sensors

MDPI

Article

Decentralized Privacy-Preserving Data Aggregation Scheme for Smart Grid Based on Blockchain

Hongbin Fan [1,2], Yining Liu [3,*] and Zhixin Zeng [3]

[1] College of Software and Communication Engineering, Xiangnan University, Chenzhou 423000, China; hongbinfan@xnu.edu.cn
[2] College of Computer Science and Technology, Hengyang Normal University, Hengyang 421002, China
[3] School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541004, China; bestzengzx@gmail.com
* Correspondence: ynliu@guet.edu.cn

Received: 13 August 2020; Accepted: 10 September 2020; Published: 15 September 2020

Abstract: As a next-generation power system, the smart grid can implement fine-grained smart metering data collection to optimize energy utilization. Smart meters face serious security challenges, such as a trusted third party or a trusted authority being attacked, which leads to the disclosure of user privacy. Blockchain provides a viable solution that can use its key technologies to solve this problem. Blockchain is a new type of decentralized protocol that does not require a trusted third party or a central authority. Therefore, this paper proposes a decentralized privacy-preserving data aggregation (DPPDA) scheme for smart grid based on blockchain. In this scheme, the leader election algorithm is used to select a smart meter in the residential area as a mining node to build a block. The node adopts Paillier cryptosystem algorithm to aggregate the user's power consumption data. Boneh-Lynn-Shacham short signature and SHA-256 function are applied to ensure the confidentiality and integrity of user data, which is convenient for billing and power regulation. The scheme protects user privacy data while achieving decentralization, without relying on TTP or CA. Security analysis shows that our scheme meets the security and privacy requirements of smart grid data aggregation. The experimental results show that this scheme is more efficient than existing competing schemes in terms of computation and communication overhead.

Keywords: decentralized; data aggregation; privacy-preservation; blockchain

1. Introduction

With the rapid development of society and economy, people's demand for electric energy is increasing, which requires that the power supply be more secure and stable. However, the traditional power system cannot keep up with the pace of technological change, the system architecture remains unchanged, which leads to the decline of power system stability and frequent safety accidents. It brings a lot of inconvenience to people's lives and causes huge economic losses to the government and enterprises. For example, in 2012, a large-scale blackout occurred in India, affecting 670 million people. Due to the low efficiency and security of the traditional power system, it cannot meet the development needs of human society. Therefore, smart grids emerge as the times require a new generation of power networks.

Smart grid is a fully-automated transmission network based on the physical grid system, which combines sensor measurement, computer, information communication, and automatic control technology [1]. The information flow between suppliers and users in smart grid is bidirectional, while the traditional power grid adopts the unidirectional centralized system. Users can control the intelligent use of household appliances and equipment at any time according to the floating

situation of electricity price in different time periods. Suppliers can automatically monitor the grid, prevent power outages, optimize grid performance, etc. Although, compared with the traditional power grid, smart grid has many excellent characteristics. However, it is easy to cause the leakage of user electricity consumption data and identity information in the process of smart grid power data collection [2,3]. For example, the blackout notification software of Vector was attacked in 2018, resulting in the disclosure of private information of thousands of customers. With the continuous integration of network, information technology, and power system; network security has become an important part of energy and power security. For example, Ukraine's power grid suffered the world's first large-scale blackout due to hacker attacks at the end of 2015, and the leakage of private information brings great security risks to the power grid and users.

In order to deal with the leakage of power consumption data and identity privacy, data aggregation, secret sharing [4–6], differential privacy [7,8], and other schemes have been proposed by predecessors. Data aggregation is one of the most common methods to solve the security and privacy problems of smart grid. In [9–19], the scheme used encryption algorithms to aggregate the power consumption data of users, and hides the data of a single user in the data of other users to protect privacy. These schemes rely on a trusted third party or a central authority, but in fact the trusted third party or the central authority is not truly reliable, and the trusted third party or the central authority can be easily knocked down by malicious attackers and leak users' private data.

Therefore, we propose a decentralized privacy-preserving data aggregation (DPPDA) scheme for smart grid based on blockchain, which is used to collect electricity consumption data without the trusted third party or the central authority in the smart grid.

We have summarized the contributions of our paper as follows:

(1) A decentralized data aggregation scheme based on blockchain is proposed. Blockchain is a new type of decentralized protocol that does not require a trusted third party or a central authority. Since the proposed scheme does not require the trusted third party or the central authority, this assumption will have a positive impact on reliability, and we can refrain from the malicious attack to the trusted third party or the central authority.
(2) The leader election algorithm is applied to select a smart meter from a residential area as a mining node (MN) to participate in the blockchain network. The MN uses Merkle hash tree to perform security authentication and data aggregation for smart meters in the residential area without any trusted third party.
(3) Paillier encryption, Boneh-Lynn-Shacham short signature, and SHA-256 function are applied to ensure the transparency of the blockchain data while achieving multiple privacy protections, which can effectively resist various security threats (such as replay attacks, tampering).

Note that the original idea has been presented in a conference [20]. In the current version, more detailed description is added to make it more easily understandable, for example, the design goals, MN election, and security analysis. Especially through the performance evaluation, it is proved that the proposed scheme is superior to the existing schemes.

The rest of this paper is organized as follows. In Section 2, the previous work in privacy-preserving data aggregation are introduced. In Section 3, blockchain, bilinear pairing, Boneh-Lynn-Shacham short signature, and the Paillier cryptosystem are given. In Section 4, the proposed system model is presented, and our scheme is proposed in Section 5. The security analysis is shown in Section 6. In Section 7, the performance of our scheme is evaluated. The research is concluded in Section 8.

2. Related Work

Privacy-preserving data aggregation in smart grids have attracted extensive attention of researchers. At present, the smart grid data aggregation solutions can be roughly divided into the following three categories.

The first category is data aggregation schemes based on traditional network architecture. Li et al. [9] proposed a privacy-preserving multi-subset data aggregation scheme (PPMA), their scheme based on Paillier cryptosystem, which enables the aggregation of electricity consumption data of different ranges. Liu et al. [10] proposed a privacy-preserving data aggregation without any TTP. This scheme uses EC-ElGamal to encrypt power consumption data and construct a virtual aggregation area for users with a certain degree of trust to shield the data of a single user. Guan et al. [11] proposed a flexible threshold for data aggregation based on the secret sharing scheme. This scheme adjusts the aggregation threshold according to the energy consumption information and time period of each specific residential area to ensure the privacy of personal data during the aggregation process, while supporting fault tolerance. Karampour et al. [12] proposed using Paillier encryption system and AV net mask to realize the aggregation of privacy protection data in smart grid can effectively protect the privacy of user data without any security channel. Chen et al. [13] proposed a data aggregation scheme based on Paillier encryption. The trusted authority generates a key for the meter to encrypt the consumed data of the meter. When a smart meter cannot work normally, the trusted authority provides the pseudo ciphertext related to the meter. The scheme solves the problem of meter failure to some extent, but it cannot completely solve the problem of privacy protection. In [14], a dynamic member data aggregation scheme based on identity signature and homomorphic encryption algorithm is proposed. The operation center obtains the sum of power consumption data in the virtual aggregation area, but knows nothing about the single user's use data. This scheme reduces the complexity of a new user joining and old user exiting. However, the above research methods do not consider the trusted environment and used a trusted third party or central authority.

The second category is data aggregation mechanism based on fog computing architecture. Lu et al. [15] proposed a privacy protection data aggregation scheme based on fog computing. In this scheme, the fog device is used as the gateway between the internet-of-things device and the control center. Lyu et al. [16] proposed a privacy-preserving aggregation scheme with the aid of fog computing architecture. This solution uses differential privacy to count user data, thereby ensuring data confidentiality. Zhu et al. [17] proposed a privacy-preserving data aggregation scheme for fog-based smart grid. Blind signature and short randomizable signature are used to provide anonymous authentication, and then fog node is used to solve the billing problem after anonymous authentication. All user data in the above solutions are concentrated in the fog layer, which inevitably brings about the problem of centralization.

The emergence of blockchain technology provides a solution to the trusted third party and centralization problems because of its decentralized characteristics. Currently, there are several studies using blockchain as privacy-preserving method for data aggregation. Guan et al. [18] proposed a privacy-preserving data aggregation scheme for power grid communications. The study divided users into different groups and each group has a private blockchain. The study uses multiple pseudonyms to hide users' identity. In this scheme, a key management center (KMC) is used to generate multiple public and private keys for users, which does not realize decentralization. Fan et al. [19] proposed a smart grid data aggregation based on consortium blockchain, and its signcryption algorithm can be applied to multidimensional data collection in the consortium blockchain. CC is the trusted third party of the scheme, which can realize key user monitoring and data recovery.

3. Preliminaries

In this section, we briefly introduce the necessary background.

3.1. Blockchain

Blockchain technology was first proposed in 2008 by Satoshi Nakamoto for Bitcoin [21]. Blockchain technology has been widely used in payment, internet of things, healthcare, finance, and so on [22]. Blockchain is a decentralized distributed ledger database maintained by network-wide nodes [23], which comprise of a chain of different data blocks in a chronological order. All hash data added to the

block is immutable. Blockchain is a new application mode of consensus mechanism, distributed data storage, encryption algorithm, and so on. The miners are responsible for creating blocks, and each block in the blockchain is identified by a hash in the header. The hash is generated by the SHA-256 hash algorithm, which uses plaintext of any size and computes a 256-bit encrypted hash of fixed size. Each header contains the address of the previous block in the chain. The information in the block cannot be deleted or changed. Blockchain has the characteristics of decentralization, anonymity, security, reliability, non-forgery, tamper resistance, and so on. Its key technologies include block structure, Merkle tree, P2P network, hash function, timestamp, asymmetric encryption mechanism, etc. [24].

(1) Merkle tree. Merkle tree is a tree that stores hash values, also known as hash tree. The value of the Merkle tree leaf node is the hash value of the data block. The value of a non-leaf node is the hash of its corresponding child node concatenation string. Merkel root is the root value of the hash tree calculated by all transactions in the current block.
(2) SHA-256. SHA-256 is the most widely used cryptographic secure hash algorithm (SHA) in the blockchain, which is used to maintain the data integrity within the block. It provides a unique 256-bit hash code, also called data file signature.
(3) Timestamp. The blockchain uses timestamp to realize that all recorded transaction data are encoded by time information, which ensures the traceability and verifiability of the recorded data in the database. The "timestamp" technology makes the blockchain database non-tamperable and unforgeable, so it is also called proof-of-existence of the block data.

3.2. Bilinear Pairing

G_1 and G_2 are two q-order prime cyclic additive groups. $e : G_1 \times G_1 \rightarrow G_2$ is a bilinear mapping [25,26] that has the following properties.

(1) Bilinearity: $e(u^a, v^b) = e(u, v)^{ab}$ for all $u, v \in G_1$, and $a, b \in Z_q^*$.
(2) Non-degeneracy: for all $u, v \in G_1$, $e(u, v) \neq 1$.
(3) Computability: there exists an efficient algorithm to compute $e(u, v)$ for all $u, v \in G_1$.

3.3. Boneh-Lynn-Shacham Short Signature

Boneh-Lynn-Shacham (BLS) short signature [27] scheme is a typical bilinear pairing scheme, which uses SHA-256 hash function $H_1 : \{0,1\}^* \rightarrow G_1$ and g is a random generator of G_1, and a bilinear map $e : G_1 \times G_1 \rightarrow G_2$. The BLS signature scheme is divided into three phases: key generation, signature, and verification.

(1) Key generation. The secret key $x \in Z_q^*$, and compute the public key $PK = x \cdot g$.
(2) Signature. The plaintext $m \in G_1$, compute the signature $\sigma = x \cdot H(m)$.
(3) Verification. If $e(\sigma, g) = e(H(m), PK)$, then the signature is verified. Otherwise fails.

3.4. Paillier Cryptosystem

Paillier cryptosystem [28] is a probabilistic public-key cryptosystem that uses asymmetric encryption algorithm, which can effectively implement homomorphic properties. The encryption algorithm satisfies homomorphism of addition and multiplication, and can operate directly on the ciphertext without needing to know the corresponding plaintext. Therefore, it is widely used in many privacy protection applications. It includes three algorithms: key generation, encryption, and decryption.

(1) Key generation. Randomly select two large primes p and q, where $|p|=|q|=|\kappa|$. Then calculate $\lambda = lcm(p-1, q-1)$. Defined a function $L(v) = \frac{v-1}{N}$, where $N = pq$. Choose a generator $g \in Z_{N^2}^*$, and calculate $\mu = \left(L(g^\lambda \bmod N^2)\right)^{-1} \bmod N$. The public key is (N, g), and the corresponding private key is (λ, μ).

(2) Encryption. Given a message $m \in Z_N$, choose a random number $r \in Z_N^*$, $\gcd(r, N) = 1$. The ciphertext is calculated as $C = Enc(m) = g^m \cdot r^N \bmod N^2$.

(3) Decryption. Given the ciphertext $C \in Z_N$, the corresponding message is decrypted with the private key (λ, μ) as $m = Dec(C) = L(C^\lambda \bmod N^2) \cdot \mu \bmod N$.

4. System Model

4.1. Communication Model

The system model of our scheme consists of operation center (OC) and smart meter (SM) in the residential area (RA), which is demonstrated in Figure 1. The system consists of L residential areas, and each residential area contains several smart meters. In our scheme, we mainly focus on removing the control center and the trusted third party while protecting the data privacy of the user's smart meter.

(1) Operation center (OC). OC reads the real-time total power consumption data aggregated by the mining nodes of L blocks through the blockchain. OC can also perform billing, power consumption trend analysis, adjustment of power generation plans, and dynamic pricing. OC is vulnerable to attacks by external adversary. Therefore, OC is not assumed to be trusted.

(2) Smart meter (SM). A SM is an electricity meter for each user's site in the residential area. The smart meter regularly and simultaneously (e.g., every 15 min) collects the power consumption data of each user's household electrical equipment. Peer-to-peer (P2P) communication is used between all SMs in each residential area. Each residential area uses leader election algorithm to select a smart meter from the smart meters as the mining node (MN), then each residential area constructs a block through a MN. The MN selected by the MN selection algorithm can replace a trusted third party or a trusted authority, it is responsible for generating system parameters, authenticates the legitimacy of the data transmitted by the smart meter, and aggregates the encrypted data. Then, SM encrypts all kinds of collected data and uploads it to the MN after a short period of time. SM is assumed to be honest-but-curious, which executes the operations according the protocol without launching the active attack. However, it perhaps tries to analyze the received data to infer some valuable information.

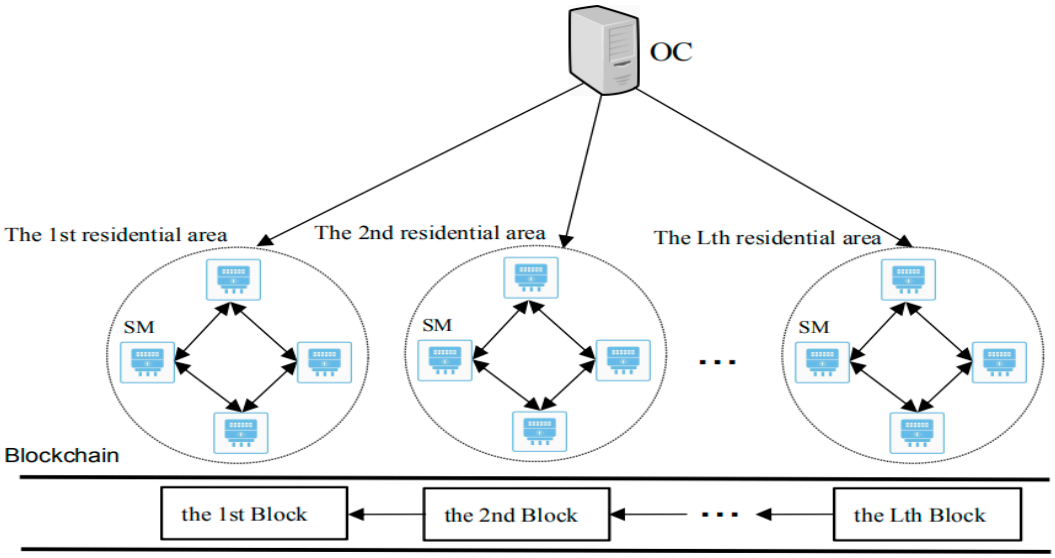

Figure 1. System model.

4.2. Design Goals

To solve the issues mentioned above, ensure the integrity and privacy of users' power consumption data while decentralizing or not relying on the trusted third parties, the design goals include five aspects.

(1) Privacy-preservation. Neither OC nor any other user has access to other user's data in the residential area. An external adversary cannot obtain the user's power consumption data, even if he knows the ciphertext. Even if the adversary and OC collude with each other, they can't get the power consumption data of a single user's smart meter.
(2) Decentralizing. Our scheme does not need a trusted third party or a central authority. The leader election algorithm is used to select a smart meter in the residential area as the mining node, which is responsible for building the Merkle tree of the block and aggregating the power consumption data of the residential area.
(3) Data unforgeability and non-repudiation. Our scheme adopts BLS short signature in blockchain, which is based on bilinear pair to ensure the unforgeability and non-repudiation of data.
(4) Data security. The proposed scheme can defend against various attacks. Even if the aggregate ciphertext of users' electricity consumption data is intercepted, the individual user's electricity consumption data cannot be recovered.
(5) Confidentiality. The data of electricity consumption belongs to personal privacy, which can reflect the real-time power consumption of users' homes. Once the data is leaked, it will be used by criminals to commit crimes. Data confidentiality should be maintained by a secure data aggregation scheme. Even if an attacker steals the ciphertext, it will not be able to obtain the power consumption data of a single user.

5. The Proposed Scheme

In this section, a decentralized smart grid privacy protection data aggregation scheme based on block chain is proposed, which consists of five phases: system initialization, ciphertext generation, ciphertext aggregation, ciphertext decryption, and data reading. The notations are listed in Table 1.

Table 1. Notations.

Symbol	Quantity
g_1, g_2	A generator of G
RA_j	the jth residential area
m_i	Power consumption data of the ith smart meter in RA_j
n	Number of smart meters in the jth residential area
H_1	Hash functions: H_1: $\{0,1\}^* \rightarrow G$
L	Number of residential areas
SM_i	Smart meter in jth residential area
MN_j	Mining node of the jth residential area
M_j	the aggregated electricity consumption data of the jth residential areas
$\|\|$	Concatenation operation

Each smart meter in the system acts as a node, and each node has three states: follower, MN, and candidate. All nodes start from the follower state. Each term begins with an election in which one or more candidates try to become MNs. If a candidate wins the election, it will be a MN for the rest of its term. The state change of MN election algorithm is shown in Figure 2.

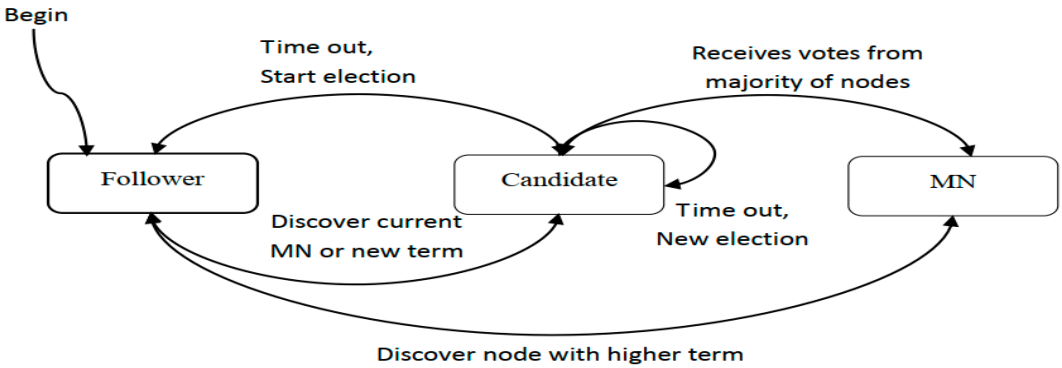

Figure 2. State transition model of MN election algorithm.

5.1. System Initialization

OC collects electricity consumption data of smart meters in L residential areas. There are n smart meters in RA_j. Through Algorithm 1, it selects a SM as a mining node from the n SMs in RA_j, then constructs the jth block, where MN_j is the root of the Merkle tree in the jth block. The consumption data of SMs in RA_j is aggregated to MN_j through Merkle tree. The structure of blockchain is shown in Figure 3.

Algorithm 1. MN Election

1. Set the initial state of SM[i] to ***Follower***,i∈[1,n],n is the number of SMs in RA;
2. Let the number of terms of SM[i] elected as MN be 0,TN = 0;
3. Set the number of votes obtained by SM[i] to 0,Nv = 0;
4. Start the Timer of Follower FT;
5. Set a random timeout of Follower FRT_{out};
6. **while** $FT > FRT_{out}$ **do**
7. The state of SM[i] has changed from ***Follower*** to ***Candidate***;
8. $TN = TN + 1$;
9. Start the Timer of Candidate CT;
10. Set a random timeout of Candidate CRT_{out};
11. $Nv = Nv + 1$;
12. SM[i] with Candidate state sends a request of voting to other SMs;
13. SM[i] counts the number k of voting responses received from other SMs;
14. $Nv = Nv + k$;
15. **if** $Nv > n/2 + 1$ **then**
16. The state of SM[i] has changed from *Candidate* to *MN*;
17. SM[i] sends messages that are selected as MN to other SMs;
18. **end if**
19. **if** SM[i] receives messages from a SM that is selected as MN **then**
20. The state of SM[i] has changed from *Candidate* to *Follower*;
21. **end if**
22. **while** $CT > CRT_{out}$ **do**
23. Repeat step 8–11 for a new election
24. **endwhile**
25. **endwhile**

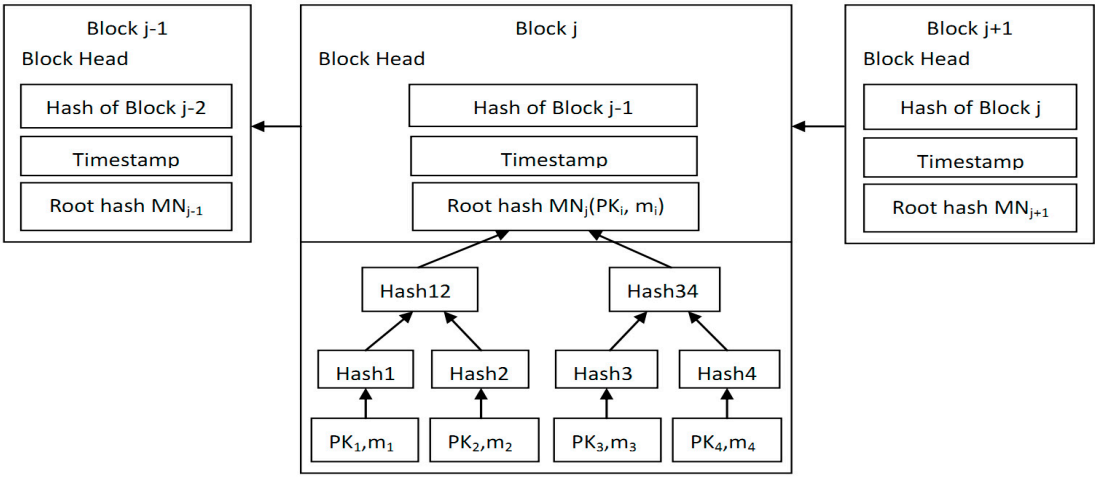

Figure 3. Blockchain structure in our scheme.

MN_j runs Bilinear parameter generator $Gen(\kappa)$ to generate (q, g_1, G_1, G_2, e), and g_1 is a generator of G_1. MN_j calculates Paillier cryptosystem public key (N, g_2), corresponding private key (λ, μ), $g_2 \in Z^*_{N^2}$. MN_j choose a SHA-256 hash function H_1 and a secure cryptographic hash function H_2, where $H_1 : \{0,1\}^* \rightarrow G_1$, $H_2 : \{0,1\}^* \rightarrow \{0,1\}^\kappa$.

MN_j publishes the system public parameter $\{q, g_1, g_2, G_1, G_2, e, N, H_1\}$.

5.2. Ciphertext Generation

Step 1 SM_i selects a random number $x_i \in Z^*_q$ as the private key and computes the corresponding public key $PK_i = x_i \cdot g_1$.

Step 2 SM_i collects electricity consumption data m_i at timestamp T, and computes the Hash value $H_2(T)$, then selects a random number $r_i \in Z^*_N$ to generate ciphertext: $C_i = g_2^{m_i} \times (r_i \times H_2(T))^N \bmod N^2$.

Step 3 SM_i generates the BLS short signature $\sigma_i = x_i \cdot H_1(C_i \parallel PK_i \parallel Ts_i)$, Ts_i is the current timestamp to prevent replay attack.

Step 4 SM_i sends $C_i \parallel PK_i \parallel Ts_i \parallel \sigma_i$ to MN through the Merkle tree.

5.3. Ciphertext Aggregation

After MN_j receives users' data $C_i \parallel PK_i \parallel Ts_i \parallel \sigma_i$, it performs the following steps for privacy-preserving data aggregation.

Step 1 MN_j verifies n signatures after receiving $C_i \parallel PK_i \parallel Ts_i \parallel \sigma_i$. If $e(\sigma_i, g_i) = e(H_1(C_i \parallel PK_i \parallel Ts_i), PK_i)$ validation is successful and fails otherwise. If it holds, the signature is valid and MN_j will accept SM_i's ciphertext. In order to make the verification more efficient, MN_j adopts batch verification

$$e(\sigma_i, g_1) \underline{?} e(H_1(C_i \parallel PK_i \parallel Ts_i), PK_i)$$

The proof is given as follows.

$$\begin{aligned} e(\sum_{i=1}^{n} \sigma_i, g_1) &= e(\sum_{i=1}^{n} x_i \cdot H_1(C_i \parallel PK_i \parallel Ts_i), g_1) \\ &= \prod_{i=1}^{n} e(x_i \cdot H_1(C_i \parallel PK_i \parallel Ts_i), g_1) \\ &= \prod_{i=1}^{n} e(H_1(C_i \parallel PK_i \parallel Ts_i), x_i \cdot g_1) \end{aligned}$$

Step 2 MN_j aggregates the ciphertext.

$$\begin{aligned} C = Enc(m) &= \prod_{i=1}^{n} C_i \\ &= \prod_{i=1}^{n} g_2 \cdot (r_i \cdot H_2(T))^N \bmod \mathrm{N}^2 \\ &= g_2^{\sum_{i=1}^{n} m_i} \cdot \prod_{i=1}^{n} (r_i \cdot H_2(T))^N \bmod \mathrm{N}^2 \end{aligned}$$

5.4. Ciphertext Decryption

MN_j uses the private key (λ, μ) to decrypt the aggregated ciphertext to obtain the aggregated electricity consumption data M_j of the jth residential district.

$$\begin{aligned} M_j &= Dec(C) = L(C^\lambda \bmod \mathrm{N}^2) \cdot \mu \bmod \mathrm{N} \\ &= \frac{L(C^\lambda \bmod \mathrm{N}^2)}{L(g_2^\lambda \bmod \mathrm{N}^2)} \bmod \mathrm{N} \\ &= \frac{L(g_2^{\sum_{i=1}^{n} m_i \lambda} \bmod \mathrm{N}^2)}{L(g_2^\lambda \bmod \mathrm{N}^2)} \bmod \mathrm{N} \end{aligned}$$

5.5. Data Reading

MN_j generates the $(j+1)$th block, and adds the jth block to the blockchain after the $(j-1)$th block. OC obtains the power consumption data through the public key read blockchain.

6. Security Analysis

The security of DPPDA in smart grid is compared with that of schemes [9,11–13], as shown in Table 2.

Table 2. Comparison between the proposed scheme and other related schemes.

Security Requirements	[9]	[11]	[12]	[13]	DPPDA
Blockchain-based	No	No	No	Yes	Yes
Decentralization	No	No	No	No	Yes
Non-repudiation	No	Yes	No	Yes	Yes
Privacy	Yes	Yes	Yes	Yes	Yes
Confidentiality	Yes	Yes	Yes	Yes	Yes
Data integrity	Yes	Yes	Yes	Yes	Yes
Replay attack resistance	No	Yes	Yes	Yes	Yes
Data unforgeability	No	Yes	Yes	Yes	Yes

6.1. Privacy-Preservation

To avoid the leakage of the power consumption data m_i, we mainly consider the external attack and the internal attack.

First, we assume that the external adversary may eavesdrops the communication between SMs and MN to obtain the electricity consumption data m_i. In DPPDA, SM_i reports m_i to MN_j in the form of $C_i = g_2^{m_i} \times (r_i \times H_2(T))^N \bmod N^2$. Let $r = r_i \times H_2(T)$, then the ciphertext expression will become $C_i = g_2^{m_i} \times r^N \bmod N^2$. The ciphertext C_i is still the legal ciphertext of the Paillier cryptosystem. Because the adversary does not know the decryption key λ of the Paillier encryption algorithm, the adversary cannot decrypt the ciphertext C_i to obtain the power consumption data of a single user. The power consumption data of a single smart meter is not disclosed, so as to protect the privacy of users.

Second, we assume that the internal adversary includes SM_1, SM_2, ⋯, SM_{n-1}, and they collude to obtain the power consumption m_n of SM_n. The expression of n SMs is expressed as: $\sum_{i=1}^{n} m_i = 0 \bmod \lambda$. For $(n-1)$ users, the expression can be rewritten as: $m_n + \sum_{i=1}^{n-1} m_i = 0 \bmod \lambda$. This means without having Paillier's secret key λ, the internal adversary will not be able to obtain m_n. We can conclude that, no matter how many SMs are colluded, the internal adversary cannot disclose the power consumption data m_i of the other users.

6.2. Decentralized

In our scheme, the blockchain can be implemented without a trusted third party or central authority, the availability and reliability of data is guaranteed by MN election. Any SM is not controlled or operated by other SMs and OC. P2P network is adopted among smart meters to realize decentralization. The whole process does not rely on a trusted third party to make our solution more reliable and convenient.

6.3. Data Security

The electricity consumption data of SM_i in RA_j is encrypted as $C_i = g_2^{m_i} \times (r_i \times H_2(T))^N \bmod N^2$, m_i is secure and privacy-preservation. Even if an adversary intercepts C_i, he/she cannot recover the power consumption data of a single smart meter. After MN collects all the smart meter power consumption data in the residential area through data aggregation, only the aggregated data can be obtained through decryption, and the plaintext of single smart meter power consumption data cannot be recovered.

6.4. Confidentiality

The power consumption data includes user privacy and business secrets. The usage data of the smart meters are encrypted by Paillier cryptosystem algorithm derived from [29]. After receiving the ciphertext of the smart meters in the residential area, only MN can decrypt the aggregated plaintext data. Since Theorem 1 of [28] represents confidentiality based on the DDH assumption, even if the adversary eavesdrops on the ciphertext of the smart meters in the residential area, the adversary still cannot infer any relevant information about usage data sent by the smart meters. The confidentiality of user power consumption data is guaranteed.

6.5. Data Integrity and Non-Repudiation

SHA is an anti-collision algorithm where different inputs (data information) cannot produce the same output (hash value), so SHA-256 can be used to check whether the data information is the same. The integrity of the data is determined by comparing the calculated "hash value" with the known hash value. Each smart meter in the scheme signs the message to be sent. MN receives the message after verifying the signature to ensure the integrity of the data and prevent tampering. Each smart meter's private key is kept by itself and cannot be denied the information it sends and signs.

6.6. Data Unforgeability

All SMs use their private keys to sign their messages before sending MN use SM public keys to verify received messages. The proposed scheme uses the BLS signature based on the CDH [30], which makes it impossible for any attacker to forge a new signature by eavesdropping on the original signature. BLS short signature and blockchain are used to verify the source and authenticity of power consumption data. Since all transactions in the blockchain have timestamps and all hash data added to the block cannot be changed, the data in the blockchain has unforgeability.

7. Performance Evaluation

The performance of our scheme is evaluated in this section, including the computation complexity of SM and OC, and the communication overhead.

7.1. Computation Complexity

Compared with multiplication operation and exponentiation operation, leader election and hash operation is negligible. In our scheme, the computations in the data aggregation process mainly include three phases, data encryption, batch verification, and aggregation, decryption. We denote the computational cost of an exponentiation operation and a multiplication operation, by T_{exp}, T_{mul}, respectively. The computation complexities of the major entities in the system are as show in Table 3.

Table 3. Comparing computation complexity between the proposed scheme and other schemes.

Scheme Ref.	[9]	[11]	[12]	DPPDA
Overhead SM	$3T_{exp} + 4T_{mul}$	$4T_{exp} + 3T_{mul}$	$2T_{exp} + nT_{mul}$	$2T_{exp} + 4T_{mul}$
Overhead GW	nT_{mul}	$3T_{exp} + (2n+1)T_{mul}$	nT_{mul}	-
Overhead CC	$T_{exp} + (4n+3)T_{mul}$	$3T_{exp} + 2T_{mul}$	$T_{exp} + 3T_{mul}$	-
Overhead MN	-	-	-	$T_{exp} + (n+1)T_{mul}$

We conduct the experiments with the cpabe0.10 [31] library on a 3.0-GHz processor and a 2-GB memory PC. As shown in Figure 4, compared with PPMA, EFFECT, and Karampour's schemes, our scheme has much less computational overhead. As the number of users increases, the advantages of our scheme become more obvious.

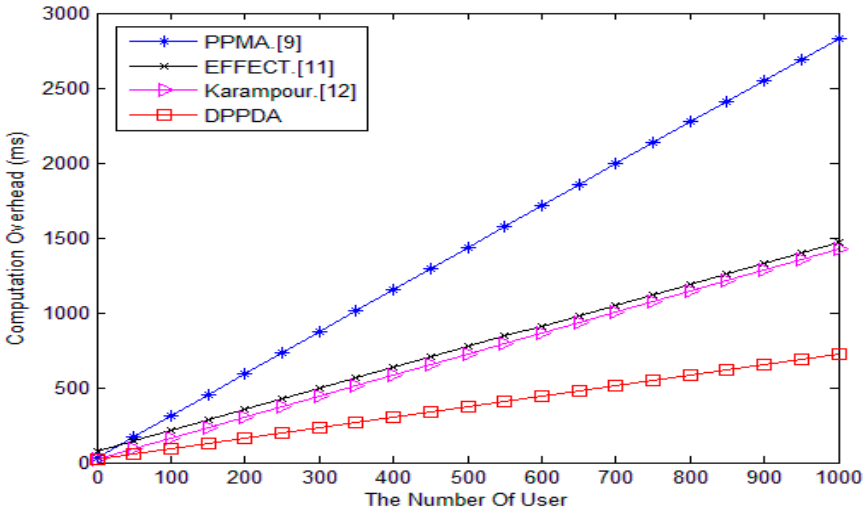

Figure 4. Comparison of computational cost.

7.2. Communication Overhead

The communication of our proposed scheme is only SM_i to MN_j, the power consumption data collected by SM_i is used to generate the report $(C_i \parallel PK_i \parallel Ts_i \parallel \sigma_i)$ and is sent to MN_j. The size of SM_i report is $S_z = |C_i| + |PK_i| + |Ts_i| + |\sigma_i|$. The maximum communication overhead is $S_{\max} = n \cdot S_z = n \cdot (|C_i| + |PK_i| + |Ts_i| + |\sigma_i|)$. Suppose that C_i generates a 2048-bit ciphertext and chooses 160-bit Z_N^*, 160-bit G.

In PPMA and EFFECT scheme, the communication cost on SM-to-GW is $2048n$ bit, the communication cost on GW-to-CC is 2048 bit, the total communication overhead is $2048(n + 1)$ bit. In Karampour's scheme, the communication cost on SM-to-SM is $n(2048(n - 1))$ bit, the communication

cost on SM-to-GW is 2048n bit, the communication cost on GW-to-CC is 2048 bit, the total communication overhead is 2048(n^2 + 1) bit.

In our scheme, the total communication overhead is 2048n bit. The comparison is shown in Table 4, the total communication cost of our scheme is less than the other schemes.

Table 4. Comparing communication cost between the proposed scheme and other schemes.

Scheme Ref.	[9]	[11]	[12]	DPPDA
SM-to-SM (bit)	-	-	n(2048 (n − 1))	-
SM-to-GW (bit)	2048n	2048n	2048n	-
GW-to-CC (bit)	2048	2048	2048	-
SM-to-MN (bit)	-	-	-	2048n

In Figure 5, we plot the communication cost in PPMA, EFFECT, Karampour's, and our scheme versus the SM number n. It is shown that our scheme does not bring too much communication overhead.

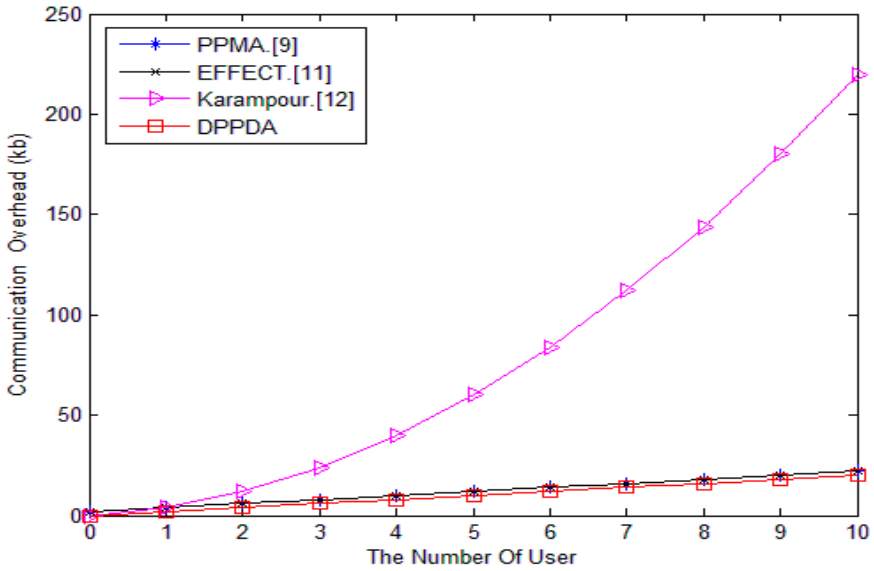

Figure 5. Comparison of communication cost.

8. Conclusions

In this paper, a decentralized smart grid privacy-preservation data aggregation scheme based on blockchain is proposed. The smart meters select a mining node through leader election algorithm, which records the data of smart meters into the blockchain. BLS signature and Paillier encryption are based on bilinear pairing, which guarantees the security and integrity of messages during transmission. Security analysis shows that our mechanism meets the requirements of privacy protection and security of smart meters. The performance evaluation shows that our scheme is superior to some popular data aggregation schemes in computational efficiency. Our scheme has low communication overhead and does not require any trusted third party, trusted authority, and secure channels. At present, we have decentralized the aggregation of one-dimensional power consumption data. In the future, we will work on the combination of blockchain and other algorithms to aggregate multidimensional power consumption data.

Author Contributions: Y.L. supervised the study; H.F. and Y.L. designed the study, H.F. wrote the manuscript, performed the experiment research and the theoretical analysis; H.F. and Z.Z. prepared the figures and tables. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Intelligent Information Processing and Application Hunan Provincial Key Laboratory Open Fund Project of Hengyang Normal University (grant no. IIPA20K05), Xiangnan University Information Teaching Application Construction Project in 2016 (grant Xiangnan University xiao fa (2016) no. 60,7).

Acknowledgments: We thank all the editors for their comments on this manuscript.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Fang, X.; Misra, S.; Xue, G.; Yang, D. Smart Grid—The New and Improved Power Grid: A Survey. *IEEE Commun. Surv. Tutor.* **2011**, *14*, 944–980. [CrossRef]
2. Xue, K.; Li, S.; Hong, J.; Xue, Y.; Yu, N.; Hong, P. Two-Cloud Secure Database for Numeric-Related SQL Range Queries With Privacy Preserving. *IEEE Trans. Inf. Forensics Secur.* **2017**, *12*, 1596–1608. [CrossRef]
3. Wu, J.; Dong, M.; Ota, K.; Liang, L.; Zhou, Z. Securing distributed storage for Social Internet of Things using regenerating code and Blom key agreement. *Peer-to-Peer Netw. Appl.* **2014**, *8*, 1133–1142. [CrossRef]
4. Guan, Z.; Si, G.; Du, X.; Liu, P. Protecting User Privacy Based on Secret Sharing with Error Tolerance for Big Data in Smart Grid. *arXiv* **2018**, arXiv:1811.06918.
5. Chen, J.; Liu, G.; Liu, Y. Lightweight Privacy-preserving Raw Data Publishing Scheme. *IEEE Trans. Emerg. Top. Comput.* **2020**, 1. [CrossRef]
6. Liu, Y.-N.; Zhao, Q. E-voting scheme using secret sharing and K-anonymity. *World Wide Web* **2018**, *22*, 1657–1667. [CrossRef]
7. Hassan, M.U.; Rehmani, M.H.; Kotagiri, R.; Zhang, J.; Chen, J. Differential privacy for renewable energy resources based smart metering. *J. Parallel Distrib. Comput.* **2019**, *131*, 69–80. [CrossRef]
8. Piao, C.; Shi, Y.; Yan, J.; Zhang, C.; Liu, L. Privacy-preserving governmental data publishing: A fog-computing-based differential privacy approach. *Future Gener. Comput. Syst.* **2019**, *90*, 158–174. [CrossRef]
9. Li, S.; Xue, K.; Yang, Q.; Hong, P. PPMA: Privacy-Preserving Multisubset Data Aggregation in Smart Grid. *IEEE Trans. Ind. Inf.* **2018**, *14*, 462–471. [CrossRef]
10. Liu, Y.-N.; Guo, W.; Fan, C.-I.; Chang, L.; Cheng, C. A Practical Privacy-Preserving Data Aggregation (3PDA) Scheme for Smart Grid. *IEEE Trans. Ind. Inf.* **2018**, *15*, 1767–1774. [CrossRef]
11. 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]
12. Karampour, A.; Ashouri-Talouki, M.; Ladani, B.T. An Efficient Privacy-Preserving Data Aggregation Scheme in Smart Grid. In Proceedings of the 2019 27th Iranian Conference on Electrical Engineering (ICEE), Yazd, Iran, 30 April–2 May 2019; pp. 1967–1971.
13. Chen, L.; Lu, R.; Cao, Z. PDAFT: A privacy-preserving data aggregation scheme with fault tolerance for smart grid communications. *Peer-to-Peer Netw. Appl.* **2015**, *8*, 1122–1132. [CrossRef]
14. Song, J.; Liu, Y.-N.; Shao, J.; Tang, C. A Dynamic Membership Data Aggregation (DMDA) Protocol for Smart Grid. *IEEE Syst. J.* **2020**, *14*, 900–908. [CrossRef]
15. Lu, R.; Heung, K.; Lashkari, A.H.; Ghorbani, A.A. A Lightweight Privacy-Preserving Data Aggregation Scheme for Fog Computing-Enhanced IoT. *IEEE Access* **2017**, *5*, 3302–3312. [CrossRef]
16. Lyu, L.; Nandakumar, K.; Rubinstein, B.; Jin, J.; Bedo, J.; Palaniswami, M.; Rubinstein, B. PPFA: Privacy Preserving Fog-Enabled Aggregation in Smart Grid. *IEEE Trans. Ind. Inform.* **2018**, *14*, 3733–3744. [CrossRef]
17. Zhu, L.; Li, M.; Zhang, Z.; Xu, C.; Zhang, R.; Du, X.; Guizani, N. Privacy-Preserving Authentication and Data Aggregation for Fog-Based Smart Grid. *IEEE Commun. Mag.* **2019**, *57*, 80–85. [CrossRef]
18. Guan, Z.; Si, G.; Zhang, X.; Wu, L.; Guizani, N.; Du, X.; Ma, Y. Privacy-Preserving and Efficient Aggregation Based on Blockchain for Power Grid Communications in Smart Communities. *IEEE Commun. Mag.* **2018**, *56*, 82–88. [CrossRef]
19. Fan, M.; Zhang, X. Consortium Blockchain Based Data Aggregation and Regulation Mechanism for Smart Grid. *IEEE Access* **2019**, *7*, 35929–35940. [CrossRef]

20. Fan, H.; Liu, Y.; Zeng, Z. Blockchain-based Decentralized Privacy-Preserving Data Aggregation (BDPDA) Scheme for Smart Grid. In Proceedings of the 2020 The 2nd International Conference on Blockchain Technology, Hilo, HI, USA, 12–14 March 2020.
21. Nakamoto, S. *Bitcoin: A Peer-to-Peer Electronic Cash System*; BN Publishing: Hawthorne, CA, USA, 2019.
22. Crosby, M.; Pattanayak, P.; Verma, S.; Kalyanaraman, V. Blockchain technology: Beyond bitcoin. *Appl. Innov.* **2016**, *2*, 71.
23. Yuan, Y.; Wang, F.-Y. Parallel blockchain: Concept, methods and issues. *Acta Autom. Sin.* **2017**, *43*, 1703–1712.
24. Xie, Q.H. Research on blockchain technology and financial business innovation. *Financ. Dev. Res.* **2017**, *5*, 77–82.
25. Boneh, D.; Franklin, M. Identity-Based Encryption from the Weil Pairing. In *Advances in Cryptology—CRYPTO 2001*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2001; Volume 2139, pp. 213–229.
26. Joux, A. A One Round Protocol for Tripartite Diffie–Hellman. In *International Algorithmic Number Theory Symposium*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2000; Volume 1838, pp. 385–393.
27. Boneh, D.; Lynn, B.; Shacham, H. Short Signatures from the Weil Pairing. In *Advances in Cryptology-ASIACRYPT 2001*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2001; Volume 2248, pp. 514–532.
28. Paillier, P. Public-Key Cryptosystems Based on Composite Degree Residuosity Classes. In *Advances in Cryptology-EUROCRYPT '99*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2007; Volume 1592, pp. 223–238.
29. Shi, E.; Chan, T.-H.; Rieffel, E.G.; Chow, R.; Song, D. Privacy-preserving aggregation of time-series data Annual Network & Distributed System Security Symposium (NDSS). *Int. Soc.* **2011**, *2*, 1–17.
30. Bao, F.; Deng, R.H.; Zhu, H. Variations of Diffie-Hellman Problem. In *Information and Communications Security*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2003; Volume 2836, pp. 301–312.
31. Bethencourt, J. Advanced Crypto Software Collection: The CPABE Toolkit. 2018. Available online: http://acsc.cs.utexas.edu/cpabe/ (accessed on 24 March 2018).