symmetry

Article

Privacy Preserving Data Aggregation for Smart Grid with User Anonymity and Designated Recipients

Liang Wu [1], Wenzheng Zhang [2] and Wei Zhao [2,*]

[1] School of Computer Science and Artificial Intelligence, Wuhan University of Technology, Wuhan 430070, China; liangwu_73@whut.edu.cn
[2] Science and Technology on Communication Security Laboratory, Chengdu 610000, China; zwz85169038@sina.com
* Correspondence: zhaowei9801@163.com

Abstract: Smart grids integrate modern Internet of Things technologies with the traditional grid systems, aiming to achieve effective and reliable electricity distribution as well as promote clean energy development. Nowadays, it is an indispensable infrastructure for smart homes, wisdom medical, intelligent transportation, and various other services. However, when smart meters transmit users' power consumption data to the control center, sensitive information may be leaked or tampered. Moreover, distributed architecture, fine-grained access control, and user anonymity are also desirable in real-world applications. In this paper, we propose a privacy-preserving data aggregation scheme for a smart grid with user anonymity and designated recipients. Smart meters collect users' power consumption data, encrypt it using homomorphic re-encryption, and then transmit the ciphertexts anonymously. Afterward, proxies re-encrypt the aggregated data in a distributed fashion so that only the designated recipients can decrypt it. Therefore, our proposed scheme provides a more secure and flexible solution for privacy-preserving data aggregation in smart grids. Security analyses prove that our scheme achieves all the above-mentioned security requirements, and efficiency analyses demonstrate that it is efficient and suitable for real-world applications.

Keywords: smart grid; user anonymity; designated recipients; homomorphic re-encryption

Citation: Wu, L.; Zhang, W.; Zhao, W. Privacy Preserving Data Aggregation for Smart Grid with User Anonymity and Designated Recipients. *Symmetry* **2022**, *14*, 847. https://doi.org/10.3390/sym14050847

Academic Editors: Zhe Xia, Yanwei Zhou and Yining Liu

Received: 25 March 2022
Accepted: 11 April 2022
Published: 19 April 2022

1. Introduction

Electricity is important for modern civilization. However, power outages occur from time to time across the world, causing significant economic losses and social impacts. For example, in 2019, the Guri Hydropower Station, which provides 80% of Venezuela's electricity, was maliciously attacked, causing 21 out of the 23 states to experience power outages [1]. At the same year, a large-scale power outage also occurred in South America, affecting more than 40 million people in Argentina, Brazil, Uruguay, and Chile. When such an accident occurs, the traffic lights ceased operation and all public transportation was suspended, making the affected cities into chaos [2]. One of the main reasons for this catastrophe is that the traditional power grid was designed more than a century ago and its effectiveness and robustness are far from satisfactory in the modern era [3].

In 2001, the concept of "smart grid" was introduced, expecting to enhance the traditional power grid using some latest information technologies, such as the Internet of Things (IoT) and computer networks [4]. In the smart grid, power transmission can be scheduled more intelligent and reliable thanks to the digitization and standardization of information [5]. Moreover, through two-way communications, the power grid can be continuously monitored in real-time, reducing the probability of power outages. To alleviate the phenomenon of isolated data islands, various cryptographic primitives, such as symmetric and asymmetric ciphers, have been employed to realize privacy-preserving and authentication in data sharing. Therefore, not only the desirable security requirements can be guaranteed for users' personal data, but also the whole society can benefit from more effective information utilization. Nowadays, many countries have adopted the development of smart grids as a national strategy.

As shown in Figure 1, the smart grid generally consists of three layers [6]. At the bottom layer, smart meters collect users' power consumption data and upload it into the grid system regularly. Based on this data, the electricity company can charge the users for power usage. This information also can be used to set-up flexible price packages to smooth the power usage, e.g., higher prices in the peak period and lower prices in the other periods, enhancing the reliability and effectiveness of the smart grid system. At the middle layer, the cloud is responsible for forwarding the aggregated power consumption data to the power station. During this process, individual users' power consumption data must be kept secret from the cloud. Otherwise, users' living habits, as well as some other private information, may be leaked. Moreover, if a malicious attacker tampers or forges the power consumption data during transmission, it will not only cause economic losses to the electricity company but also affect the power distribution of the entire grid. At the top level, the power station generates electricity based on demands and the power is distributed through the substations. As it is expensive to store electricity, it requires that the amount of electricity generated by the power station roughly matches the real-time demands. Otherwise, it will reduce the reliability of the smart grid or even cause catastrophic events.

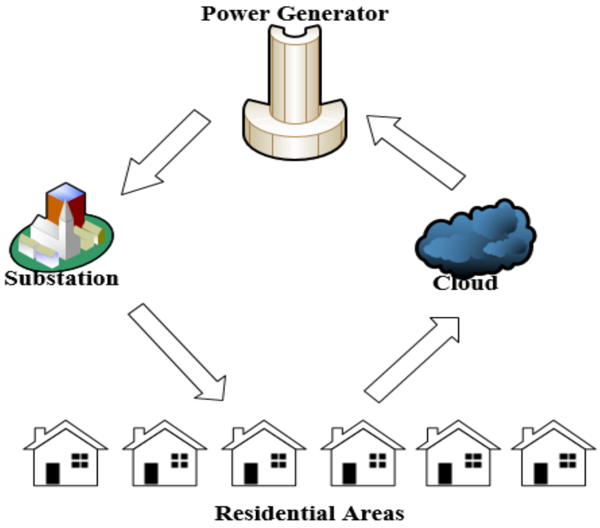

Figure 1. The traditional architecture of the smart grid.

At present, the design and implementation of a smart grid need to consider the following security features:

- *Confidentiality:* The data collected by the smart meters may contain users' sensitive information. If an attacker obtains this data, users' living habits could be leaked, so the power consumption data must be protected.
- *Authentication:* Power consumption data transmitted in the smart grid can be tampered with by a malicious adversary, so it is necessary to ensure that the adversary cannot modify, fabricate or delete the transmitted data without being detected.
- *User anonymity:* The power consumption data is normally sent with the user's identity. When the cloud collects the data, users' identities may be exposed to the cloud. In many circumstances, such exposure is undesirable and users' identities should also be protected.
- *No single point of trust:* The decryption power should not be possessed by a single party. Otherwise, it could become a single point of trust in the system. For example, if

this party is compromised, all sensitive information within the system can be read or leaked by this party. Instead, a distributed architecture should be employed.

- *Designated recipients:* Based on the minimum disclosure principle, fine-grained access control should be posed on the aggregated power consumption data, e.g., its access should be strictly restricted to the designated recipients.

To address the above problems, this paper proposes a privacy-preserving data aggregation scheme with user anonymity and designated recipients. In our proposed scheme, smart meters first collect users' power consumption data, encrypt it using homomorphic re-encryption and then send the ciphertexts anonymously. The control centers aggregate the received data and re-encrypt it in a distributed fashion so that only the designated recipients can decrypt it. Moreover, novel verification techniques are employed to ensure that only legitimate users' data is accepted and an adversary cannot tamper with this data during transmission. Our main contributions can be summarized as follows.

1. Apart from the traditional security requirements, such as confidentiality and authentication, our proposed scheme also achieves user anonymity and no single point of trust. Moreover, it can ensure that the aggregated data can only be accessed by the designated recipients, realizing fine-grained access control. Therefore, it provides a more secure and flexible solution for privacy-preserving data aggregation in smart grid.
2. Security analyses prove that our scheme achieves all these desirable security requirements, and efficiency analyses demonstrate that it is efficient to be implemented in real-world applications.

The rest of the paper is organized as follows. In Section 2, we briefly review some related works in the literature. The notations and preliminaries are outlined in Section 3. In Section 4, models and definitions are described. Then, our proposed scheme is introduced in Section 5, and its security and efficiency analyses are presented in Sections 6 and 7, respectively. Finally, we conclude in Section 8.

2. Related Works

Nowadays, it is widely accepted that smart grids are a fundamental infrastructure for renewable energy [7]. Smart meters are important devices to realize two-way communication in smart grids, so they are vulnerable targets for attackers [8,9]. Hence, it is worth investigating methods that securely transmit information within smart grids and build a flexible smart grid architecture [10,11]. It is necessary to build a security model to meet the security demands of a smart grid [12,13]. To address this issue, various privacy-preserving data aggregation schemes have been proposed in the literature [14–16]. Moreover, these works can be divided into two main categories: one protects users' power consumption data and the other protects users' identities.

In the first category, homomorphic encryption [17] is used as a popular building block, thanks to its feature of allowing operations on the ciphertexts. Lu et al. [18] have proposed a data aggregation scheme EPPA that uses hyper-increasing sequences to record the multi-dimensional data and Paillier encryption to encrypt the data. The local gateway aggregates the encrypted data and sends it to the control center, which can then decrypt the aggregated data without learning any individual data. Later, Shen et al. [19] have proposed a modified data aggregation scheme in which the aggregated data of different regions can be aggregated in a hierarchical manner. Ding et al. [20] have proposed a novel encryption scheme that supports homomorphic re-encryption, in which the ciphertexts can be either decrypted or re-encrypted, both requiring two parties to operate in a distributed fashion. However, the majority of existing data aggregation solutions need to employ a trusted third party (TTP) [21–24]. To address this issue, Liu et al. [25] have proposed a scheme without a TTP. The trick is to select some users to construct a virtual aggregation area to mask the power consumption data of a particular user. Xue et al. [26] have proposed another data aggregation scheme without a TTP using secret sharing. However, it suffers heavy communication overheads and it is vulnerable to the man-in-the-middle attack.

To improve efficiency of data sharing in smart grids, Zhao et al. [27] have introduced a fog-assisted data aggregation scheme that can reduce network bandwidth and realize smart pricing. Su et al. [28] proposed a lightweight data aggregation scheme for smart grid with forwarding secrecy. However, its limitation is that if any user's data is missing, the aggregated data will become unreadable. To solve this issue, Huang et al. [29] have proposed a lightweight data aggregation scheme with fault tolerance. Xu et al. [30] have proposed a similar scheme that allows collusion between the aggregator and some entities, achieving a high level of fault tolerance. Although the above-mentioned schemes can achieve privacy protection for individual users' power consumption data, very few have considered fine-grained access control for the aggregated data.

In the second category, smart meters have to send the power consumption data anonymously. A pseudonym is a common technique used to achieve user anonymity. Tan et al. [31] suggest using pseudo IDs instead of real identities, where these pseudo IDs are generated using a function with inputs of the group key, the time, and the number of smart meters. To hide the relationship between a user's identity and her pseudonym, Guan et al. [32] suggested using the user's public key as her pseudonym. Each user can be associated with many pseudonyms, and the Bloom filter is used to verify the validity of a user's pseudonym. Liu et al. [33] have proposed a solution using a blind signature, but it has not considered the protection of individual user's power consumption data. Sui et al. [34] have proposed a method to realize strong anonymity through anonymous networks. Moreover, a reward mechanism is designed where the user who requests a reduction in power usage can revoke her anonymity and gets some rewards. Yu et al. [35] have proposed a privacy-preserving power request scheme. Each smart meter is associated with a unique identifier, and a ring signature is used to protect their identities. Cheung et al. [36] have proposed a scheme achieve user privacy and data authentication, in which users generate a group of credentials and the control center signs them blindly. However, as the control center needs to generate many signatures, its computational overheads are very high.

3. Notations and Preliminaries

In this section, we describe the notations and briefly review some cryptographic primitives, such as ElGamal encryption, Schnorr signature, and Homomorphic re-encryption.

3.1. Notations

The notations used in the proposed scheme and their meanings are outlined in Table 1.

Table 1. Some notations.

Notations	Meaning
p, q	Two large primes such that $q \| p-1$
G	A finite cyclic group with order q
g	The generator of group G
Z_p^*	A multiplicative group modulo p
H	A collision resistant hash function
PK	The public key
m_i	User's power consumption data
M_i	The power consumption data recorded by RMM_i
M	The total amount of power usage across all areas
ℓ, S	The number of SM_i in each area and RMM_i
RID	The smart meter's real identity
T^*	The current time stamp of the RMM_i
ΔT	The allowed time delay in the system
CID	Computation identifier
$L(*)$	Bit length of the input data
$\|\|$	The message concatenation operation

3.2. Preliminaries

3.2.1. ElGamal Encryption

- *Setup:* Randomly choose $x \in \mathbb{Z}_q$ and compute $y \equiv g^x \pmod{p}$. The public key is (y, p, g) and the private key is x.
- *Encryption:* Given the plaintext m, randomly choose a value $r \in \mathbb{Z}_q$ and calculate the ciphertext as $C = (C_1, C_2) = (g^r, m \cdot y^r)$.
- *Decryption:* The entity with the private key x can decrypt the ciphertext as:

$$m = \frac{C_2}{C_1^x} = \frac{m \cdot y^r}{g^{rx}} = \frac{m \cdot y^r}{y^r} \pmod{p}$$

3.2.2. Schnorr Signature

- *KeyGen:* Randomly choose $x \in \mathbb{Z}_q$ and compute $y \equiv g^x \pmod{p}$. The public key is (y, p, g) and the private key is x.
- *Signing:* Given the message m, the signer randomly selects $k \in \mathbb{Z}_q$ and computes $r \equiv g^k \pmod{p}$, $e = H(r, m)$ and $s = xe + k \pmod{q}$. Now, (e, s) is the signature for m.
- *Verifying:* After receiving the signature (e, s), the verifier computes $r' \equiv g^s y^{-e} \pmod{p}$ and $H(r', m)$. Then the following equation is verified:

$$Ver(y, (e, s), m) = True \Leftrightarrow H(r', m) = e$$

3.2.3. Homomorphic Re-Encryption

The ciphertext can be either decrypted or re-encrypted, while both operations need two entities to collaborate. The homomorphic property permits users to perform computations on the encrypted data. The computation results are left in encrypted form. But when decrypted, the value is identical as the operations are performed on the plaintext data. The re-encryption property allows the ciphertext to be re-encrypted to another one containing the same plaintext but under a different public key. Note that the re-encryption operation ensures that only the designated recipients can derive the plaintext. Its operation works as follows:

- *Setup:* p' and q' are two safe primes, where $p' = 2p'' + 1, q' = 2q'' + 1$ and $n = p' \cdot q'$. Denote QR as the cyclic group of quadratic residues in $\mathbb{Z}_{n^2}^*$, and g is a generator of QR.
- *KeyGen:* The data center (DC) and the access control server (ACS) generate their public and private key pairs $(SK_{DC} = a, PK_{DC} = g^a)$ and $(SK_{ACS} = b, PK_{ACS} = g^b)$. These two parties execute the Diffie-Hellman key exchange to obtain the system public key $PK = PK_{DC}^{SK_{ACS}} = PK_{ACS}^{SK_{DC}} = g^{ab}$. Every designated recipient generates its public and private key pair $\left(sk_i = k_i, pk_i = g^{k_i}\right)$.
- *Encryption:* Given a message $m_i \in \mathbb{Z}_n$, one randomly chooses $r \in [1, \frac{n}{4})$ and generates the ciphertext as

$$\begin{aligned} [m_i]_{PK} &= (T_i, T_i') \\ &= \{(1 + m_i \cdot n) \cdot PK^r, g^r\} \pmod{n^2} \end{aligned}$$

- *Re-Encryption Phase I:* DC chooses and publishes a computation identifier CID. It then computes $h_1 = H\left(\left(pk_j\right)^{SK_{DC}} || CID\right)$ and re-encrypt the ciphertext as

$$[m_i]_+ = \left(\hat{T}_i, \hat{T}_i'\right) = \left\{T_i, (T_i')^{SK_{DC}} \cdot g^{h_1}\right\}$$

- *Re-Encryption Phase II:* ACS calculates $h_2 = H\left(\left(pk_j\right)^{SK_{ACS}}||CID\right)$ after receiving $[m_i]_+$. It then re-encrypts the ciphertext as

$$[m_i]_{pk_j} = (\bar{T}_i, \bar{T}_i') = \left\{\hat{T}_i, \left(\hat{T}_i'\right)^{SK_{ACS}} \cdot g^{h_2}\right\}$$

- *Decryption:* The designated recipient can decrypt the ciphertext $[m_i]_{pk_j}$ as

$$\begin{aligned} {h_1}' &= H\left((PK_{DC})^{sk_j}||CID\right) \\ {h_2}' &= H\left((PK_{ACS})^{sk_j}||CID\right) \\ m_i &= L\left(\frac{\bar{T}_i \cdot PK_{ACS}^{h_1'} \cdot g^{h_2'}}{\bar{T}_i'} \pmod{n^2}\right) \end{aligned}$$

where $L(u) = \frac{u-1}{n}$.

4. Models and Definitions

In this section, we describe the system model, adversary model, and security requirements.

4.1. System Model

Our proposed system, as shown in Figure 2, consists of five types of participants: smart meters (SM), regional master meters (RMM), grid company (GC), operation center (OC), and power transmission units (PTU).

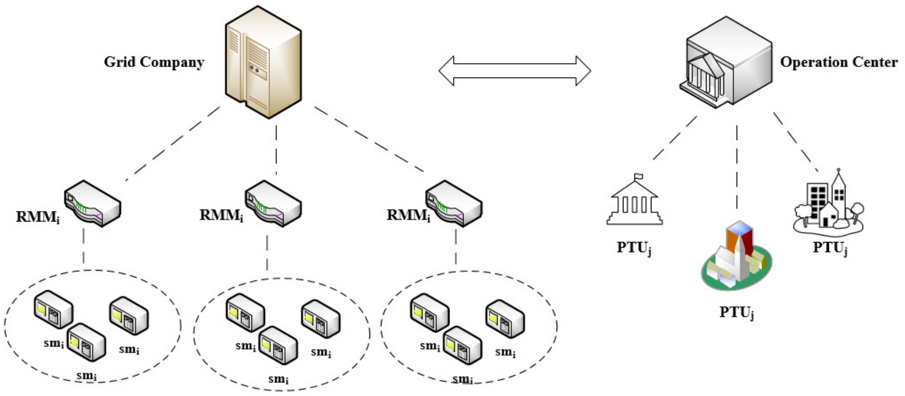

Figure 2. System model in our scheme.

1. SM: It collects user power consumption data and sends it to the RMM regularly. Note that this data needs to be sent anonymously in our proposed scheme. Moreover, each SM is assumed to contain some tamper-proof device, and its internal states can be protected.
2. RMM: It is responsible for aggregating users' power consumption data in some regions and it will forward the aggregated result to the GC.
3. GC: Once it receives the aggregated power consumption data from the RMMs, it aggregates the received data again and then performs the first phase of proxy re-encryption.
4. OC: It executes the second phase of proxy re-encryption and sends the outputs to the designated recipients.
5. PTU: They are the designated recipients of power usage data, such as power plants and data analysts. Each of them will use its private key to decrypt the received ciphertexts.

4.2. Communication Model

We assume that public channels are used to transmit data from SMs to RMMs and from RMMs to the GC. Moreover, we assume that a secure channel exists between the GC and the OC, and authenticated channels are used to transmit data from the OC to PTUs. In smart grids, there might be a large number of SMs and RMMs. Hence, it is impractical to assume that secure channels or authenticated channels exist among them. Moreover, as there is only one GC, one OC, and a few PTUs, the assumption of a secure channel between the GC and the OC, and some authenticated channels from the OC to PTUs is feasible. Note that the assumption of these channels allows us to focus on the protocol design without digging into the low level of technical details. It is well known how these channels can be implemented in practice using standard cryptographic primitives, e.g., encryption and digital signatures.

4.3. Adversary Model

In our proposed scheme, we assume that all participants are honest-but-curious. In other words, these participants will follow the protocol, but they will try to learn some sensitive information beyond their authorization. Moreover, we assume that the GC and the OC will not collude. The adversary $\mathcal{A}$ can eavesdrop on the exchanged messages through the public channel and the authenticated channel. In addition, it can also tamper with the data through the public channel but it neither intercepts nor falsifies the data through the secure channel.

4.4. Security Requirements

Under the above models, our design goal is to develop a privacy-preserving data aggregation scheme for smart grids with user anonymity and designated recipients. Specifically, the following security requirements are considered.

1. *Correctness*: If all participants follow the protocol, it will output the correct aggregated power consumption data to the designated recipients.
2. *Confidentiality*: The adversary $\mathcal{A}$ cannot learn the power consumption data of any individual user.
3. *Authentication*: Only data from legitimate participants will be accepted. If the data is tampered with during transmission, it can be detected.
4. *User anonymity and un-linkability*: The adversary $\mathcal{A}$ cannot extract the real identities of the smart meters. Moreover, $\mathcal{A}$ cannot link two messages that are sent by the same smart meter.
5. *No single point of trust*: The secret key is distributed among multiple entities, i.e., no single party can decrypt or leak sensitive information within the smart grid.
6. *Designated recipients*: The aggregated power consumption data can only be accessed by the designated recipients but no one else.

5. The Proposed Scheme

In this section, the privacy-preserving data aggregation scheme with user anonymity and designated recipients is introduced, which mainly consists of the following six algorithms: initialization, key generation, identity anonymization and encryption, verification and aggregation, proxy re-encryption, and decryption.

5.1. Initialisation

In this phase, GC generates the system parameters. It first randomly chooses two large primes p' and q', and then computes $n = p' \cdot q'$. Denote G as a cyclic group of quadratic residues modulo n^2, and g as a generator of G. GC also selects a secure hash function H: $\{0,1\}^* \rightarrow G$.

5.2. KeyGen

All entities generate their own public and private key pairs. In addition, GC and OC jointly negotiate a key using Diffie–Hellman key exchange.

1. GC and OC randomly chooses α and β respectively as its private key. Their public and private key pairs are $(SK_{GC} = \alpha, PK_{GC} = g^{\alpha})$ and $(SK_{OC} = \beta, PK_{OC} = g^{\beta})$.
2. Each power transmission unit PTU_j generates its public and private key pair $(sk_j = d_j, pk_j = g^{d_j})$.
3. OC negotiates the key with GC to obtain the system public key:

$$PK = PK_{OC}^{SK_{GC}} = PK_{GC}^{SK_{OC}} = g^{\alpha\beta} (\mathrm{mod}\, n^2) \tag{1}$$

4. Finally, the system parameters $pp = (g, G, PK)$ are made public.

5.3. Identity Anonymization and Encryption

In this phase, the smart meter encrypts its real identity and then sends the anonymous identity, power consumption data, and digital signature to the RMM. The following steps are executed during this phase.

1. Before smart meter SM_i sending the power consumption data m_i to RMM_i, SM_i needs to encrypt the data m_i and hide its real identity. And SM_i generates its public and private key pair $(sk_i = x_i, pk_i = g^{x_i})$.
2. In each period, SM_i randomly chooses $\eta_i \in Z_q$ and calculates $HID_{i,1} = g^{\eta_i} \pmod p$, $HID_{i,2} = RID * (PK_{GC})^{\eta_i}$. Then SM_i uses the public key PK to encrypt data and sign, $C_{m_i} = Enc_{PK}(m_i) = \{(1 + m_i \cdot n) \cdot PK^r, g^r\} \pmod{n^2}, \sigma_i = \eta_i + x_i \cdot H(HID_i || T_i || C_{m_i})$, where T_i is the current timestamp and $HID_i = \{HID_{i,1}, HID_{i,2}\}$. Then SM_i sends the message $\{C_{m_i}, \sigma_i, HID_i, T_i\}$ to RMM_i.

5.4. Batch Verification and Aggregation

In this phase, RMM_i checks the validity of received messages. In addition to the traditional verification methods, it also allows a batch of data to be verified simultaneously.

1. *Traditional verification:*
   (a) Once the message $\{C_{m_i}, \sigma_i, HID_i, T_i\}$ from SM_i is received, RMM_i checks the validity of T_i first. If $T^* - T_i > \Delta T$, RMM_i will reject the message.
   (b) RMM_i checks the validity of σ_i using the following equation:

$$g^{\sigma_i} = HID_{i,1} \cdot pk_i^{H(HID_i || T_i || C_{m_i})} \tag{2}$$

2. *Batch Verification:* The above verification can be made more efficient using the small exponent test technology [37].
   (a) Upon receiving multiple data $\{C_{m_1}, \sigma_1, HID_1, T_1\}$, $\{C_{m_2}, \sigma_2, HID_2, T_2\}, \ldots,$ $\{C_{m_\ell}, \sigma_\ell, HID_\ell, T_\ell\}$ sent by some SM_i, RMM_i checks the freshness of T_i, where $i = 1, 2, \ldots, \ell$. When the check fails, RMM_i rejects the message.
   (b) RMM_i selects a random vector $v = v_1, v_2, \ldots, v_\ell$, where v_i is a small random integer in $[1, 2^t]$ and t is a small integer. Then, RMM_i verifies through the following equation:

$$\sum_{i=1}^{\ell} (g^{\sigma_i})^{v_i} = \sum_{i=1}^{\ell} (HID_{i,1} \cdot pk_i^{H(HID_i || T_i || C_{m_i})})^{v_i} \tag{3}$$

   If the above equation does not hold, RMM_i rejects the messages.

3. *Aggregation:* RMM_i aggregates the encrypted data C_{m_i} by calculating $C_{M_i} = \prod_{i=1}^{\ell} C_{m_i}$, where ℓ is the number of SM_i in the current area. Finally, RMM_i sends C_{M_i} and its corresponding signature and current timestamp T_j to GC.

$$\begin{aligned} C_{M_i} &= \{C_1, C_2\} \\ &= \{(PK^r)^{\ell}(1 + n \cdot \sum_{i=1}^{\ell} m_i), (g^r)^{\ell}\} \\ &\quad (\bmod\, n^2) \end{aligned} \tag{4}$$

5.5. Proxy Re-Encryption

After receiving the message, GC first verifies the freshness and validity of the signature. It then aggregates and stores the received power consumption data. When a designated recipient requests electricity data, proxy re-encryption is performed.

1. GC verifies the freshness and correctness of the received data C_{M_i} and it then aggregates them:

$$\begin{aligned} C_M &= \{A, B\} = \prod_{i=1}^{S} C_{M_i} \\ &= \{(PK^{r\ell})^S(1 + n \cdot \sum_{i=1}^{S} M_i), (g^{r\ell})^S\} \\ &\quad (\bmod\, n^2) \end{aligned} \tag{5}$$

2. The PTU_j issues a request to the electricity data. After verifying that it is a legitimate designated recipient, the proxy re-encryption will be performed as follows:
   (a) GC calculates $h_1 = H((pk_j)^{SK_{GC}}||CID)$. Then it converts C_M to C'_M and send it to OC, where $C'_M = \{A', B'\} = \{A, g^{h_1} \cdot B^{SK_{GC}}\}$.
   (b) OC calculates $h_2 = H((pk_j)^{SK_{OC}}||CID)$. Then computes C''_M and sends it to PTU_j, where $C''_M = \{A'', B''\} = \{A', g^{h_2} \cdot (B')^{SK_{GC}}\}$.

5.6. Decryption

Once the PTU_j has received the C''_M from OC, it can be decrypted using its private key.

1. PTU_j first calculates

$$h'_1 = H((PK_{GC})^{sk_j}||CID) = H(g^{\alpha \cdot sk_j}||CID) = h_1$$

$$h'_2 = H((PK_{OC})^{sk_j}||CID) = H(g^{\beta \cdot sk_j}||CID) = h_2$$

2. The aggregated electricity data M can be decrypted as follows:

$$M = L(\frac{A'' \cdot PK_{OC}^{h'_1} \cdot g^{h'_2}}{B''}(\bmod\, n^2)) \tag{6}$$

3. Once PTU_j obtains the aggregated power consumption data M, it can perform dynamic power distribution according to the power consumption across the area.

6. Security Analyses

In this section, we analyze the security properties of the proposed scheme, proving that it meets the aforementioned security requirements.

6.1. Correctness

Theorem 1. *If the data sent by the SM_i were not tampered by the adversary $\mathcal{A}$, the RMM_i would accept it.*

Proof 1. Once the RMM_i receives the message $\{C_{m_i}, \sigma_i, HID_i, T_i\}$ from the SM_i, it can verify its authenticity using the following equation. Therefore, if the data sent by the SM_i was not tampered by the adversary $\mathcal{A}$, the RMM_i will accept it.

$$\begin{aligned} g^{\sigma_i} &= g^{\eta_i + x_i \cdot H(HID_i||T_i||C_{m_i})} \\ &= g^{\eta_i} \cdot g^{x_i \cdot H(HID_i||T_i||C_{m_i})} \\ &= HID_{i,1} \cdot pk_i^{H(HID_i||T_i||C_{m_i})} \end{aligned} \tag{7}$$

□

Theorem 2. *Given multiple messages and their corresponding valid signatures $\{\sigma_i\}_{1 \le i \le n}$ from different smart meters, the batch verification technique (3) can be used to verify their authenticity simultaneously.*

Proof 2. The correctness of Equation (3) can be proved as follows:

$$\begin{aligned} \sum_{i=1}^{\ell} (g^{\sigma_i})^{v_i} &= \sum_{i=1}^{\ell} g^{v_i(\eta_i + x_i \cdot H(HID_i||T_i||C_{m_i}))} \\ &= \sum_{i=1}^{\ell} (g^{v_i \cdot \eta_i} \cdot g^{v_i \cdot x_i \cdot H(HID_i||T_i||C_{m_i})}) \\ &= \sum_{i=1}^{\ell} (HID_{i,1} \cdot pk_i^{H(HID_i||T_i||C_{m_i})})^{v_i} \end{aligned} \tag{8}$$

□

Theorem 3. *The designated recipients can decrypt the received message with their own private key to obtain the correct electricity data.*

Proof 3. The correctness of Equation (6) can be proved as follows:

$$\begin{aligned} M &= L(\frac{A'' \cdot PK_{OC}^{h_1'} \cdot g^{h_2'}}{B'} (\bmod n^2)) \\ &= L(\frac{(PK^{r\ell})^S (1 + n \cdot \sum_{i=1}^{S} M_i) \cdot (g^{\beta})^{h_1{}'} \cdot g^{h_2{}'}}{g^{h_2} \cdot (g^{h_1} \cdot (g^{r\ell S})^{SK_{GC}})^{SK_{OC}}} (\bmod n^2)) \\ &= L((1 + n \cdot \sum_{i=1}^{S} M_i)(\bmod n^2)) \\ &= \sum_{i=1}^{S} M_i \end{aligned} \tag{9}$$

□

6.2. User Anonymity and Un-Linkability

Theorem 4. *Our proposed scheme achieves user anonymity and un-linkability, i.e., the adversary $\mathcal{A}$ with probability polynomial-time resources cannot link the identity sent by the same smart meter.*

Proof 4. When a SM_i sends its power consumption data to RMM_i, it firstly hides its identity RID to achieve anonymous transmission HID_i, where $HID_i = \{HID_{i,1}, HID_{i,2}\} = \{g^{\eta_i}, RID \cdot (PK_{GC})^{\eta_i}\}$. As the ElGamal encryption is semantic secure, i.e., the adversary cannot learn any plaintext information from the given ciphertext. Hence, $\mathcal{A}$ cannot learn the real identity of the smart meter. Moreover, the ElGamal ciphertext, which encodes the pseudo-identity, can be re-encrypted. The re-encryption can be performed multiple times, and it does not require the knowledge of the private key. After re-encryption, the ciphertext appears random and it cannot be linked to its previous form because ElGamal encryption is semantic secure. Therefore, if this pseudo-identity is refreshed regularly, $\mathcal{A}$ cannot link the identity to the same smart meter. □

6.3. Confidentiality

In our scheme, all transmissions are encrypted, so the adversary $\mathcal{A}$ cannot eavesdrop on the smart meter to get electricity data.

Theorem 5. *If the semantic security of the encryption scheme [38] holds, our proposed scheme satisfies confidentiality against malicious GC or OC.*

Proof 5. Assume that there is a probabilistic polynomial-time adversary $\mathcal{A}$ that can break the confidentiality of our proposed scheme. Our goal is to use $\mathcal{A}$ to construct an algorithm $\mathcal{S}$ to break the semantic security of the encryption scheme in [38]. $\mathcal{S}$ is given the public parameters $(n, g, pk_2 = g^a (\bmod n^2))$, the adversary $\mathcal{A}$ can construct $pk_1 = g^b$. Then the adversary $\mathcal{A}$ choose two messages of the same length m_0 and m_1, we randomly select $\beta \leftarrow \{0, 1\}$ and encrypt m_β as follow: $Enc(m_\beta) = \{(1 + m_\beta \cdot n)(pk_2)^r, g^r\} \bmod n^2$. The The encrypted ciphertext is sent to the adversary $\mathcal{A}$. Adversary $\mathcal{A}$ performs further calculations as follows:

1. $A = \{(1 + m_\beta \cdot n)(pk_2)^r\}^b \bmod n^2 = \{(g^{ab})^r (1 + b \cdot m_\beta \cdot n)\} \bmod n^2, B = g^r$;
2. Based on (A, B), $\mathcal{A}$ further construct a re-encryption ciphertext (A', B'), where $A' = A \cdot g^{h_1} = \{(g^{ab})^r (1 + b \cdot m_\beta \cdot n) \cdot g^{h_1}\} \bmod n^2, h_1 = H((pk_1)^a || CID), B' = B$.

We can observe that adversary $\mathcal{A}$ can obtain two raw data $m_0{}' = b \cdot m_0$ and $m_1{}' = b \cdot m_1$. We set $m_\beta{}' = b \cdot m_\beta$. We can observe that (A', B') is one HRES ciphertext. It has already been proved that if the encryption scheme is semantically secure, then the HRES scheme is also semantically secure. Because the HRES is semantically secure, adversary $\mathcal{A}$ cannot guess the value of β'. Hence, our proposed scheme satisfies confidentiality. □

6.4. No Single Point of Trust

The secret key of the system is shared by the GC and OC, and it is assumed that these two participants will not collude. Hence, neither of them can obtain the sensitive information within the system that is encrypted under the corresponding public key.

6.5. Designated Recipients

In the decryption phase, only the designated recipients can decrypt the ciphertext outputted by OC. The designated recipients have the private key sk_j to compute $h_1' = H((PK_{GC})^{sk_j} || CID)$ and $h_2' = H((PK_{OC})^{sk_j} || CID)$. Hence, designated recipients can obtain the aggregated power consumption data M. Although A'' and B'' are transmitted over the communication network, and the adversary $\mathcal{A}$ is assumed to be able to intercept this information, $\mathcal{A}$ cannot decrypt C_M'' because it cannot calculate $h_1' = h_1$ and $h_2' = h_2$. Therefore, only the designated recipients can obtain the computational results, but no one else.

6.6. Comparison of Security Properties

Our proposed scheme is compared with several related schemes, such as [20,33–35]. The following table presents the comparison results. As shown in Table 2, our scheme is the only one that can satisfy all of the desirable security properties, such as user anonymity, un-linkability, confidentiality, correctness, and designated recipients.

Table 2. Comparison with Existing Related Schemes.

Schemes	Ding [20]	Liu [33]	Sui [34]	Yu [35]	Ours
Anonymity	N	Y	Y	Y	Y
Un-linkability	N	N	N	N	Y
Confidentiality	Y	N	N	N	Y
Correctness	Y	Y	Y	Y	Y
Designated recipients	Y	N	N	N	Y

7. Efficiency Analyses

In this section, we evaluate the performance of our proposed scheme in terms of computation and communication.

7.1. Computation Costs

The following notations are used to denote different operations in our scheme. Let C_e, C_m and C_H denote one exponentiation operation, one multiplication operation, and a hash function, respectively. The bilinear pairing C_p incurs the most computation costs. The other operations are much faster, such as the hash operation and the addition operation. ℓ is the number of SM_i in each area. S is the number of the regional master meters. In Table 3, the computation cost of all entities are listed, where "-", GW, OA and CC denote non-considered, gateway, trusted operation organization and control center, respectively.

Table 3. Comparison of Computation Costs.

Schemes	Our Scheme	EPPA [18]	Guan [21]	Shen [39]
SM	$2C_e + C_m$	$lC_e + C_m + 4C_p$	$4C_e + 3C_m$	$2C_e + C_m$
RMM/GW	$lC_e + lC_m$	$wC_p + C_m$	$3C_e + 2lSC_m$	$n_iC_p + (n - n_i)C_e + C_m$
GC	$5C_e + 2C_m$	-	-	$(S + 2)C_p + C_m$
OC	$3C_e + C_m$	-	-	-
PTU/OA/CC	$4C_e + 3C_m$	$2C_p + C_e + 4C_m$	$3C_e + 2C_m$	$2C_p$

When smart meter SM_i generates power consumption data $\{C_{m_i}, \sigma_i, HID_i, T_i\}$, the computational costs of user anonymity are considered negligibly. Then, 2 exponentiation operations and a multiplication operation are required to encrypt electricity data, and a hash operation are required to generate σ_i. Thus, the computation costs of a smart meter is $2C_e + C_m + C_H$. After receiving the power consumption data from ℓ smart meters, the RMM first verifies the received data by performing a batch verification, including ℓ exponentiation operations, ℓ multiplication operations, and ℓ hash operations. In addition, the RMM should aggregate the data from different SM_i and encrypt the data, in which the computation costs are $2C_e + C_m$. As follows, the GC aggregates the data from different RMM, which costs $2C_e + C_m$.

When a designated recipient requests electricity data from the OC, OC forwards the request to the GC. It costs 3 exponentiation, a multiplication, and a hash operation. Then OC also needs to perform 3 exponentiation, a multiplication and a hash operation. After the designated recipient receives the data, it needs to spend $4C_e + 3C_m + 2C_H$ to perform the decryption operation. As hash function can be computed much faster than the other computations, we will ignore the computational costs of hash function evaluation.

7.2. Communication Costs

The communication overheads of our proposed scheme can be divided into two parts, power consumption data transmission, and electricity data request. In Figure 3, we compare the communication overheads of our scheme with some related schemes, such as EPPA [18], Shen's scheme [39], and Jo's scheme [40].

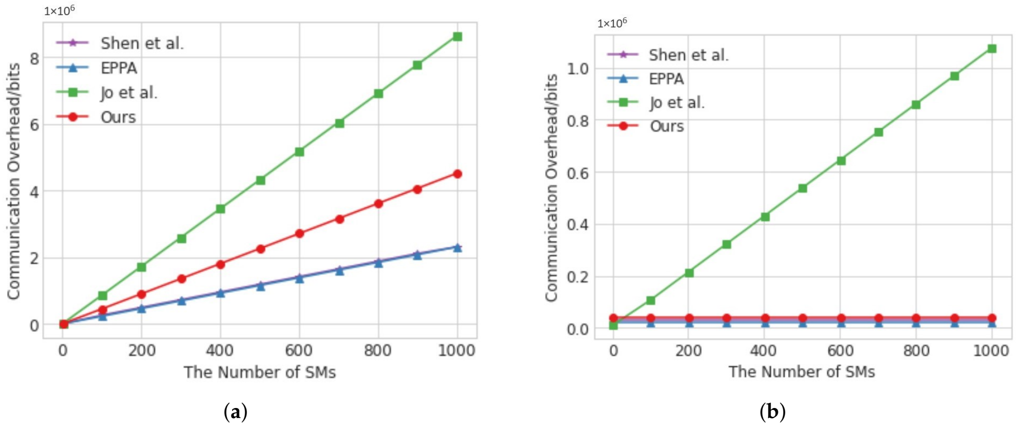

Figure 3. Comparison of the communication overhead. (**a**) SM-to-RMM. (**b**) OC-to-PTU.

We first consider power consumption data transmission phase, where smart meters transmit the power consumption data to the RMM. The data is in the form of $\{C_{m_i}, \sigma_i, HID_i, T_i\}$. Thus, the size of power consumption data is $S_s = |C_{m_i}| + |\sigma_i| + |HID_i| + |T_i|$. The group element in G is of 160 bits and Z_p^* contains elements of 160 bits. Each ciphertext is composed of two parts, we have $4L(n) = 4096$ bits if we choose 1024-bit n. When we set $|T_i|$ = 100-bit length, the communication overheads of SM_i-to-RMM are S_s = 4516 bits. Then the communication overheads of RMM-to-GC is $S_R = |C_{m_i}| + |\sigma| + |T_j|$ = 4356 bits.

Next, we consider the electricity data request phase. Electricity data sent by GC to OC is in ciphertext, so the size of the communication overheads are $S_G = 4096$ bits. OC still sends PTU_j encrypted data after re-encrypting. Thus the communication overheads are also $S_O = 4096$ bits.

In Figure 3a,b, we plot the communication overheads versus the number of smart meters. We set the number of smart meters from 1 to 1000 and increased it by an interval of 100. As shown in Figure 3a,b, the communication overheads in the grid increase linearly with the number of smart meters. In Figure 4, we present a graph of the relationship between the number of regions and the communication overheads. In our scheme, the encryption mode with long ciphertext length is used, so the communication overheads of our scheme are about twice compared with the scheme proposed by Shen et al. [39]. However, the increase in the number of regions does not affect the communication overheads sent by the OC to the designated recipients.

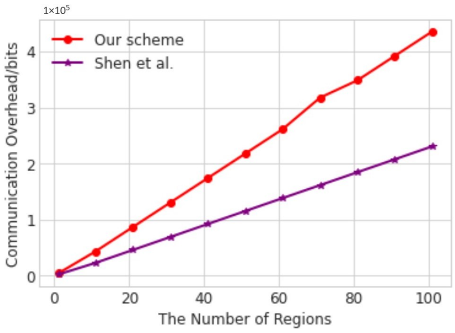

Figure 4. Communication overhead incurred by regions.

8. Conclusions

In this paper, we have proposed a privacy-preserving data aggregation scheme for smart grids with user anonymity and designated recipients. The smart meters collect users' power consumption data but this data is encrypted using homomorphic re-encryption so that the adversary cannot intercept it and only the designated recipients can obtain the aggregated results. Moreover, users' identities are protected and there is no single point of trust. Therefore, it provides a more secure and flexible solution for privacy-preserving data aggregation in the smart grid. Performance analysis demonstrates that it is generally as efficient as the existing related schemes, achieving more desirable security features.

In future work, we would like to investigate further how to remove the assumption that all participants are honest-but-curious, and introduce novel verification techniques to ensure those dishonest participants can be detected and identified. Moreover, the security proof for the authentication property suffers a loose security reduction because security arguments for the Schnorr signature require to use the Forking Lemma. In the future, we would like to explore efficient authentication techniques with a tight security reduction.

Author Contributions: Conceptualization, L.W., W.Z. (Wenzheng Zhang), and W.Z. (Wei Zhao); methodology, L.W.; software, L.W.; validation, W.Z. (Wenzheng Zhang); formal analysis, W.Z. (Wei Zhao); investigation, W.Z. (Wei Zhao); resources, W.Z. (Wenzheng Zhang); data curation, L.W.; writing—original draft preparation, L.W.; writing—review and editing, W.Z. (Wenzheng Zhang) and W.Z. (Wei Zhao); visualization, L.W.; supervision, W.Z. (Wenzheng Zhang); project administration, W.Z. (Wenzheng Zhang); funding acquisition, W.Z. (Wenzheng Zhang) and W.Z. (Wei Zhao). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Science and Technology on Communication Security Lab (Grant No. 6142103010710).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fang, L.; Huang, L.; Zhao, Q. Discussion on megalopolis power grid safety from the perspective of Venezuelan blackout. *Power Energy* **2019**, *40*, 674–677.
2. Gao, K.; Han, F.; Dong, P.; Xiong, N.; Du, R. Connected vehicle as a mobile sensor for real time queue length at signalized intersections. *Sensors* **2019**, *19*, 2059. [CrossRef] [PubMed]
3. Arnold, G.W. Challenges and Opportunities in Smart Grid: A Position Article. *Proc. IEEE.* **2011**, *99*, 922–927. [CrossRef]
4. Farhangi, H. The path of the smart grid. *IEEE Power Energy Mag.* **2010**, *8*, 18–28. [CrossRef]
5. Liu, H. A Review on Development Practice of Smart Grid Technology in China. *IOP Conf. Ser. Mater. Sci. Eng.* **2017**, *199*, 012062. [CrossRef]
6. Northcotegreen, J. *Control and Automation of Electrical Power Distribution Systems*; CRC Press: Boca Raton, FL, USA, 2007.
7. Sheha, M.; Mohammadi, K.; Powell, K. Solving the duck curve in a smart grid environment using a non-cooperative game theory and dynamic pricing profiles. *Energy Convers. Manag.* **2020**, *220*, 113102. [CrossRef]
8. Shen, H.; Liu, Y.; Xia, Z.; Zhang, M. An efficient aggregation scheme resisting on malicious data mining attacks for smart grid. *Inf. Sci.* **2020**, *526*, 289–300. [CrossRef]
9. Yang, P.; Xiong, N.; Ren, J. Data security and privacy protection for cloud storage: A survey. *IEEE Access* **2020**, *8*, 131723–131740. [CrossRef]
10. Aloqaily, M.; Boukerche, A.; Bouachir, O.; Khalid, F.; Jangsher, S. An energy trade framework using smart contracts: Overview and challenges. *IEEE Netw.* **2020**, *34*, 119–125. [CrossRef]
11. Lopez, J.; Rubio, J.E.; Alcaraz, C. A resilient architecture for the smart grid. *IEEE Trans. Ind. Inform.* **2018**, *14*, 3745–3753. [CrossRef]
12. Serrano, D.; Ruíz, J.F.; Muñoz, A.; Maña, A.; Armenteros, A.; Crespo, B.G.N. Development of applications based on security patterns. In Proceedings of the 2009 Second International Conference on Dependability, Athens, Greece, 18–23 June 2009; pp. 111–116.
13. Sánchez-Cid, F.; Mana, A.; Spanoudakis, G.; Kloukinas, C.; Serrano, D.; Munoz, A. Representation of security and dependability solutions. In *Security and Dependability for Ambient Intelligence*; Springer: Boston, MA, USA, 2009; pp. 69–95.
14. Li, S.; Xue, K.; Yang, Q.; Hong, P. PPMA: Privacy-preserving multi-subset aggregation in smart grid. *IEEE Trans. Ind. Informat.* **2018**, *14*, 462–471. [CrossRef]
15. Zhang, J.; Zhao, Y.; Wu, J.; Chen, B. LVPDA: A lightweight and verifiable privacy-preserving data aggregation scheme for edge-enabled IoT. *IEEE Internet Things J.* **2020**, *7*, 4016–4027. [CrossRef]

16. Ding, Y.; Wang, B.; Wang, Y.; Zhang, K.; Wang, H. Secure metering data aggregation with batch verification in industrial smart grid. *IEEE Trans. Ind. Inform.* **2020**, *16*, 6607–6616. [CrossRef]
17. Paillier, P. Public-key cryptosystems based on composite degree residuosity classes. In *International Conference on the Theory and Applications of Cryptographic Techniques*; Springer: New York, NY, USA, 1999; pp. 223–238.
18. Lu, R.; Liang, X.; Li, X.; Lin, X.; Shen, X. EPPA: An efficient and privacy preserving aggregation scheme for secure smart grid communications. *IEEE Trans. Parallel Distrib. Syst.* **2012**, *23*, 1621–1631.
19. Shen, H.; Zhang, M.; Wang, H. A lightweight privacy-preserving fair meeting location determination scheme. *IEEE Internet Things J.* **2020**, *7*, 3083–3093. [CrossRef]
20. Ding, W.; Yan, Z.; Deng, R.H. Encrypted data processing with homomorphic re-encryption. *Inf. Sci.* **2017**, *409*, 35–55. [CrossRef]
21. Guan, Z.; Zhang, Y.; Zhu, L.; Wu, L.; Yu, S. EFFECT: An efficient flexible privacy-preserving data aggregation scheme with authentication in smart grid. *Sci. China Inf. Sci.* **2019**, *62*, 32103. [CrossRef]
22. Li, H.; Lin, X.; Yang, H.; Liang, X.; Lu, R.; Shen, X. EPPDR: An efficient privacy-preserving demand response scheme with adaptive key evolution in smart grid. *IEEE Trans. Parallel Distrib. Syst.* **2014**, *25*, 2053–2064. [CrossRef]
23. Zhang, M.; Chen, Y.; Lin, J. A privacy-preserving optimization of neighborhood-based recommendation for medical-aided diagnosis and treatment. *IEEE Internet Things J.* **2021**, *8*, 10830–10842. [CrossRef]
24. Zhang, M.; Chen, Y.; Xia, Z.; Du, J.; Susilo, W. PPO-DFK a privacy-preserving optimization of distributed fractional knapsack with application in secure footballer configurations. *IEEE Syst. J.* **2020**, *15*, 759–770. [CrossRef]
25. Liu, Y.; Guo, W.; Fan, C.; Chang, L.; Cheng, C. A practical privacy-preserving data aggregation (3pda) scheme for smart grid. *IEEE Trans. Ind. Inf.* **2019**, *15*, 1767–1774. [CrossRef]
26. Xue, K.; Zhu, B.; Yang, Q.; Wei, D.S.L.; Guizani, M. An efficient and robust data aggregation scheme without a trusted authority for smart grid. *IEEE Internet Things J.* **2020**, *7*, 1949–1959. [CrossRef]
27. Zhao, S.; Li, F.; Li, H.; Lu, R.; Ren, S.; Bao, H.; Lin, J.H.; Han, S. Smart and practical privacy-preserving data aggregation for fog-based smart grids. *IEEE Trans. Inf. Forensics Secur.* **2021**, *16*, 521–536. [CrossRef]
28. Su, Y.; Li, Y.; Li, J.; Zhang, K. LCEDA: Lightweight and Communication-Efficient Data Aggregation Scheme for Smart Grid. *IEEE Internet Things J.* **2021**, *8*, 15639–15648. [CrossRef]
29. Huang, C.; Wang, X.; Gan, Q.; Huang, D.; Yao, M.; Lin, Y. A lightweight and fault-tolerable data aggregation scheme for privacy-friendly smart grids environment. *Clust. Comput.* **2021**, *24*, 3495–3514. [CrossRef]
30. Xu, C.; Zhang, L.; Zhu, L.; Zhang, C.; Du, X.; Guizani, M.; Sharif, K. Aggregate in my way: Privacy-preserving data aggregation without trusted authority in ICN. *Future Gener. Comput. Syst.* **2020**, *111*, 107–116. [CrossRef]
31. Tan, X.; Zheng, J.; Zou, C.; Niu, Y. Pseudonym-based privacy-preserving scheme for data collection in smart grid. *Int. J. Hoc Ubiquitous Comput.* **2016**, *22*, 120–127. [CrossRef]
32. Guan, Z.; Si, G.; Zhang, X. Privacy-preserving and Efficient Aggregation based on Blockchain for Power Grid Communications in Smart Communities. *IEEE Commun. Mag.* **2018**, *56*, 82–88. [CrossRef]
33. Liu, X.; Zhang, Y.; Wang, B.; Wang, H. An anonymous data aggregation scheme for smart grid systems. *Secur. Commun. Netw.* **2014**, *7*, 602–610. [CrossRef]
34. Sui, Z.; Alyousef, A.; de Meer, H. IAA: Incentive-based anonymous authentication scheme in smart grids. In *International Conference on Internet Science*; Springer: Cham, Switzerland, 2015; pp.133–144.
35. Yu, C.; Chen, C.; Kuo, S.; Chao, H. Privacy-preserving power request in smart grid networks. *IEEE Syst. J.* **2013**, *8*, 441–449. [CrossRef]
36. Cheung, J.; Chim, T.; Yiu, S.; Hui, L. Credential-Based Privacy-Preserving Power Request Scheme for Smart Grid Network. In Proceedings of the IEEE Global Telecommunications Conference, Houston, TX, USA, 5–9 December 2011; pp. 1–5. [CrossRef]
37. Hornget, S. b-SPECS+: Batch verification for secure pseudonymous authentication in VANET. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 1860–1875. [CrossRef]
38. Bresson, E.; Catalano, D.; Pointcheval, D. A simple public-key cryptosystem with a double trapdoor decryption mechanism and its applications. In *Advances in Cryptology-ASIACRYPT*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 37–54.
39. Shen, H.; Zhang, M.; Shen, J. Efficient privacy-preserving cube-data aggregation scheme for smart grids. *IEEE Trans. Inf. Forensics Secur.* **2017**, *12*, 1369–1381. [CrossRef]
40. Jo, H.; Kim, I.; Lee, D. Efficient and Privacy-Preserving Metering Protocols for Smart Grid Systems. *IEEE Trans. Smart Grid* **2016**, *7*, 1732–1742. [CrossRef]