future internet

MDPI

Article

A Data Sharing Scheme for GDPR-Compliance Based on Consortium Blockchain

Yangheran Piao, Kai Ye and Xiaohui Cui *

Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, Wuhan 430072, China; CDra_90n@whu.edu.cn (Y.P.); kai_ye@whu.edu.cn (K.Y.)
* Correspondence: xcui@whu.edu.cn

Abstract: After the General Data Protection Regulation (GDPR) was introduced, some organizations and big data companies shared data without conducting any privacy protection and compliance authentication, which endangered user data security, and were punished financially for this reason. This study proposes a blockchain-based GDPR compliance data sharing scheme, aiming to promote compliance with regulations and provide a tool for interaction between users and service providers to achieve data security sharing. The zero-knowledge Succinct Non-Interactive Arguments of Knowledge (zk-SNARK) algorithm is adopted for protecting data and ensure that the user's private data can satisfy the individual requirements of the service provider without exposing user data. The proposed scheme ensures mutual authentication through the Proof of Authority consensus based on the Committee Endorsement Mechanism (CEM-PoA), and prevents nodes from doing evil using the reputation incentive mechanism. Theoretical analysis and performance comparison indicate that the scheme meets the confidentiality, availability, and other indicators. It has superiority in efficiency and privacy protection compared with other schemes.

Keywords: data security; privacy-preservation; big data; blockchain; GDPR

Citation: Piao, Y.; Ye, K.; Cui, X. A Data Sharing Scheme for GDPR-Compliance Based on Consortium Blockchain. *Future Internet* **2021**, *13*, 217. https://doi.org/10.3390/fi13080217

Academic Editor: Maumita Bhattacharya

Received: 30 July 2021
Accepted: 19 August 2021
Published: 21 August 2021

1. Introduction

In the era of big data, the security and privacy issues of user data are becoming more and more serious. GDPR has been enforced since May 2018, and it has triggered extensive discussions and profound changes in personal data protection legislation [1], law enforcement, corporate compliance, and international cooperation globally. Numerous companies believe that they are not fully prepared to comply with the regulations [2], and even have a significant gap. GDPR poses a considerable challenge to companies from an administrative and legal perspective and a technical perspective. The fine for violations may be as high as EUR 20 million [3], accounting for 4% of the company's annual global revenue. Public information statistics reveal that in the first year of GDPR, regulatory authorities in Europe have made 48 penalties of approximately EUR 51,833,345. Since the second half of 2019 [4], regulatory authorities have issued British Airways, Marriott's hundreds of millions of euros and Austrian Post, Deutsche Wohnen SE, and other fines of tens of millions of euros. Most of the penalties are due to the failure to ensure data security or secure data sharing, which lead to data leakage.

The application of blockchain in GDPR compliance appears to be a hot direction since it owns the advantages of traceability and non-tamperability, it naturally satisfies most of the GDPR requirements. Yet conversely, the existing GDPR blockchain solutions mainly concentrate on how to use blockchain to enhance the transparency of data sources, fine-grained access control, and GDPR compliance [5]. Until the present, few studies have comprehensively considered user privacy and data exchange security.

In this study, our motivation is to protect users' privacy during the entire process of GDPR-compliant data sharing and transactions. We propose a solution for privacy

protection based on the consortium blockchain, it will play a practical role in the personal data sharing scenario under the GDPR. The main contributions of this paper are listed as follows:

1. This paper designs a blockchain-based data sharing scheme to achieve GDPR compliance and secure data sharing. In the proposed scheme, the blockchain ensures traceability.
2. The proposed scheme employs the zk-SNARK algorithm to guarantee the availability and privacy-preservation of data when interacting with smart contracts. Users construct zero-knowledge proofs and submit them to the smart contract to ensure that privacy data can meet service provider's requirements and will not leak. In addition, confidentiality is guaranteed by symmetric encryption.
3. After the success of verification, transactions between the user and the service provider are published on the blockchain. We have designed a PoA consensus algorithm based on a committee endorsement mechanism to ensure data integrity.
4. In order to prevent nodes from doing nothing or doing evil, we apply identity reputation as an incentive mechanism to ensure mutual authentication.
5. Performance analysis shows that this solution can ensure regulatory traceability, mutual authentication, and other indicators presented in Section 5. Compared with previous work, the proposed scheme in this paper is feasible and efficient.

The proposed solution can ensure the security and privacy of data, without affecting the company's or organization's ability to provide personalized services. Based on the results of our contributions, we believe the design conception of our scheme is helpful to researchers who continue their research in the field of data sharing.

The structure of this paper is presented as follows: Section 2 introduces the background and related work, including GDPR and blockchain-related technologies. In Section 3, we clarify the model of the scheme. Section 4 illustrates the specific implementation details. The performance analysis and evaluation of the proposed scheme are performed in Section 5. Section 6 summarizes the work of this paper and discusses the future research directions.

2. Background and Related Work

2.1. GDPR

On 25 May 2018, the GDPR bill formally took effect in all EU member states. This move indicates that the EU has achieved unprecedented stringency in the protection and supervision of personal information. As released by the statistics, from July 2018 to June 2020, a total of 264 GDPR fines were publicly available [6], and the number and scale of fines increased exponentially. According to the provisions of GDPR [7], each country in the EU member states should set up a corresponding regulatory authority to effectively supervise the personal data processing activities within the country and coordinate data processing activities among other member states. Roles defined by GDPR are presented as:

Data Subject: The Data Subject is the actual owner of the personal data, which shares the personal data with the data controller in order to obtain services.

Data Controller: Organizations or companies that collect and manage personal data for profit or service purposes are service providers [8] in this article (since service providers may share collected data with third parties for profit). Its primary responsibilities include the necessity to prove that it collects personal information within the scope of GDPR. Moreover, it is legal, truthful, transparent, and accurate, which has done a minimum of collecting personal information.

Data Processor: Organizations or companies that provide services to users through a service provider's infrastructure and obtain the required personal data. When processing personal data for their own business, it can be the service provider itself. When it wants to share the personal data of other companies, it plays the role of a third party.

2.2. Blockchain and Smart Contract

Blockchain [9] technology comes from Bitcoin proposed by Satoshi Nakamoto, and its most significant feature refers to the fact that it cannot be tampered with. This property is

mainly obtained through global sharing and hash chain structure. Based on the timestamps of its transactions and messages, the blockchain provides universally verifiable proof of transactions in the distributed database. The underlying cryptographic primitives using hash functions and digital signatures provide the reliability of these proofs. Therefore, it is guaranteed that there are computational security and verifiability [10] at any point in time. The definition of a smart contract is a program that is event-driven and stateful, which runs on a replicated and shared ledger, and can keep assets on the ledger.

2.3. Zero-Knowledge Proof

Zero-Knowledge proof (ZK-proof) [11] is an agreement that makes the verifier believe that a certain assertion is correct when the prover does not provide any valuable information for the verifier. The scheme of this article adopts non-interactive proof. The zk-SNARK [12] is a kind of ZK-proof, which we used in the proposed solution. The zk-SNARK can provide the corresponding proof for the calculation that generates a specific output, the speed of verifying the proof is much faster than the speed of performing the corresponding calculation, the algorithm can enable the smart contract to confirm the legality of the transaction when the transaction information is ciphertext. The specific process will be introduced in Section 3.2.

2.4. Related Work

According to the requirements of GDPR, the study [13] proposed a model, through the use of smart contracts to enhance the transparency and source tracking of data usage, a blockchain-based personal data accountability system was designed, but the solution is limited to smart contracts. The study [14] designed and implemented a privacy authentication mechanism based on GDPR and further discussed legal issues regarding blockchain technology, while there is no specific design in terms of data sharing. The study [15] implemented a blockchain-based personal data management model based on the Hyperledger Fabric framework. The model interacts with resource servers, service providers, data processors, and data subjects to ensure the configuration file data stored in the server. However, it neither introduced regulatory authorities, nor was designed for privacy protection. The study [16] discussed several frameworks, methods, and architectures that can be used to meet the GDPR requirements to withdraw consent and permanently delete widely distributed personal data. However, these methods are just examples that can be used to implement the Right to be Forgotten (RtbF) in a digital ecosystem.

The study [17] proposed ADvoCATE, a user-centered framework to deal with GDPR requirements, but it only addresses the main needs of user privacy in the IoT ecosystem. Moreover, there is no regulatory authority in ADvoCATE. In the study [18], the authors argued that the existing consent mechanisms in online social networks did not conform to GDPR, and advocated a consent management approach based on blockchain, they focused on transparency and privacy. The study [19] discussed the design of a blockchain system towards GDPR compliance in the healthcare sector, proving that blockchain could enhance GDPR in some aspects. The study [20] designed a Blockchain-Based Personally Identifiable Information Management System (BcPIIMS), which could reduce the risk of personally identifiable information leaking, but only theoretically introducing the system.

In the field of data sharing, researchers have proposed many advanced block-chain-based methods for data privacy protection. In the study [21], authors designed a Smart Factory Big Data (SFBD) sharing system for secure storage and access control in the Industrial Internet of Things (IIoT). It uses identity-based encryption to provide access control while providing confidentiality, and supported user identity-based revocation. In our scheme, the symmetric encryption we used can only ensure confidentiality. Compared with the SFBD sharing system, we use zk-SNARK to provide perfect zero-knowledge, which is the advantage of our solution. The study [22] proposed PrivySharing to ensure the confidentiality of personal/critical user data, which is a framework for secure data sharing in smart cities. The multi-channel blockchain used by PrivySharing is more scalable

than the consortium blockchain in our scheme. Unlike our reward method that relies on reputation incentives, it provides the token 'PrivyCoin' for rewards. The CEM-PoA consensus algorithm we proposed in this article can provide data integrity, which is not possible with the Kafka algorithm used in PrivySharing.

Through the summary of related works under the GDPR, it can be found that the application of blockchain in GDPR compliance scenarios has many advantages. However, there are still some problems with data sharing under the supervision of GDPR through blockchain. On the premise of ensuring availability, it is also necessary to ensure confidentiality, regulatory traceability, privacy, integrity, and mutual authentication between entities. These problems are exactly what this article is dedicated to solving. Compared with the related works, our work proposes a new privacy and security architecture for data sharing based on a consortium blockchain. The proposed scheme uses Zero-Knowledge proof to protect the privacy of personal data, which has not been considered in previous works. We designed the CEM-PoA algorithm, which introduces regulators into the blockchain, it can ensure data integrity and implement regulatory-traceability. The reputation-based incentive mechanism we used can limit malicious behavior of nodes to ensure mutual authentication. A detailed comparison is given in Section 5.1.

3. Scheme Design

3.1. Data Sharing Scheme

This scheme employs the consortium blockchain, ZK-proof algorithm, CEM-PoA, and reputation incentive mechanism to make the data sharing process between users and service providers more secure. The system model is exhibited in Figure 1. Participating entities include data subjects, data controllers, data processors, regulatory authorities, consortium blockchain, smart contract, and private key generation center. The specific description of each entity is presented as follows.

Data Subject: Users who have personal data allow the data controller to collect data that can meet the requirements when requesting services and verify it through smart contracts.

Data Controller: The data controller is the user's service provider in GDPR, who collects the personal data of the data subject according to the business, and releases data requirements, guaranteeing the validity and availability of data through smart contracts.

Data Processor: An entity that provides users with data analysis services, but must depend on the data controller's infrastructure to develop the service and acquire the desired personal data.

Regulatory Authority: They can trace the source of data sharing compliance through the blockchain. The Regulatory authority is registered as a regulatory node on the consortium blockchain and used to form a committee in the CEM-PoA in order to verify the generated blocks.

Consortium Blockchain: The consortium blockchain is the basis of our solution. Since the blockchain is tamper-proof and traceable, multiple certified entities participate in the management of the blockchain.

Smart Contract: Smart contracts in the proposed scheme specify the format, size, and content of the data to be collected by the controller or data processor. Then, the validity of the data in the absence of a trusted organization can be automatically determined.

Private Key Generation Center: As a completely trusted entity, the private key generation center is responsible for generating keys and system parameters, as well as distributing public and private keys to data subjects, data controllers, and data processors. The private key generation center is completely credible, and it will not perform illegal operations.

All parties will register as different nodes in the blockchain in accordance with their different identities. If the data subject wants to use the service, he will encrypt his personal data in the format required by the data controller and submit the ZK-proof π. The smart contract will verify the validity of π, which will determine the availability of data.

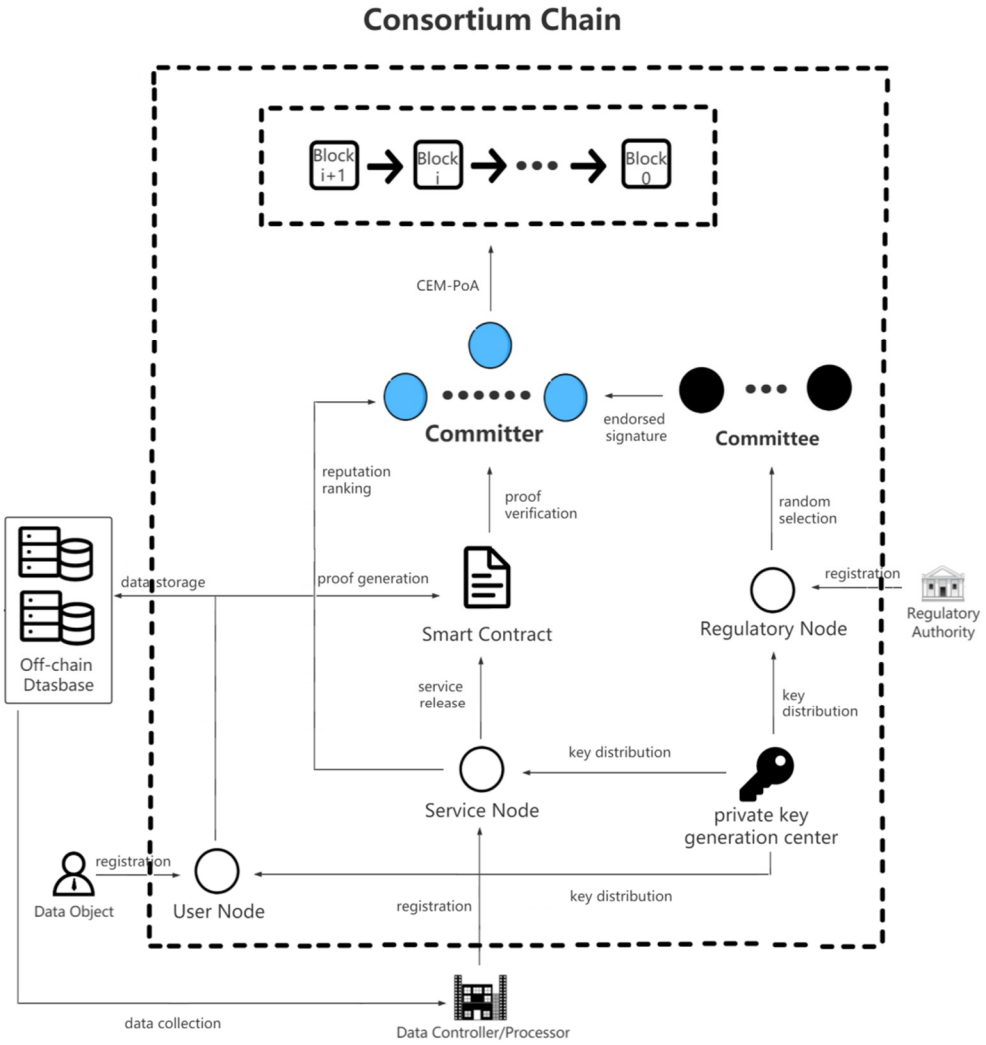

Figure 1. Consortium chain-based data sharing scheme for GDPR-compliance.

After the successful verification of π, the smart contract notifies the data controller. The encrypted data will be stored in the off-chain database, and the decryption key will be transmitted to the data processor. The data sharing process between the data controller and the data processor is similar to the above process. All transactions will be distributed through CEM-PoA for submitting blocks and will be recorded in the consortium chain.

The reputation incentive mechanism will reward the honest Committer. Otherwise, it will be punished. Interactions of all parties and violations of service providers will be recorded in a constant ledger, which can facilitate the regulatory authorities to conduct GDPR compliance checks and traceability. Regulatory authorities can track GDPR compliance through the consortium blockchain, and punish service providers in a timely manner in accordance without the need to collect evidence and information.

The main reasons we propose this architecture are as follows. Blockchain is one of the effective solutions to solve the problem of data sharing [15], it has the characteristics of decentralization, traceability, immutability, etc. According to the requirements of GDPR, we decided to use the consortium blockchain. First of all, there are multiple registration authentication entities such as regulators and service providers, so it is more appropriate to use a PoA consensus mechanism, we designed the CEM-PoA for this purpose. Secondly, for efficiency and applicable considerations, we used reputation incentives to prevent any unauthorized or malicious transactions initiated by them. At last, only compliance with the GDPR is not enough, it is also needed to ensure the privacy of personal data, so we use

zk-SNARK to verify through the Zero-Knowledge proof before sharing without revealing any privacy.

As indicated in Figure 2, the implementation of the proposed scheme can be divided into the following seven phases, including (1) Initialization; (2) Build the Network; (3) Service Released; (4) Data Storage; (5) ZK-proof Generation and Verification; (6) Node Consensus; and (7) Work Incentive. The specific content will be introduced in Section 4. The notations used in this paper can be found in Table A1 of Appendix A.

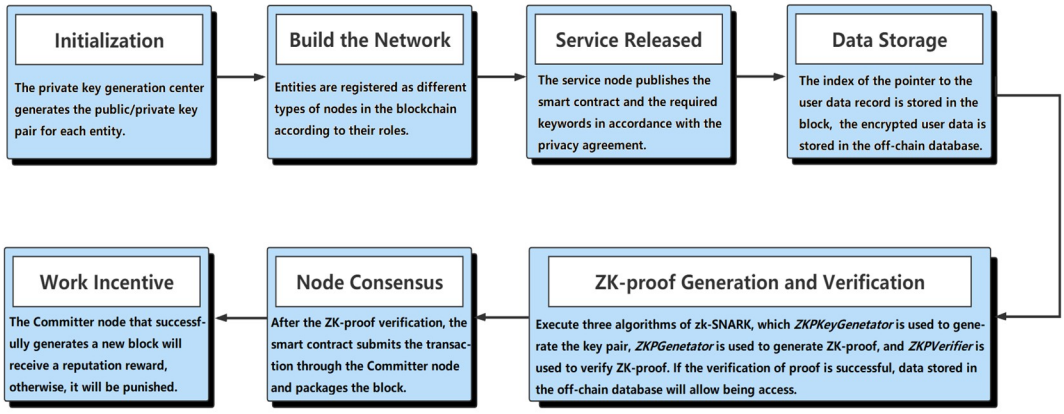

Figure 2. Seven phases in the proposed scheme.

3.2. On-Chain Privacy

Identity Verification: Identity verification mechanisms are of great importance since they are directly in association with the security and privacy of data [23]. In the proposed scheme, asymmetrically encrypted public keys or public key hashes are used to uniquely identify individuals in the blockchain, providing it with pseudo-anonymity. Key pairs are generated through *KeyGen*.

zk-SNARK: Let $C : F^n \times F^h \rightarrow F^l$ be an arithmetic circuit, and $R_C = \{(\vec{x}, \vec{w})\} \subseteq F^n \times F^h$ be the corresponding circuit satisfaction relation, where $\vec{x} \in F^n$ is called the statement and $\vec{x} \in F^h$ is the witness. A zk-SNARK for circuit satisfiability is a triple of polynomial-time algorithms $(ZKPKeyGenerator, ZKPGenerator, ZKPVerifier)$. They will be defined in Section 4.5. A zk-SNARK satisfies the necessary properties including completeness, soundness, and perfect zero-knowledge:

- **Completeness:** Given $(\vec{x}, \vec{w})$, the prover P can produce a proof π, and thus the verifier V accepts $(\vec{x}, \pi)$ with probability 1.
- **Soundness:** No Probabilistic Polynomial Time (PPT) adversary can generate a proof π for $\vec{x}$ that fools the verifier V to accept $(\vec{x}, \pi)$.
- **Perfect Zero-knowledge**: There exists a (randomized) polynomial simulator S. Therefore, for any $\vec{x}$,$S(x)$ generates a proof that is computationally indistinguishable from an honestly generated one.

3.3. CEM-PoA

Compared with proof of work, proof of authority does not involve mining mechanisms [24]. It is an energy-saving consensus algorithm with higher throughput and fast transaction processing. We designed CEM-PoA, introduced a committee endorsement mechanism, and added supervisory nodes to ensure the correctness of the generated blocks. Figure 3 shows the CEM-PoA architecture:

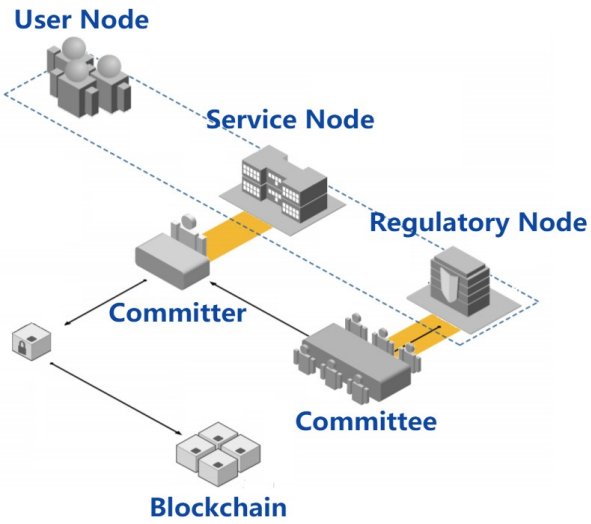

Figure 3. CEM-PoA architecture.

As a consortium blockchain, the Committer used to generate blocks will be selected from the service nodes. All service providers are supposed to disclose their true identities, and the blockchain dynamically maintains a set of rankings based on identity credibility. One of the service nodes with the highest reputation ranking will be randomly selected as a Committer for several rounds. The essence behind the reputation mechanism is the certainty of the identity of the service node. The work of the Committer involves collecting transactions in the network, verifying the transaction, and packaging the transaction into a block. The block is added to its local blockchain after passing the endorsement.

Additionally, the committee endorsement mechanism is introduced. After the verification of a new block, several nodes will be randomly selected from the regulatory nodes to form a committee to sign the submitted block as an endorsement. It is required that the Committer must collect enough regulatory node endorsements before adding relevant information to the new block to be generated. Moreover, if the Committer wants to destroy the blockchain in the current round of consensus, he must unite a certain number of regulatory nodes to do evil together:

- The Committer selected from the service node must be complicit with the regulatory node.
- Among these conspiracy nodes, enough nodes must be randomly selected for endorsement.
- The above two conditions make it challenging for the Committer to utilize his rights in order to generate different blocks.

3.4. Evaluation of Node Reputation

The incentive mechanism aims to motivate the nodes abided by the rules, participate in the blockchain work, and punish the nodes that do not follow the rules. As a result, the entire network can develop in the direction of a virtuous circle since the security and availability of the blockchain require the participation of most honest nodes. In the consortium blockchain, to prevent the service node from doing nothing or doing evil, our scheme adopts the identity credibility as an incentive mechanism, which will reward the nodes abided by the rules and participate in the authentication, and punish the nodes that do not follow the rules. The specific factor thresholds in the reputation evaluation mechanism vary with the size of the company, so there is no clear definition. The method of calculating the initial reputation of a service node is presented as follows:

- Server Factor S_1: S_1 includes the processor, memory, service speed, and operating performance, etc.
- Network Factor S_2: S_2 includes blocking the size of the internal network, server band-width, network latency, and server vulnerability, etc.

- Reputation Factor S_3: S_3 includes company size, revenue, user size, whether having been punished or not recently, and compliance history, etc.
- Assign a different weight W_{ij} to each S_{ij}, and then calculate the weighted average as the node reputation, giving:

$$R = \frac{(\sum_{i=1}^{I} S_{1i} W_{1i} + \sum_{j=1}^{J} S_{1j} W_{1j} + \sum_{k=1}^{K} S_{1k} W_{1k})}{I + J + K} \quad (1)$$

4. Specific Implementation

4.1. Initialization

The private key generation center sets safety parameter λ and random number r, the user provides ID_u as his unique identifier to the private key generation center, which will use $KeyGen(ID_u) \rightarrow (PK_u, SK_u)$ to generate the public/private key pair (PK_u, SK_u) of the data subject. Service providers and regulatory agencies can obtain key pairs separately through the same process.

4.2. Build the Network

Multiple entities, including data subjects, data controllers, and data processors, are registered as consortium chain nodes. Blockchain will assign a blockchain address based on participating roles and grant different permissions. As displayed in Figure 4, nodes are divided into light nodes and full nodes: Light nodes are all ordinary nodes, and full nodes include ordinary nodes, service nodes, and regulatory nodes. The data subjects are all user nodes (nodes that do not participate in commit and endorsement).

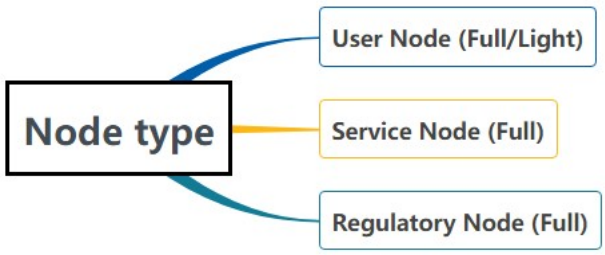

Figure 4. Types of nodes in the scheme.

Light Node: Light nodes only store block headers. By storing block headers, it can be demonstrated that certain transactions have not been changed without using most of the memory to store the blockchain. Additionally, light nodes can also access the specific data they want. For the purpose of saving storage resources, the data subject can choose to become a light node.

Full Node: Full nodes need to save every transaction and block in the blockchain to store complete blockchain information. User nodes can also apply to become full nodes. The service node is composed of service providers (that is, data controllers and data processors) that comply with GDPR. The Committer is selected to verify the block generation, while the regulatory node is composed of regulatory authorities (such as the Data Protection Agency) that operates and endorses the blocks generated by the Committer.

4.3. Service Released

According to the data requirements of data controller's privacy agreement, the data subject generates ZK-proof π. We take the data subject's proof π generation and verification stage as an example. The similar process of the data controller's proof π' is not described in detail. The data controller's ZK-proof and related parameters will be stored in the consortium blockchain. Simultaneously, the smart contract will publish some keywords required by the service provider.

4.4. Data Storage

We store indexes containing pointers to user's data records in the consortium blockchain, it can be observed from Figure 5 that a data block consists of a block header and a block body. The endorsement signature of the block represents the regulatory node endorsing the Committer.

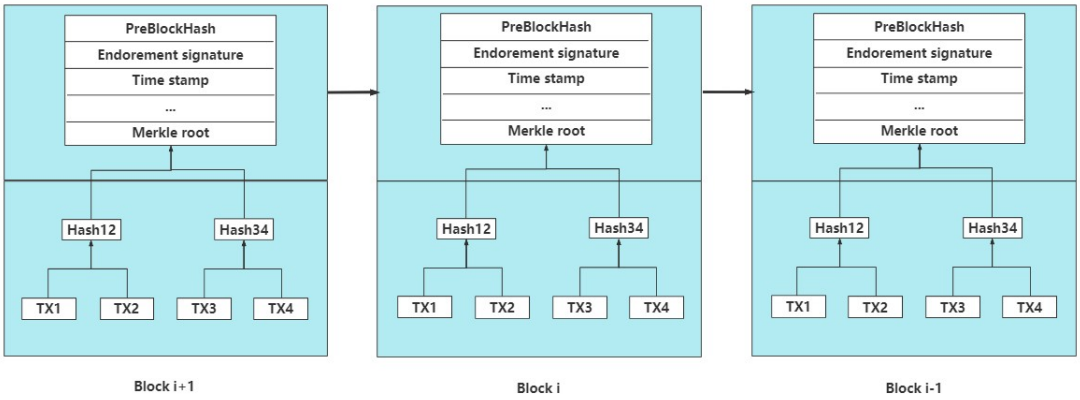

Figure 5. Block structure and transaction form.

If the committee verifies the transaction, it will be recorded on the blockchain. The personal data of the data subject will be encrypted by PK_u before being stored in the off-chain database rather than being directly stored on the blockchain since storing data on the blockchain will not be capable of modifying or deleting the data, which may violate the right of modification and forgetting in GDPR and generate a potential privacy leak.

4.5. ZK-Proof Generation and Verification

As shown in Figure 6, the circuit C takes the PK_u, the users' private data $D=<d_1, d_2, \ldots, d_n>$ and $< ID_u,\ T >$ as input, where the ID_u is node's identification, T is the timestamp, and r is the random number. A result R and a hash value h can be the output from the circuit to verify the data.

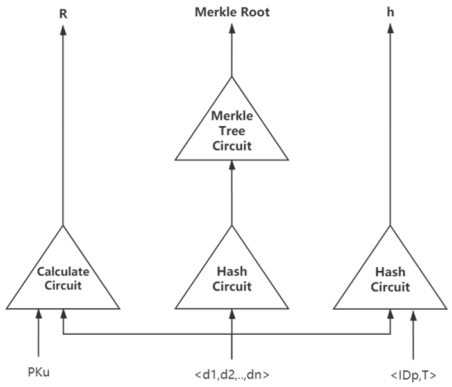

Figure 6. Circuit structure of our zk-SNARK algorithm.

The security parameter λ and the circuit C will be taken as input parameters in order to calculate the key pair (EK_c, VK_c), where the proving key EK_c is used to generate a ZK-proof, and verification key VK_c is used to verify the ZK-proof.

$$ZKPKeyGenerator(1^{\lambda},\ C) \rightarrow (EK_c, VK_c) \tag{2}$$

The smart contract will receive the ZK-proof π which is sent by a data subject. Taking the proving key EK_c, the result R, the hash value h, and data D as input, we get the proof π:

$$ZKPGenerator(EK_c,\ D,\ R,\ h) \rightarrow\ \pi \tag{3}$$

The smart contract will verify whether π is consistent with the data controller's requirements. If the verification is completed, it will output rejection (false) or acceptance (true) results:

$$ZKPVerifier(VKc,\ PK_u, \pi,\ R,\ h,) \rightarrow (true\ or\ false) \tag{4}$$

If the verification of proof is successful, the data subject's data, which is stored in the off-chain database will allow access by the data controller, otherwise the transaction fails.

4.6. Node Consensus

After the data controller decrypts and obtains the data of the data subject, the smart contract will submit the transaction through the Committer and package the block. As shown in Figure 7, all rounds included in each period of CEM-PoA are displayed as follows:

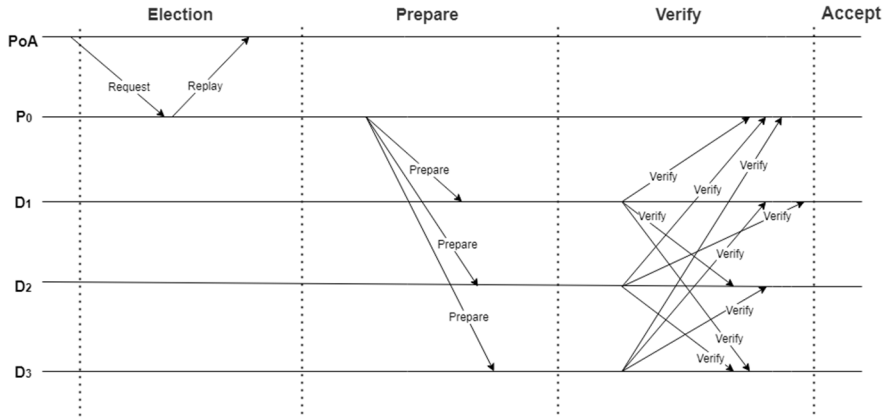

Figure 7. CEM-PoA work period.

4.6.1. Election Round

According to the current k-th round of reputation ranking list, CEM-PoA randomly sends task O and timestamp T in $Request(k,\ O,\ T)$ to the top 3% service nodes (including P_0). If P_0 returns the result $Reply(k,\ T,\ O,\ P_0)$, it will become the Committer of this round of generating blocks. After being selected, the service node was elected to the Committer for several rounds. In addition, the CEM-PoA will reselect the Committer and resend the request in case of timeout if the selected service node does not respond.

4.6.2. Prepare Round

After being elected as the Committer, P_0 starts to collect transaction information $Tx = \{Tx_1, \ldots, Tx_n\}$ in consortium chain and package it into block B_k, and then encapsulates it as $Prepare(P_0,\ TX,\ Sign_{P_0}(Request))$. $Sign_{P_0}(Request)$ is the Committer's digital signature of $Request(k,\ O,\ T)$, functioning as a proof of being elected Committer.

4.6.3. Verify Round

CEM-PoA selects $D_i \mid \mathrm{i} = 1, 2, 3, \ldots, n$ from the regulatory nodes to form a committee. The Committer sends *Prepare* to the committee members for verification. After the verification, D_i will send $Verify(P_0,\ E_{P_0}(Sign_{D_i}(Prepare)))$, where ${}_{P_0}(Sign_{D_i}(Prepare))$ is the D_i to *Prepare*'s endorsement signature, which is encrypted with P_0's public key. After P_0 collects m committee signatures (for example, $3 \leq m \leq n$), it packs the endorsement signature into block B_k.

4.6.4. Accept Round

P_0 obtains the reputation reward r_z, and broadcasts B_k to the nodes of the entire network in order to enter the next round. Conditions for receiving new blocks on the blockchain:

- The new block is generated by the current Committer.
- Currently, the Committer has no other blocks generated.
- The block has been endorsed.
- The block has been generated and signed correctly.

4.7. Work Incentive

The reword reputation score r_z of Committer P_i is calculated as:

$$r_z = \sum_{e=1}^{E} S(z)T(z), \tag{5}$$

where E represents the number of data records contained in this block, $T(z)$ is the proof size executed by the smart contract to verify the record z, the factor $S(z)$ responds to the behaviour of different reputations and is employed to perform different degrees on the node's rewards or punishments.

When users select a service, they can check the service provider's identity and credibility at any time. Undertaking the work of verifying nodes will be rewarded and improve the service provider's identity and credibility. The rules are presented as follows:

- If the Committer successfully produces a block and finally confirms the block, the node will receive reputation rewards.
- If the Committer fails to generate blocks within a certain period or fails to generate the expected number of blocks, it will be punished.

5. Analysis and Discussion

5.1. Comparative Analysis

We make a comparison between our scheme and some other related solutions, the result in Table 1 shows that our scheme is more comprehensive:

Table 1. Qualitative comparisons of the proposed scheme with the existing works (Support ✓, Non-Support ×).

Features / Approaches	Confidentiality	Regulatory-Traceability	Privacy-Preservation	Availability	Integrity	Mutual Authentication
Ours	✓	✓	✓	✓	✓	✓
Truong et al. [15]	✓	×	×	✓	×	×
Rantos et al. [17]	×	✓	×	✓	×	×
Hasselgren et al. [19]	×	✓	✓	✓	✓	×

Confidentiality: The security of the data on the blockchain in our solution is mainly based on asymmetric encryption. As long as the system parameter is large enough, the security of blockchain can be guaranteed. The asymmetric encryption algorithm is statistically secure in the proposed scheme. If the corrupt node or the attacker does not have the decryption key, the data submitted to the blockchain cannot be decrypted. The study [17] proposed a GDPR-compliant framework ADVoCATE, its design covers most of the rights in the GDPR, but the framework does not introduce customized encryption measures for shared data, so data confidentiality cannot be guaranteed.

Regulatory-Traceability: Compared with traditional solutions, all data recorded in blockchain are traceable and cannot be changed once written. Thus, GDPR compliance investigations can be conducted. Non-compliance activities will cause regulators to conduct formal investigations and audits of service providers. In the study [15], the authors have not introduced regulatory authorities, their solution is unable to supervise service nodes, the timely regulatory traceability of the blockchain cannot be guaranteed. ADVoCATE in the study [17] is mainly oriented to IoT, which introduces regulatory authorities in the entity to ensure regulatory traceability.

Privacy-Preservation: Blockchain will assign a unique pseudonym to every participating entity, and thus the pseudo-identity will be used in the subsequent process without revealing the real identity. In the process of data sharing, any entity participating in the interaction with the smart contract will not disclose data privacy. Smart contracts can only receive the ZK-proof rather than the privacy data of the data subject. In addition, the service provider will only publish some keywords, not the entire requirement, which guarantees data privacy protection. In the study [15], the authors only adopt digital signatures to prove the node submitting the block. In addition, their data does not adopt ZK-proofs when interacting with smart contracts, which may be the source of privacy leaks. Although ADVoCATE [17] uses an intelligent policy analysis mechanism to support the so-called user privacy policy, it does not take measures other than encryption for privacy protection during user data interaction. The GDPR-compliant blockchain framework for healthcare in the study [19] achieves partial privacy-preservation by enabling patients to anonymize their names and identities.

Availability: In the proposed scheme, only authorized entities can decrypt the specific data of the data subject. Data subjects can generate proof by private data which follows the data controller's requirement. Proof can be sent to the smart contract to determine if the user's data meets the service provider's established requirements or not, without revealing the user's privacy. This feature ensures data availability. The availability of data is the prerequisite for all other features, all the schemes proposed in the comparison work have achieved availability in their designs.

Integrity: After the successful verification, the transaction between the user and the service provider will be published on the blockchain. The committee composed of supervisory nodes will endorse the new block to ensure that transaction records will not be tampered with during the block generation process. CEM-PoA ensures data integrity. The consensus protocol used in [15] does not guarantee that internal opponents cannot tamper with the transaction, nor does it use an incentive mechanism, it cannot guarantee that those dishonest nodes will be affected by punishment. In the study [17], there is no specific design for ensuring the integrity of transaction data. The proposed scheme in [19] concentrates on achieving data integrity by separating the identity of each patient from the provider through anonymous metadata.

Mutual Authentication: Each party in the consortium blockchain is able to confirm the legal identity of the other party in the network. We use reputation ranking and introduce regulatory authorities to achieve mutual authentication. All service nodes are authenticated before entering the blockchain, the private key generation center ensures that each node in the network has a unique and verifiable identity. Additionally, their reputation behavior is recorded in real-time by ranking. The ADVoCATE in [17] does not authenticate service providers or authorized third parties. The research work in [19] does not use any authentication mechanism. Therefore, mutual authentication between on-chain entities cannot be guaranteed.

As can be seen from Table 1, our solution is more suitable for GDPR. The comparison shows that the scheme proposed in this paper can meet multiple security indicators of GDPR-compliance data sharing, which is exactly the goal of our research.

5.2. Performance Evaluation

In this paper, we propose a blockchain-based scheme, we implemented the prototype of the proposed scheme based on python. Transaction records between users and service providers are stored in blocks. The zk-SNARK algorithm is implemented by the PySNARK library. The experiments were performed in the following environment: (i) Intel Core i7-10700K CPU (3.8GHz); (ii) 256GB RAM; (iii) Ubuntu Linux 20.04.1 Server (64-bit); and (iv) Python 3.7.0. Initially, we conducted a comparative experiment on the block generation time between the proposed scheme and others.

Figure 8 represents the experimental result of our scheme compared with the existing works, our solution has the best performance among the three schemes. In the experiment,

performance is affected by the consensus mechanism. The study [15] applies Proof of Work (PoW) in their system, nodes obtain the right to create a new block through competition and mining. PoW requires high computing resources and equipment. The Practical Byzantine Fault Tolerance (PBFT) algorithm is used in [17], which is suitable for permitted systems, but its communication complexity is relatively high. The authors in [19] use the Proof of Stack (PoS) as their consensus mechanism. PoS is hosted by the node with the most tokens, and uses a random process to determine which node will generate the next block.

The CEM-PoA consensus proposed in our scheme is designed for consortium blockchain, it does not require nodes to spend computing resources to solve complex mathematical tasks, but relies on reliable nodes with a high reputation to generate blocks. It can be seen from the result in Figure 8 that the growth rate of the proposed scheme is more stable and faster, which has a 41% block generation growth rate. Therefore, compared with PoW, PBFT, and PoS, the proposed CEM-PoA is highly efficient. Additionally, the efficiency of the consortium blockchain is higher than the public blockchain in most instances.

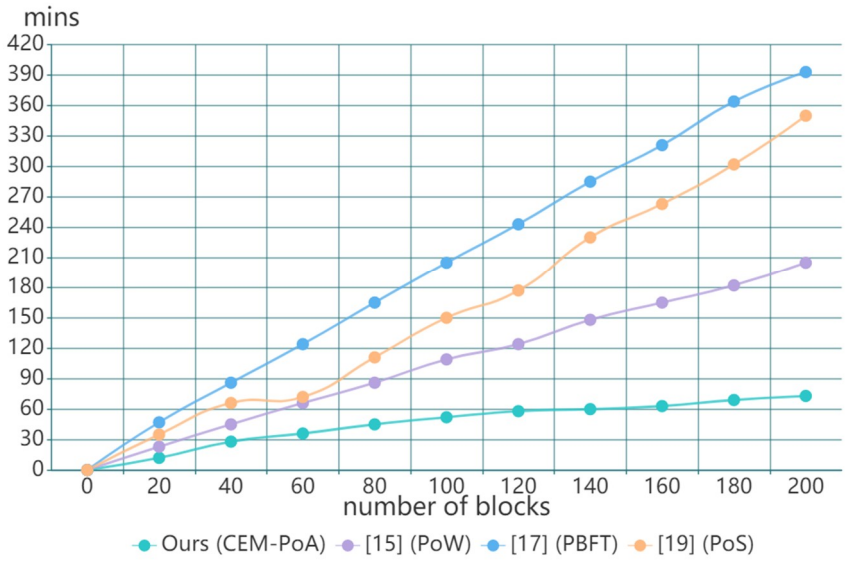

Figure 8. Comparison of block generation time.

In addition to the block generation time of the scheme, we also focus on the time consume related to zk-SNARK. We simulated nine phases of the proposed scheme in the same environment, and mainly evaluated the time to generate ZK-proof key pairs, ZK-proof generation, and verification time under different inputs. Each experiment was repeated 200 times and their average value was calculated. The challenge of the proposed scheme is ZK-proof's time consumption. We set the security parameter λ with a security level of 128 bits. The number of inputs is set to 200, 400, 600, 800, and 1000, respectively.

The following findings show the efficiency and feasibility of our zk-SNARK algorithm. Figure 9 illustrates the generation time of a ZK-proof key pair under different inputs. It shows that the time to generate a ZK-proof key pair (EK_c, VK_c) is about 16.5 s. Regardless of the number of inputs, the generation time is almost constant. Even the number of circuit inputs is 1000, the difference almost remains within milliseconds, which is an acceptable time constraint for overall performance. The ZK-proof key pair generation only needs to be executed once, so the related costs are not a problem for the whole blockchain.

Figure 10 illustrates the generation time of ZK-proofs under different inputs. The proof generation is only run by a part of the nodes, not the whole blockchain system. Therefore, it will not significantly affect the overall performance of the system. The generation proof time only fluctuates slightly and has no apparent relationship with the number of inputs, the experimental results remain within acceptable limits.

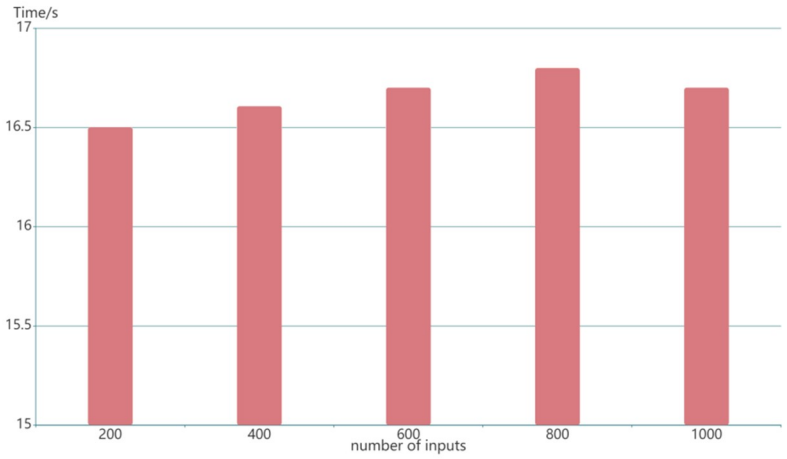

Figure 9. The time to generate a ZK-proof key pair.

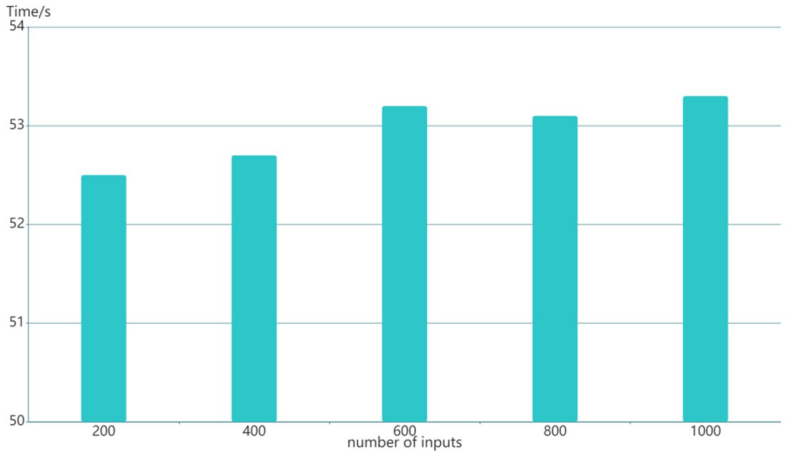

Figure 10. The time to generate a proof.

We provide the experiment results of ZK-proof verification in Figure 11. The ZK-proof verification is run by every smart contract in the blockchain, and it can be found that the time of verification increases linearly as the input increases, but the whole is less than 1 s. The time fluctuation is within milliseconds, and thus the result still can be considered feasible. These findings show that smart contracts can easily verify the validity of the relevant transactions posted in the blockchain.

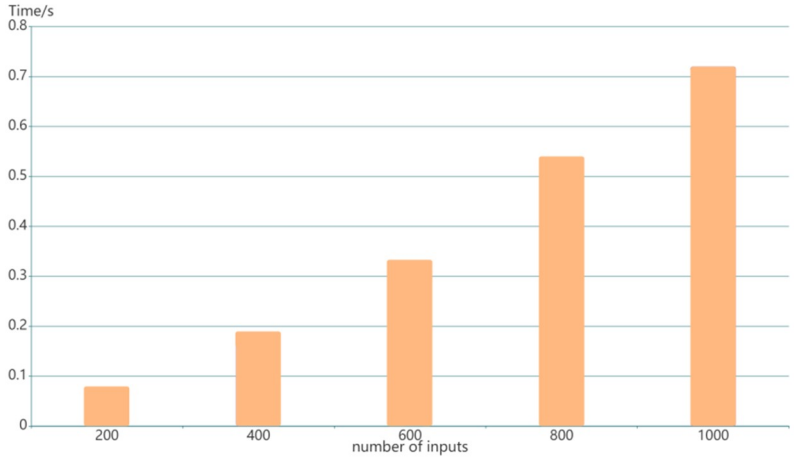

Figure 11. The time to verify a proof.

Finally, we evaluate the zk-SNARK circuit's cost size. We measure the size of EK_c, VK_c, and the proof. As shown in Table 2, the EK_c has a constant size (14.7 MB), the VK_c size increases from 8.6 to 40.9 KB, and the proof size increases from 33.5 to 160.5 KB. The size of VK_c increases directly and is caused by the increase of inputs, which results in that the proof size gradually rises with the increase in the size of VK_c. The result remains within an acceptable range, these findings demonstrate that the cost of the zk-SNARK circuit is small enough to ensure the efficiency of our scheme.

Table 2. The size of EK_c, VK_c, and the proof.

Inputs	EK_c Size	VK_c Size	Proof Size
200	14.7 MB	8.6 KB	33.5 KB
400	14.7 MB	17.2 KB	66.7 KB
600	14.7 MB	24.1 KB	92.3 KB
800	14.7 MB	32.5 KB	127.1 KB
1000	14.7 MB	40.9 KB	160.5 KB

5.3. Limitation

The proposed scheme is only designed and tested for general personal data sharing scenarios, while different data also have different transmission requirements [25]. For example, there are special standards for health care and crime data in GDPR, which is not within the scope of our research.

Although there are regulatory agencies, the consortium blockchain cannot guarantee complete decentralization, posing a certain threat to user data and records. Additionally, the concept of our scheme is not the definitive solution for GDPR personal data sharing. For example, an attacker can find the decryption key through nefarious means or by discovering the limitations of the encryption technology itself, which is the future research direction of blockchain-based solutions. Moreover, for the other data protection regulation such as Data Security Law of the People's Republic of China (DSL), our scheme is not completely applicable.

6. Conclusions and Future Work

When applying blockchain to GDPR compliance, how to ensure as much as possible the security and privacy protection in data sharing is a challenging issue. Our research designed the first data sharing solution using ZK-proof in the combination of technologies such as consortium blockchain, CEM-PoA, and zk-SNARK. The analysis shows that confidentiality, availability, and other indicators are met in the proposed scheme. Compared with the existing GDPR blockchain solution, the proposed solution has an efficiency advantage in block generation speed while achieving ideal privacy protection. Precisely, by automatically verifying the ZK-proof, the smart contract can determine whether the user's data meets the service provider's established requirements or not, without revealing the user's privacy. After verification, the transaction is recorded in the blockchain through CEM-POA, and the reputation incentive mechanism can ensure the honesty of the node. Any violations of regulations will be immutably recorded in the blockchain, and can be easily detected by regulatory agencies.

The proposed scheme can be applied to various data transactions or sharing scenes with privacy protection and compliance requirements. For example, in the medical area, this solution can share medical data among patients, hospitals, and research institutions while protecting patient data from being leaked. Additionally, it can provide a privacy-preserving solution from technical design for contact tracing for the COVID-19 pandemic. In the Internet of Vehicles (IoV) area, our scheme can solve the anonymity and authentication problems, such as traffic services or location services in vehicular social networks.

In addition to the technology used in our scheme, there are numerous technologies that can ensure the security and privacy of the blockchain. For instance, Differential Privacy (DP) [26] technology can be used to scramble data before sharing it or adding noise in

the calculation process. Fully Homomorphic Encryption (FHE) [27] technology is a major breakthrough in cryptography, which implements various operations requested by users without decrypting data. However, the overhead must be considered.

The contribution of this article explores how the blockchain-based solution can better comply with the GDPR regulations. Our new GDPR-compliance consortium blockchain scheme supports many features, such as regulatory-traceability, mutual authentication, and others, all defects that exist in previous works are eliminated. Our experimental results show the practicality of our scheme. To put it in a nutshell, we leave the following questions for future research:

- Further optimizing the security and privacy of the proposed scheme based on the above technologies.
- Adding more functions to better comply with GDPR requirements.
- Exploring how to improve the efficiency of generation and verification of zero- knowledge proof.

Author Contributions: Writing—review and editing, Y.P.; visualization, K.Y.; supervision, X.C.; project administration, Y.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Key Research and Development Program of China (grant no. 2018YFC1604000) and Fundamental Research Funds for the Central Universities (grant no. 2042017gf0035).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

Appendix A

Table A1. Nomenclature.

Notation	Description
π	Zero-Knowledge proof
ID_i	User unique identifier
PK_u	Public key of u
SK_u	Private key of u
λ	Security parameter
C	The circuit of the zk-SNARK
EK_c	The key used to generate the zero-knowledge proof
VK_c	The key used to verify the zero-knowledge proof
h	The hash value output by the circuit
R	The result output by the circuit
T	The timestamp used for circuit input
r	The random number used for circuit input
P_i	The i-th service node
D_i	The i-th regulatory node

References

1. Lee, G.Y.; Cha, K.J.; Kim, H.J. Designing the GDPR Compliant Consent Procedure for Personal Information Collection in the IoT Environment. In Proceedings of the 2019 IEEE International Congress on Internet of Things (ICIOT), Milan, Italy, 8–13 July 2019; pp. 79–81.
2. Farshid, S.; Reitz, A.; Roßbach, P. Design of a Forgetting Blockchain: A Possible Way to Accomplish GDPR Compatibility. In Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS), Hawaii, UK, 8 January 2019.
3. Martin, Y.-S.; Kung, A. Methods and Tools for GDPR Compliance Through Privacy and Data Protection Engineering. In Proceedings of the 2018 IEEE European Symposium on Security and Privacy Workshops (Euro S&P), London, UK, 24–26 April 2018; pp. 108–111.
4. Aridor, G.; Che, Y.K.; Salz, T. *The Economic Consequences of Data Privacy Regulation: Empirical Evidence from GDPR*; National Bureau of Economic Research: Cambridge, MA, USA, 2020; p. 26900.

5. Politou, E.; Alepis, E.; Patsakis, C. Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions. *J. Cybersecur.* **2018**, *4*, tyy001. [CrossRef]
6. Furey, E.; Blue, J. Can I Trust Her? Intelligent Personal Assistants and GDPR. In Proceedings of the 2019 International Sym-posium on Networks, Computers and Communications (ISNCC), Istanbul, Turkey, 18–20 June 2019; pp. 1–6.
7. Pandit, H.J.; O'Sullivan, D.; Lewis, D. Exploring GDPR Compliance Over Provenance Graphs Using SHACL. In Proceedings of the 14th International Conference on Semantic Systems (SEMANTiCS), Vienna, Austria, 10 September 2018.
8. Badii, C.; Bellini, P.; Difino, A.; Nesi, P. Smart City IoT Platform Respecting GDPR Privacy and Security Aspects. *IEEE Access* **2020**, *8*, 23601–23623. [CrossRef]
9. Nakamoto, S. Bitcoin: A Peer-to-Peer Electronic Cash System. Available online: https://bitcoin.org/bitcoin.pdf (accessed on 20 May 2021).
10. Fan, H.; Liu, Y.; Zeng, Z. Decentralized Privacy-Preserving Data Aggregation Scheme for Smart Grid Based on Blockchain. *Sensors* **2020**, *20*, 5282. [CrossRef]
11. Wang, Y.; Kogan, A. Designing confidentiality-preserving Blockchain-based transaction processing systems. *Int. J. Acc. Inf. Syst.* **2018**, *30*, 1–18. [CrossRef]
12. Ben Sasson, E.; Chiesa, A.; Garman, C.; Green, M.; Miers, I.; Tromer, E.; Virza, M. Zerocash: Decentralized Anonymous Payments from Bitcoin. In Proceedings of the 2014 IEEE Symposium on Security and Privacy, Berkeley, CA, USA, 18–21 May 2014; pp. 459–474. [CrossRef]
13. Zyskind, G.; Nathan, O.; Pentland, A. Decentralizing Privacy: Using Blockchain to Protect Personal Data. In Proceedings of the 2015 IEEE Security and Privacy Workshops, San Jose, CA, USA, 21–22 May 2015; pp. 180–184. [CrossRef]
14. Wirth, C.; Kolain, M. Privacy by blockchain design: A blockchain-enabled GDPR-compliant approach for handling personal data. In Proceedings of the 1st ERCIM Blockchain Workshop, Amsterdam, The Netherlands, 8–9 May 2018.
15. Truong, N.B.; Sun, K.; Lee, G.M.; Guo, Y. GDPR-Compliant Personal Data Management: A Blockchain-Based Solution. *IEEE Trans. Inf. Forensics Secur.* **2019**, *15*, 1746–1761. [CrossRef]
16. Camilo, J. Blockchain-based consent manager for GDPR compliance. In Proceedings of the Open Identity Summit, Garmisch-Partenkirchen, Germany, 28–29 March 2019.
17. Rantos, K.; Drosatos, G.; Demertzis, K.; Ilioudis, C.; Papanikolaou, A.; Kritsas, A. ADvoCATE: A Consent Management Platform for Personal Data Processing in the IoT Using Blockchain Technology. In Proceedings of the International Conference on Security for Information Technology and Communications, Bucharest, Romania, 8–9 November 2018; pp. 300–313.
18. Ahmed, J.; Yildirim, S.; Nowostaki, M.; Ramachandra, R.; Elezaj, O.; Abomohara, M. GDPR Compliant Consent Driven Data Protection in Online Social Networks: A Blockchain-Based Approach. In Proceedings of the 2020 3rd International Conference on Information and Computer Technologies (ICICT), San Jose, CA, USA, 9 March 2020; pp. 307–312.
19. Hasselgren, A.; Wan, P.K.; Horn, M.; Kralevska, K.; Gligoroski, D. GDPR Compliance for Blockchain Applications in Healthcare. *arXiv* **2020**, arXiv:2009.12913. Available online: https://arxiv.org/abs/2009.12913 (accessed on 20 May 2021).
20. Al-Zaben, N.; Onik, M.H.; Yang, J.; Lee, N.-Y.; Kim, C.-S. General Data Protection Regulation Complied Blockchain Architecture for Personally Identifiable Information Management. In Proceedings of the 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE), Southend, UK, 16 August 2018; pp. 77–82.
21. Yu, K.-P.; Tan, L.; Aloqaily, M.; Yang, H.; Jararweh, Y. Blockchain-Enhanced Data Sharing With Traceable and Direct Revocation in IIoT. *IEEE Trans. Ind. Inform.* **2021**, *17*, 7669–7678. [CrossRef]
22. Makhdoom, I.; Zhou, I.; Abolhasan, M.; Lipman, J.; Ni, W. PrivySharing: A blockchain-based framework for privacy-preserving and secure data sharing in smart cities. *Comput. Secur.* **2020**, *88*, 101653. [CrossRef]
23. Ramachandran, A.; Kantarcioglu, D. Using blockchain and smart contracts for secure data provenance management. *arXiv* **2017**, arXiv:1709.10000. Available online: https://arxiv.org/abs/1709.10000 (accessed on 20 May 2021).
24. Brogan, J.; Baskaran, I.; Ramachandran, N. Authenticating Health Activity Data Using Distributed Ledger Technologies. *Comput. Struct. Biotechnol. J.* **2018**, *16*, 257–266. [CrossRef] [PubMed]
25. Benhamouda, F.; Halevi, S.; Halevi, T. Supporting private data on Hyperledger Fabric with secure multiparty computation. *IBM J. Res. Dev.* **2019**, *63*, 3–10. [CrossRef]
26. Dwork, C. Differential Privacy: A Survey of Results. In Proceedings of the International Conference on Theory and Applications of Models of Computation, Xi'an, China, 25–29 April 2008; pp. 1–19.
27. Zhou, L.; Wang, L.; Ai, T.; Sun, Y. BeeKeeper 2.0: Confidential Blockchain-Enabled IoT System with Fully Homomorphic Computation. *Sensors* **2018**, *18*, 3785. [CrossRef] [PubMed]