{ECC}} = 3.2264n + 5.003\,ms$.

Chen et al.'s scheme [21] utilizes $2n$ bilinear pairing operations, $n + 1$ map-to-point hash operations, and one scale multiplication operation in ECC. Consequently, the running time is $2nT_{bp} + (n+1)T_{\text{mtp}} + T_{m-\text{ECC}} = 4.8942n + 1.6674\,ms$.

Guan et al.'s scheme [27] demands $n + 2$ exponentiation operations in $\mathbb{Z}_N$ and n exponentiation operations in $\mathbb{G}_1$. As a consequence, the running time is $(n+2)T_{e-N} + nT_{e-\mathbb{G}_1} = 0.7905n + 0.3558\,ms$.

Ding et al.'s scheme [28] needs $n + 3$ exponentiation operations in $\mathbb{G}_1$. Hence, the running time is $(n+3)T_{e-\mathbb{G}_1} = 0.6126n + 1.8378\,ms$.

In the report aggregation phase, the proposed scheme executes $0.1n + 2$ scale multiplication operations in ECC. As a matter of fact, the running time is $(0.1n+2)T_{m-ECC} = 0.0109n + 0.2176\,ms$.

Table 3: Time cost of cryptographic operations (millisecond).

Notations	Descriptions	Run time
T_{bp}	Bilinear pairing operation	1.6678
T_{mtp}	Map-to-point hash operation	1.5586
$T_{e-\mathbb{G}_T}$	Exponentiation operation in $\mathbb{G}_T$	0.1491
T_{e-N}	Exponentiation operation in $\mathbb{Z}_N$	0.1779
T_{m-N}	Multiplication operation in $\mathbb{Z}_N$	0.0268
$T_{\log}$	Solving the discrete logarithm operation	0.1845
$T_{e-\mathbb{G}_1}$	Exponentiation operation in $\mathbb{G}_1$	0.6126
$T_{e-\overline{q}}$	Exponentiation operation in $\mathbb{Z}_{\overline{q}}$	0.0063
T_{Enc-P}	Paillier public key encryption operation	0.9466
T_{Dec-P}	Paillier public key decryption operation	1.2129
T_{e-N^2}	Exponentiation operation in $\mathbb{Z}_{N^2}$	0.4999
T_{m-ECC}	Scale multiplication operation in ECC	0.1088
$T_{\text{mod}\,\widehat{p}}$	Modular $\widehat{p}$ operation	0.0138
$T_{\text{mod}\,\mathscr{L}}$	Modular $\mathscr{L}$ operation	0.0028

Table 4: Comparison of computation cost.

Schemes	Report generation phase	Report aggregation phase	Report reading phase
Li et al.'s scheme [19]	$3nT_{e-N} + 2nT_{m-N} + nT_{e-\mathbb{G}_1} + nT_{\text{mtp}} = 2.7585n\,ms$	$(n+1)T_{bp} + (n+1)T_{mtp} + nT_{m-N} + T_{e-\mathbb{G}_1} + T_{e-N} = 3.2532n + 4.0169ms$	$2T_{bp} + T_{mtp} + T_{e-N} + T_{\log} = 5.2566ms$
Shen et al.'s scheme [20]	$nT_{\text{Enc}-P} + 2nT_{e-\bar{q}} + nT_{m-\text{ECC}} + nT_{\text{mtp}} = 2.6266n\,ms$	$(n+2)T_{bp} + (n+1)T_{\text{mtp}} + T_{m-\text{ECC}} = 3.2264n + 5.003ms$	$2T_{bp} + T_{\text{mtp}} + T_{\text{Dec}-P} + 2T_{e-\bar{q}} = 6.1197ms$
Chen et al.'s scheme [21]	$nT_{bp} + 2nT_{\text{mtp}} + nT_{e-\mathbb{G}_T} + nT_{m-\text{ECC}} = 5.0429n\,ms$	$2nT_{bp} + (n+1)T_{\text{mtp}} + T_{m-\text{ECC}} = 4.8942n + 1.6674ms$	$3T_{bp} + 2T_{\text{mtp}} + T_{\log} = 8.3051ms$
Guan et al.'s scheme [27]	$nT_{e-N^2} + nT_{\text{Enc}-P} + nT_{e-\mathbb{G}_1} + nT_{e-N} = 2.237n\,ms$	$(n+2)T_{e-N} + nT_{e-\mathbb{G}_1} = 0.7905n + 0.3558ms$	$T_{e-N} + T_{\text{Dec}-P} = 1.3908ms$
Ding et al.'s scheme [28]	$2nT_{e-\mathbb{G}_T} + nT_{e-\mathbb{G}_1} = 0.9108nms$	$(n+3)T_{e-\mathbb{G}_1} = 0.6126n + 1.8378ms$	$3T_{e-\mathbb{G}_1} + T_{bp} + T_{\log} = 3.6901ms$
The proposed scheme	$2.2nT_{m-\text{ECC}} + 3nT_{m-N} = 0.3198n\,ms$	$(0.1n+2)T_{m-\text{ECC}} = 0.0109n + 0.2176\,ms$	$2T_{m-\text{ECC}} + T_{\text{mod}\widehat{p}} + T_{\text{mod}\mathscr{L}} = 0.2342\,ms$

Ultimately, the computation cost of the report reading phase is summarized.

Li et al.'s scheme [19] employs two bilinear pairing operations, one map-to-point hash operation, one exponentiation operation in $\mathbb{Z}_N$, and one solving the discrete logarithm operation. As a result, the running time is $2T_{bp} + T_{\text{mtp}} + T_{e-N} + T_{\log} = 5.2566ms$.

Shen et al.'s scheme [20] applies two bilinear pairing operations, one map-to-point hash operation, one Paillier public key decryption operation, and two exponentiation operations in $\mathbb{Z}_{\overline{q}}$. In this way, the running time is $2T_{bp} + T_{\text{mtp}} + T_{\text{Dec}-P} + 2T_{e-\overline{q}} = 6.1197\,ms$.

Chen et al.'s scheme [21] utilizes three bilinear pairing operations, two map-to-point hash operations, and one solving discrete logarithm operation. Consequently, the running time is $3T_{bp} + 2T_{\text{mtp}} + T_{\log} = 8.3051ms$.

Guan et al.'s scheme [27] demands one exponentiation operation in $\mathbb{Z}_N$ and one Paillier public key decryption operation. As a consequence, the running time is $T_{e-N} + T_{\text{Dec}-P} = 1.3908\,ms$.

Ding et al.'s scheme [28] needs three exponentiation operations in $\mathbb{G}_1$, one bilinear pairing operation, and one solving discrete logarithm operation. Hence, the running time is $3T_{e-\mathbb{G}_1} + T_{bp} + T_{\log} = 3.6901ms$.

In the report reading phase, the proposed scheme executes two scale multiplication operations in ECC, one modular $\widehat{p}$ operation, and one modular $\mathscr{L}$ operation. As a matter of fact, the running time is $2T_{m-\text{ECC}} + T_{\text{mod}\widehat{p}} + T_{\text{mod}\mathscr{L}} = 0.2342\,ms$.

The total running time of the other schemes [19–21, 27, 28] and the proposed scheme are $6.0117n + 9.2735$, $5.853n + 11.1227$, $9.9371n + 9.9725$, $3.0275n + 1.7466$, $1.5234n + 5.5279$, and $0.3307n + 0.4518$, respectively. Figure 4 displays that the overall computation cost varies with the number of smart meters. Apparently, the overall computation cost of all schemes has a linear relationship with the number of smart meters. The proposed scheme requires the minimum overall computation cost and shows slower growth than other schemes. Specifically, the proposed scheme reduced the cost by 94.5%, 94.3%, 96.7%, 89.1%, and 78.3%, respectively, compared with other schemes [19–21, 27, 28]. Consequently, the proposed data aggregation scheme is more appropriate for the smart meters with limited computation resources because it involves no time-consuming operations, such as map-to-point hash and bilinear pairing operation.

7.2. Communication Overhead. The communication overhead will be compared with the schemes [19–21, 27, 28]; the details are shown in Table 5, where $|x|$ denotes bit size of x. In the smart grid, the size of the transmitted report is analyzed, including two parts, communication overhead from SM to GW and from GW to CC. Same as before, the length of $\mathbb{G}$, $\mathbb{G}_1$, $\mathbb{G}_T$, $\mathbb{Z}_{\overline{q}}$, $\mathbb{Z}_N$, and $\mathbb{Z}_{N^2}$ are 160 bits, 512 bits, 1024 bits, 160 bits, 1024 bits, and 2048 bits, respectively. Furthermore, assume that the identity and the timestamp are both defined as 32 bits.

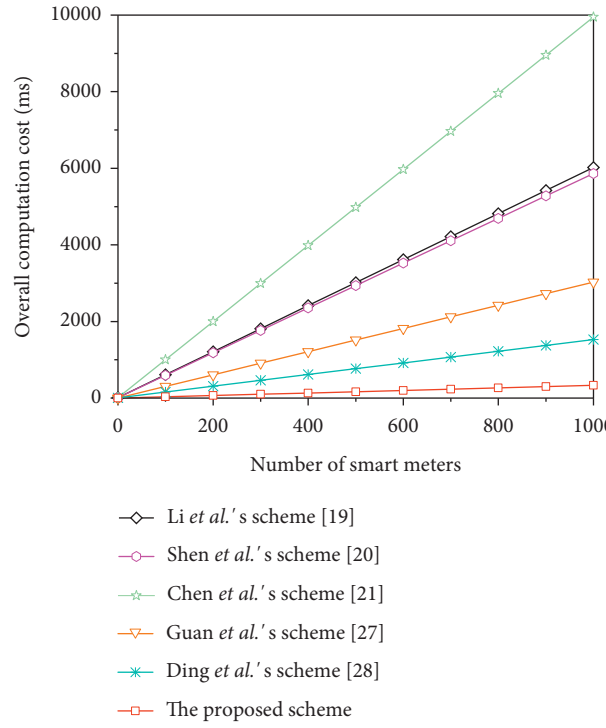

Figure 4: Comparison of overall computation cost.

For the first part, the communication process from SM to GW is analyzed.

In Li et al.'s scheme [19], the electricity transmission data are $\langle ID_{TD_{ij}}, t_{ij}, C_{ij}, \sigma_{ij}\rangle$, where $ID_{TD_{ij}}$ is 32 bits identity and t_{ij} is 32 bits timestamp, $C_{ij} \in \mathbb{Z}_N$ and $\sigma_{ij} \in \mathbb{G}_1$. Consequently, the size of communication overhead is $(|ID_{TD_{ij}}| + |t_{ij}| + |C_{ij}| + |\sigma_{ij}|)\,n = (32 + 32 + 1024 + 512)n = 1600n$ bits.

In Shen et al.'s scheme [20], the electricity transmission data are $\langle \overline{C}_i, g_2^{sk_i}, vk_{U_i}, T, \sigma_i\rangle$, where $\overline{C}_i \in \mathbb{Z}_{N^2}$, $g_2^{sk_i} \in \mathbb{G}_1$, $vk_{U_i} \in \mathbb{G}_1$, T is 32 bits timestamp, and $\sigma_i \in \mathbb{G}$. In this way, the size of communication overhead is $(|\overline{C}_i| + |g_2^{sk_i}| + |vk_{U_i}| + |T| + |\sigma_i|)n = (2048 + 512 + 512 + 32 + 160)n = 3264n$ bits.

In Chen et al.'s scheme [21], the electricity transmission data are $\langle c_i, S_i, t_i, id_i\rangle$, where $c_i \in \mathbb{G}_T$, $S_i \in \mathbb{G}_1$, t_i is 32 bits timestamp, and id_i is 32 bits identity. Therefore, the size of communication overhead is $(|c_i| + |S_i| + |t_i| + |id_i|)n = (1024 + 512 + 32 + 32)n = 1600n$ bits.

In Guan et al.'s scheme [27], the electricity transmission data are $\langle C_i, \delta_i, S_{ID_{U_i}}^{a_i}, l_j(0), G(x_i)\rangle$, where $C_i \in \mathbb{Z}_{N^2}$, $\delta_i \in \mathbb{Z}_N$, $S_{ID_{U_i}}^{a_i} \in \mathbb{G}_1$, $l_j(0)$, and $G(x_i)$ are considered as 32 bits. Thus, the size of communication overhead is $(|C_i| + |\delta_i| + |S_{ID_{U_i}}^{a_i}| + |l_j(0)| + |G(x_i)|)\;\; n = (2048 + 1024 + 512 + 32 + 32)n = 3648n$ bits.

In Ding et al.'s scheme [28], the electricity transmission data are $\langle CT_{i,j}, S_{i,j}, R_i, T_{i,j}, ID_i\rangle$, where $CT_{i,j} = (c_{i,j,1} \in \mathbb{G}_1, c_{i,j,2} \in \mathbb{G}_T)$, $S_{i,j} \in \mathbb{Z}_{\overline{q}}$, $R_i \in \mathbb{G}_1$, $T_{i,j}$ is 32 bits timestamp, and ID_i is 32 bits identity. Hence, the size of communication overhead is $(|CT_{i,j}| + |S_{i,j}| + |R_i| + |T_{i,j}| + |ID_i|)n = (1536 + 160 + 512 + 32 + 32)n = 2272\;n$ bits.

In the proposed scheme, SM_{ij} submits the data report $\langle ID_{ij}, C_{ij}, E_{ij}, \sigma_{ij}, T\rangle$ to SM_{i1}^H, where

Table 5: Comparison of communication overhead.

Schemes	Communication overhead SM-GW	Communication overhead GW-CC
Li et al.'s scheme [19]	$1600n$ bits	1600 bits
Shen et al.'s scheme [20]	$3264n$ bits	3264 bits
Chen et al.'s scheme [21]	$1600n$ bits	1600 bits
Guan et al.'s scheme [27]	$3648n$ bits	3104 bits
Ding et al.'s scheme [28]	$2272n$ bits	2784 bits
The proposed scheme	$1725n$ bits	1408 bits

$ID_{ij} = \{ID_{ij,1} \in \mathbb{G}, ID_{ij,2} \in \{0,1\}^{32}\}$, $C_{ij} \in \mathbb{Z}_N$, $E_{ij} \in \mathbb{G}$, $\sigma_{ij} \in \mathbb{Z}_{\overline{q}}$, T is 32 bits, and the size is $(|ID_{ij}| + |C_{ij}| + |E_{ij}| + |\sigma_{ij}| + |T|)n = (192 + 1024 + 160 + 160 + 32)n = 1568n$ bits. SM_{i1}^H submits the data report $\langle ID_{i1}, C_i, E_{iH}, \sigma_{iH}, T_{iH}\rangle$ to GW, and the size is $(|ID_{i1}| + |C_i| + |E_{iH}| + |\sigma_{iH}| + |T_{iH}|)0.1n = (192 + 1024 + 160 + 160 + 32)0.1n = 157n$ bits. As a matter of fact, the communication overhead is $1568n + 157n = 1725n$ bits.

Figure 5 intuitively reflects the relationship between the communication overhead from SM to GW and the number of smart meters. It is clear that the communication overhead raises linearly with the increase in the number of smart meters. The proposed scheme demands $1725n$ bits, so it reduces the communication overhead by 47.2%, 52.7%, and 24.1%, respectively, compared with the other schemes [20, 27, 28]. Although the related schemes [19, 21] are slightly better than our work in terms of the communication overhead from SM to GW, this is negligible because the proposed scheme would meet the fault tolerance and anonymity that they did not. In general, the proposed scheme consumes less communication resources.

For the second part, the communication process from GW to CC is analyzed.

In Li et al.'s scheme [19], the electricity transmission data are $\langle ID_{ES_i}, t_i, C_i, \sigma_i\rangle$, where ID_{ES_i} is 32 bits identity and t_i is 32 bits timestamp, $C_i \in \mathbb{Z}_N$ and $\sigma_i \in \mathbb{G}_1$. Consequently, the size of communication overhead is $|ID_{ES_i}| + |t_i| + |C_i| + |\sigma_i| = 32 + 32 + 1024 + 512 = 1600$ bits.

In Shen et al.'s scheme [20], the electricity transmission data are $\langle \overline{C}, R, vk_G, T, \sigma\rangle$, where $\overline{C} \in \mathbb{Z}_{N^2}$, $R \in \mathbb{G}_1$, $vk_G \in \mathbb{G}_1$, T is 32 bits timestamp, and $\sigma \in \mathbb{G}$. In this way, the size of communication overhead is $|\overline{C}| + |R| + |vk_G| + |T| + |\sigma| = 2048 + 512 + 512 + 32 + 160 = 3264$ bits.

In Chen et al.'s scheme [21], the electricity transmission data are $\langle c_j, S_j, t_i, id_j\rangle$, where $c_j \in \mathbb{G}_T$, $S_j \in \mathbb{G}_1$, t_i is 32 bits timestamp, and id_j is 32 bits identity. Therefore, the size of communication overhead is $|c_j| + |S_j| + |t_i| + |id_j| = 1024 + 512 + 32 + 32 = 1600$ bits.

In Guan et al.'s scheme [27], the electricity transmission data are $\langle C_a, \sigma_j, TS\rangle$, where $C_a \in \mathbb{Z}_{N^2}$, $\sigma_j \in \mathbb{Z}_N$, and TS is 32 bits timestamp. Thus, the size of communication overhead is $|C_a| + |\sigma_j| + |TS| = 2048 + 1024 + 32 = 3104$ bits.

In Ding et al.'s scheme [28], the electricity transmission data are $\langle CT_{k,j}, S_{k,j}, R_k, U_{k,j}, T_{k,j}, ID_k\rangle$, where $CT_{k,j} = (\widehat{C}_{k,j,1} \in \mathbb{G}_1, \widehat{C}_{k,j,2} \in \mathbb{G}_T)$, $S_{k,j} \in \mathbb{Z}_{\overline{q}}$, $R_k \in \mathbb{G}_1$, $U_{k,j} \in \mathbb{G}_1$, $T_{k,j}$ is 32 bits timestamp, and ID_k is 32 bits identity. Hence, the size of communication overhead is $|CT_{k,j}| + |S_{k,j}| + |R_k| + |U_{k,j}| + |T_{k,j}| + |ID_k| = 1536 + 160 + 512 + 512 + 32 + 32 = 2784$ bits.

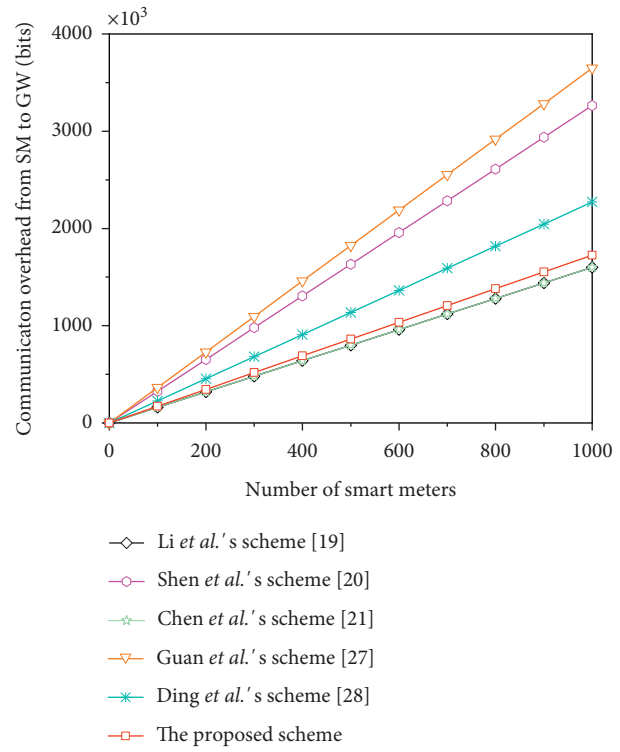

Figure 5: Comparison of communication overhead from SM to GW.

In the proposed scheme, the electricity report is $\langle \mathrm{RID}_G, C, E_G, \sigma_G, T_G\rangle$, where RID_G is 32 bits identity, $C \in \mathbb{Z}_N$, $E_G \in \mathbb{G}$, $\sigma_G \in \mathbb{Z}_{\overline{q}}$, and T_G is 32 bits timestamp. As a matter of fact, the size of communication overhead is $|\mathrm{RID}_G| + |C| + |E_G| + |\sigma_G| + |T_G| = 32 + 1024 + 160 + 160 + 32 = 1408$ bits.

The last column of Table 5 directly illustrates the communication overhead from GW to CC of the related schemes. According to the size of the transmission data report, the proposed scheme utilizes 1408 bits, which is reduced by 12.0%, 56.9%, 12.0%, 54.6%, and 49.4%, respectively, compared with other schemes [19–21, 27, 28]. Consequently, the proposed scheme realizes lower communication overhead from GW to CC, which is more beneficial for GW with limited communication resources.

8. Conclusion

In this paper, we have employed the symmetric homomorphic encryption technology and the elliptic curve signature to design a lightweight and privacy-preserving data aggregation scheme in smart grid. In the proposed scheme, even though the smart meters produce malfunction, the system can still run normally to get aggregated data. Besides, it does not restrict the space of electricity data. The security analysis has demonstrated that the proposed scheme is IND-CPA and EUF-CMA secure and satisfies all security requirements. Ultimately, the performance analysis has reflected the lightweight of the proposed scheme in terms of computation cost and communication overhead. Judging from the results, the proposed scheme is more practical for the smart grid with limited computation and communication capabilities.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (no. 62072054) and Key Research and Development Program of Shaanxi Province (no. 2021GY-047).

References

[1] A. Cheshmehzangi and H. Chen, "Key suggestions and steps ahead for China's carbon neutrality plan," in *China's Sustainability Transitions*, pp. 189–206, Springer, China, 2021.

[2] F. Rahimi and A. Ipakchi, "Demand response as a market resource under the smart grid paradigm," *IEEE Transactions on Smart Grid*, vol. 1, no. 1, pp. 82–88, 2010.

[3] 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.

[4] 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 2012 IEEE INFOCOM*, pp. 1674–1682, IEEE, Orlando, FL, USA, Mar 2012.

[5] R. Deng, Z. Yang, M.-Y. Chow, and J. Chen, "A survey on demand response in smart grids: mathematical models and approaches," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 3, pp. 570–582, 2015.

[6] R. Deng, G. Xiao, R. Lu, and J. Chen, "Fast distributed demand response with spatially and temporally coupled constraints in smart grid," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 6, pp. 1597–1606, 2015.

[7] 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.

[8] 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.

[9] D. Niyato, L. Ping Wang, 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.

[10] C. Greer, D. Wollman, D. Prochaska et al., *NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 3.0*, NIST, U.S, 2014.

[11] Y. Yan, Y. Qian, H. Sharif, and D. Tipper, "A survey on smart grid communication infrastructures: motivations, requirements and challenges," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 5–20, 2012.

[12] A. Anzalchi and A. Sarwat, "A survey on security assessment of metering infrastructure in smart grid systems," in *Proceedings of the SoutheastCon*, pp. 1–4, IEEE, Fort Lauderdale, FL, USA, Apr 2015.

[13] Y. Yan, Y. Qian, H. Sharif, and D. Tipper, "A survey on cyber security for smart grid communications," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 4, pp. 998–1010, 2012.

[14] C.-I. Fan, S.-Y. Huang, and Y.-L. Lai, "Privacy-enhanced data aggregation scheme against internal attackers in smart grid," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 1, pp. 666–675, 2014.

[15] H. Bao and R. Lu, "Comment on "privacy-enhanced data aggregation scheme against internal attackers in smart grid"," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 1, pp. 2–5, 2015.

[16] D. He, S. Zeadally, H. Wang, and Q. Liu, "Lightweight data aggregation scheme against internal attackers in smart grid using elliptic curve cryptography," *Wireless Communications and Mobile Computing*, vol. 2017, 2017.

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

[18] Y. Zhang, J. Zhao, D. Zheng et al., "Privacy-preserving data aggregation against false data injection attacks in fog computing," *Sensors*, vol. 18, no. 8, p. 2659, 2018.

[19] X. Li, S. Liu, F. Wu, S. Kumari, and J. J. P. C. Rodrigues, "Privacy preserving data aggregation scheme for mobile edge computing assisted IoT applications," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4755–4763, 2019.

[20] H. Shen, Y. Liu, Z. Xia, and M. Zhang, "An efficient aggregation scheme resisting on malicious data mining attacks for smart grid," *Information Sciences*, vol. 526, pp. 289–300, 2020.

[21] Y. Chen, J. Martínez-Ortega, P. Castillejo, and L. López, "An elliptic curve-based scalable data aggregation scheme for smart grid," *IEEE Systems Journal*, vol. 14, no. 2, pp. 2066–2077, 2019.

[22] L. Chen, R. Lu, and Z. Cao, "PDAFT: a privacy-preserving data aggregation scheme with fault tolerance for smart grid communications," *Peer-to-Peer Networking and Applications*, vol. 8, no. 6, pp. 1122–1132, 2015.

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

[24] B. Pan, P. Zeng, and K.-K. R. Choo, "A new multidimensional and fault-tolerable data aggregation scheme for privacy-preserving smart grid communications," in *Proceedings of the International Conference on Applications and Techniques in Cyber Security and Intelligence*, pp. 206–219, Springer, Berlin,Germany, Oct 2017.

[25] S. Ge, P. Zeng, R. Lu, and K.-K. R. Choo, "FGDA: fine-grained data analysis in privacy-preserving smart grid communications," *Peer-to-Peer Networking and Applications*, vol. 11, no. 5, pp. 966–978, 2018.

[26] K. Xue, Q. Yang, S. Li et al., "PPSO: a privacy-preserving service outsourcing scheme for real-time pricing demand response in smart grid," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2486–2496, 2018.

[27] Z. Guan, Y. Zhang, L. Zhu, L. Wu, and S. Yu, "EFFECT: an efficient flexible privacy-preserving data aggregation scheme with authentication in smart grid," *Science China Information Sciences*, vol. 62, no. 3, p. 32103, 2019.

[28] Y. Ding, B. Wang, Y. Wang, K. Zhang, and H. Wang, "Secure metering data aggregation with batch verification in industrial smart grid," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 10, pp. 6607–6616, 2020.

[29] X. Wang, Y. Liu, and K. R. Choo, "Fault-tolerant multisubset aggregation scheme for smart grid," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 6, pp. 4065–4072, 2020.

[30] X. Liu, Y. Zhang, B. Wang, and H. Wang, "An anonymous data aggregation scheme for smart grid systems," *Security and Communication Networks*, vol. 7, no. 3, pp. 602–610, 2014.

[31] M. Badra and S. Zeadally, "Design and performance analysis of a virtual ring architecture for smart grid privacy," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 2, pp. 321–329, 2014.

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

[33] Y. Gong, Y. Cai, Y. Guo, and Y. Fang, "A privacy-preserving scheme for incentive-based demand response in the smart grid," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1304–1313, 2015.

[34] D. He, N. Kumar, and J.-H. Lee, "Privacy-preserving data aggregation scheme against internal attackers in smart grids," *Wireless Networks*, vol. 22, no. 2, pp. 491–502, 2016.

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

[36] X. Shen, L. Zhu, C. Xu, K. Sharif, and R. Lu, "A privacy-preserving data aggregation scheme for dynamic groups in fog computing," *Information Sciences*, vol. 514, pp. 118–130, 2020.

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

[38] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Proceedings of the International Conference on The Theory and Applications of Cryptographic Techniques*, pp. 223–238, Springer, Berlin,Germany, Apr 1999.

[39] D. Boneh, E.-J. Goh, and K. Nissim, "Evaluating 2-DNF formulas on ciphertexts," in *Proceedings of the Theory of Cryptography Conference*, pp. 325–341, Springer, Berlin,-Germany, Apr 2005.

[40] V. Miller, "Use of elliptic curves in cryptography," in *Proceedings of the Conference on The Theory and Application of Cryptographic Techniques*, pp. 417–426, Springer, Berlin,-Germany, Dec 1985.

[41] A. J. Menezes, T. Okamoto, and S. A. Vanstone, "Reducing elliptic curve logarithms to logarithms in a finite field," *IEEE Transactions on Information Theory*, vol. 39, no. 5, pp. 1639–1646, 1993.

[42] H. Mahdikhani, R. Lu, Y. Zheng, J. Shao, and A. A. Ghorbani, "Achieving O ($\log^3$n) communication-efficient privacy-preserving range query in fog-based IoT," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 5220–5232, 2020.

[43] Y. Zheng, R. Lu, Y. Guan, J. Shao, and H. Zhu, "Efficient and privacy-preserving similarity range query over encrypted time series data," *IEEE Transactions on Dependable and Secure Computing*, 2021.

[44] J. K. Liu, T. H. Yuen, M. H. Au, and W. Susilo, "Improvements on an authentication scheme for vehicular sensor networks," *Expert Systems with Applications*, vol. 41, no. 5, pp. 2559–2564, 2014.

[45] D. Pointcheval and J. Stern, "Security proofs for signature schemes," in *Proceedings of the International Conference on The Theory and Applications of Cryptographic Techniques*, pp. 387–398, Springer, Saragossa, Spain, May 1996.

[46] Shamus Software, "Multiprecision integer and rational arithmetic cryptographic library (MIRACL)," 2021, http://www.certivox.com/miracl/.