Towards on Blockchain Data Privacy Protection with Cryptography and Software Architecture Approach

Zexu Wang[1,3], Bin Wen[1,2,3(✉)], and Ziqiang Luo[1,2]

[1] School of Information Science and Technology, Hainan Normal University, Haikou, China
z.x.wang1060@qq.com, binwen@hainnu.edu.cn
[2] Cloud Computing and Big Data Research Center, Hainan Normal University, Haikou, China
[3] Key Laboratory of Data Science and Intelligence Education of Ministry of Education, Hainan Normal University, Haikou 571158, China

Abstract. The essence of blockchain is to solve trust problems and realize value transfer. The traditional centralized processing method adopts centralized transmission and storage of users' data privacy, which improves the security and reliability of processing. The non central blockchain technology uses distributed ledger technology to realize the characteristics of disintermediation, data tampering, traceability, work audit, etc. encryption algorithm is used to encrypt the data, and consensus mechanism makes the data sharing of blockchain system more fair and stable. The current passive data privacy schemes are basically based on cryptography. On the premise of satisfying the constraint mechanism of blockchain, this paper studies the technical framework, encryption mechanism and empirical analysis of blockchain, and discusses the information privacy protection method of hiding the original big data as much as possible, so as to improve the system performance and protect the security and privacy of user data. At the same time, we study data privacy protection from the perspective of software architecture, and propose a data privacy protection scheme through algorithm decomposition multi center collaboration method, which provides guidance for big data sharing and transaction based on blockchain.

Keywords: Blockchain technology architecture · Cryptography · Cryptographic hash function · Privacy protection · Software architecture

1 Introduction

Blockchain is a computing infrastructure to realize value transfer in the digital economy era, because of its features of decentralization and anonymity, the traditional transmission of various types of value has a subversive improvement

Z. Zheng et al. (Eds.): BlockSys 2020, CCIS 1267, pp. 205–217, 2020.
https://doi.org/10.1007/978-981-15-9213-3_16

[1,2]. As an integrated system with point-to-point network, cryptography, consensus mechanism, smart contract and other technologies, blockchain provides a trusted channel for information and value transfer and exchange in untrusted networks. With its unique trust building mechanism, blockchain technology has cross innovation with new technologies and applications such as cloud computing, big data, artificial intelligence, etc., and has integrated and evolved into a new generation of network infrastructure to reconstruct the industrial ecology of digital economy [3–5].

Blockchain is the result of the integration of cryptography, consensus mechanism, computer science and other disciplines. Bitcoin [6] is one of blockchain's most successful apps, but blockchain isn't just about issuing coins. Just as the real value of blockchain lies in the safer and more efficient transfer of value than the traditional [11,13]. Its system is composed of the storage layer and the network layer of the protocol layer, including all kinds of scripts,[1] and the intelligent contract extension layer, which encapsulates the application layer of all kinds of applications. All levels coordinate with each other to maintain the stability of the system [5].

The lack of data privacy is one of the main limitations of blockchain services. The current protection measures mainly include coin shuffle, ring signature, zero knowledge proof and so on.

Bitcoin is the first application of blockchain and also a very successful practice. As a kind of digital currency based on blockchain, it is precisely because the mechanism of block chain cryptography guarantees that the security of products based on blockchain technology such as bitcoin and litecoin has been greatly improved. The rest of the article is organized as follows. In the second 2, we will introduce the basic architecture of the blockchain and how to interact with each other to complete the daily work. Section 3 introduces how the digital currency represented by bitcoin can guarantee the security of the transaction through the cryptography mechanism in the blockchain. In the Sect. 4, we will verify and analyze the script examples to illustrate the significant advantages of the blockchain in decentralization. In the Sect. 5, the method of block chain privacy protection based on software architecture is proposed. Conclusions with main contributions of proposed approach and further work plans are also touched upon in Sect. 6.

2 Technical Architecture

2.1 Protocol Layer

This layer consists of a storage layer and a network layer. The storage layer is the lowest technology of the block chain, which mainly ensures the security and tradability of data while storing data. Data storage is mainly realized by using Markle tree [6], block chain storage and other data structures, and the high efficiency of data storage greatly determines the performance of the upper layer. Of course, there are certain requirements for programming ability, but the

[1] https://bitcoin.org/bitcoin.pdf.

logical structure of data storage can be achieved in most languages, such as Java, Python, GO, and so on.

The main task of the network layer is to realize the information exchange between users through distributed storage, asymmetric encryption, digital signature, multi-signature and other technologies in the point-to-point network. The network layer also includes a common algorithm for encapsulating network node voting and an incentive mechanism for mining and other economic factors to jointly guarantee the security of each node. Distributed network data transmission system verification algorithm, consensus algorithm, incentive mechanism together constitute the content of the protocol layer.

2.2 Extensions Layer

The problem of "intelligent contract" is the most important content in this aspect. The contract is embodied in the form of programming, which can only be executed after certain conditions are met, and finally some requirements can be intelligently realized.This level allows for more sophisticated intelligent contract types and value delivery using more complete scripting languages, allowing the development of the extension layer to be unconstrained [7,8]. The main task of this layer is to realize intelligent operation by extending the intelligent contract of the layer.

2.3 Application Layer

The application layer, as the name implies, is the layer to realize business applications. On the basis of the block chain, online shopping, games, digital assets, ownership certificates, digital currency and so on are realized. It can not only realize the function of traditional centralized server, but also has remarkable characteristics on the security guard of user information through anonymity [9].

3 Cryptography Mechanism in Blockchain

3.1 Key Pair

Blockchain uses asymmetric encryption to encrypt and decrypt data. The private key and the corresponding public key derived from the private key form a group of key pairs. In the bitcoin system, the private key determines the ownership of the bitcoin address property generated by the corresponding public key. The private key is mainly used for signature verification during the transaction. When the verification passes, the corresponding bitcoin address has the corresponding asset, otherwise it cannot be owned [10,12]. The private key must be well preserved, and once lost, the property on the corresponding bitcoin address cannot be recovered.

The private key is just a 256-bit binary number, and you can find a random number from 0 to 2^{256}. However, in order to exclude personal factors [5,14]: for

example, the selected number is special, such as 123456789, etc., it is recommended to use a pseudo-random number generator ($CSPRNG$) to generate a random number as the private key is more secure, but it must be backed up, otherwise the asset will be lost with the loss of the private key.

Using elliptic curve cryptography (ECC), the private $key(m)$can be generated into the public $key(M)$, $M = m * G$, G is the constant point of the generation point of the elliptic curve. This process is one-way and irreversible [15]. Since G is the same, the relationship between the private key and the public key is fixed. Due to the irreversibility of the operation, the private key cannot be obtained from the public key. Thus, it is guaranteed that the address generated by the public key will not expose the private key to anyone. Asymmetric encryption uses the key pair, which greatly improves the security, enables the value to be transferred between users, and achieves anonymity and security.

3.2 Address

A bitcoin address can be seen by anyone, including anyone who wants to give you bitcoin. In the transaction of bitcoin, only the receipt address is needed, but no information about the payment address is known. The address of bitcoin represents the direction of capital inflow and issuance. The appearance of bitcoin address makes bitcoin very flexible, which can be put in various occasions to achieve the purpose of collection, without worrying about whether their private key will be exposed.

The address of bitcoin is obtained from the public key through k→SHA256→ RIPEMD160(double hash) to get a 160-byte public key hash, and then the public key hash is obtained through BaseCheck58 encoding to get the address of bitcoin. Basecheck58 encoding enables Base58 encoding with version, validation format, and explicit encoding format. Combining the 160-byte public key hash (data) with the version prefix, we get a 32-byte hash value through the double hash SHA256 (SHA256 (prefix+data)), we take the first 4 bits as our check code. Prefix + public key hash + check code, together with Base58 encoding to get the bitcoin address. The version prefix is used to indicate the type of encoded data, and the checksum is used to avoid and detect whether there are any errors caused by transcription and input. When coding Basecheck58, the checksum of the original data will be compared with the checksum in the result. If it is the same, the checksum will be successful, otherwise it will fail. It is the version prefix, checksum, and so on that makes the results of the Basecheck58 encoding easier to classify.

3.3 Wallet

There are no bitcoins in the wallet. It's a collection of private keys (See footnote 1). The owner of the wallet uses a key to sign the transaction, enabling the transfer of assets.According to the relationship between private keys in wallets, wallets can be divided into two categories: non-deterministic wallets and deterministic wallets.

The first kind of wallet: the non-deterministic wallet. This kind of wallet is just a collection of randomly generated private keys, each private key is generated by a random number, there is no relationship between the two private keys, is a discrete existence. The discrete existence of private keys in such wallets makes the loss of one private key not a threat to others. However, because the number of private keys in a wallet is limited, the number of bitcoin addresses generated is also limited. If frequently traded on a few addresses, the user's property will be threatened. Some people will say: we generate a lot of private keys can not it? Because of a large number of private keys, a large number of backups are required, but since there is no association between the private keys, it is difficult to backup and manage the private keys. Therefore, this kind of wallet is only suitable for special situations, but not suitable for general situations.

The second kind of wallet: the certainty wallet, also known as the seed wallet. Through the combination of random Numbers, indexes and chain codes, the subprivate keys can be derived, which in turn can be derived from the grandson private keys, forming a tree structure. (Merkle tree) thus, the entire purse can be recovered from a single seed. With the seed wallet, key transfer and management is very convenient, See Fig. 1 for details.

Determine that the wallet is currently using the $bip-32$ standard HD wallet. The result generated by the root seed (512bits) through the one-way HMAC-SHA512 hash function is 512 bytes, and the left 256 bits are used as the main private key m, and the remaining 256 bytes are used as the right The bit-byte portion is encoded as the main chain. The main purpose of the chain code is to introduce random Numbers into the generation of this subkey. The corresponding master public key (M) is obtained by using elliptic curve cryptography $(ECC)m * G$.

From Parent Key to Child Key. The parent private key (m), the main chain code and the index number are combined to obtain a 512-bit Hash value after the three-hash of the one-way HMAC-SHA512 hash function. Where the right 256 bits encode the child chain, the remaining left 256 bits act with the index value on the parent private key (m) to generate the child private key. The subprivate key can be obtained by using elliptic curve cryptography M = m*G, and then the address can be generated by using the subpublic key.

Expanded Key. In the process of generating and secret key, we add 256-bit private key and 256-bit chain code together to form a sequence of 512 bytes. We call it the extension key, that is, the private key and chain code are combined to form the extension key.

Keys include public and private keys, so there are two types of extended keys: extended private keys and extended public keys. The chain code and the private key are combined to form the extended private key, which can be used to generate the sub-private key. The chain code and the public key are combined to form the extended public key, which can be used to generate the child public key. An extended private key can derive an entire branch, whereas an extended public key can only create a branch containing a public key because it cannot produce a

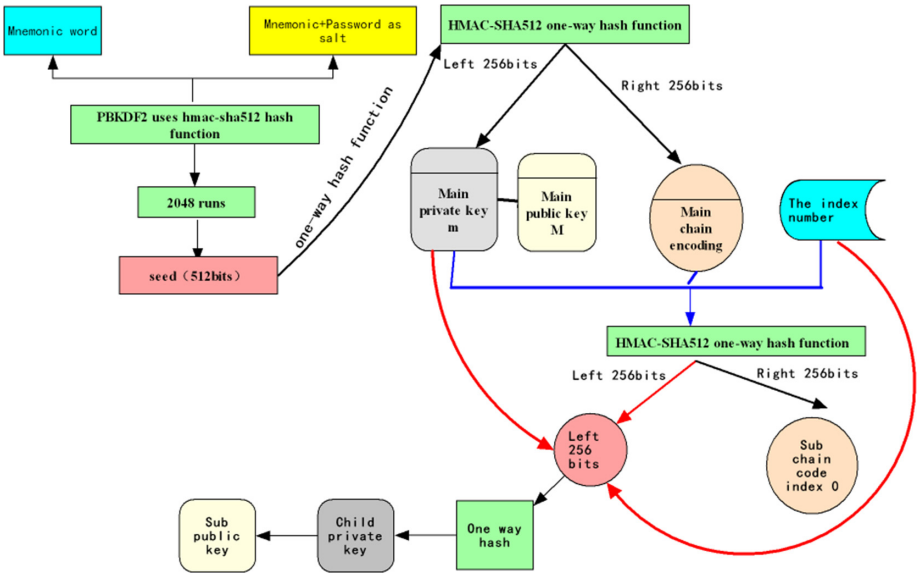

Fig. 1. Tree structure of HD wallet.

private key. As shown in the Fig. 2, the 512-bit bytes generated by combining the extended public key and index number together through the one-way HMAC-SHA512 hash function, the right 256 bit bytes are encoded as the child chain, while the left 256 bit bytes act on the parent public key to generate the child public key. However, since the extended public key contains chain code, if a child private key is exposed or lost, it will cause the insecurity of other brother private keys and even make the parent private key leak.

In this case, you can use the enhanced derivation of the child key: the method of generating the child chain code by the parent private key to solve the problem. Because we can't know the parent private key, we can't infer the parent or the brother private key by exposing the operator private key. In general, the extended public key is generally used, and the enhanced derivation method of using the child key on the parent node is generally used to ensure security, while the extended public key method is used for other nodes.

Generate Subpublic Key. To sum up, there are two ways to generate the subpublic key: one is to generate the subprivate key and then regenerate it into the subpublic key; Second, the public key can be generated directly by extending the public key. Using the extended public key to derive the HD wallet can be very convenient and secure value transfer. Using the extended public key, you can create a large number of addresses online, use the private key to sign and trade offline, and then complete the transaction by broadcasting. This ensures that safe transactions can be made even in unsafe situations.

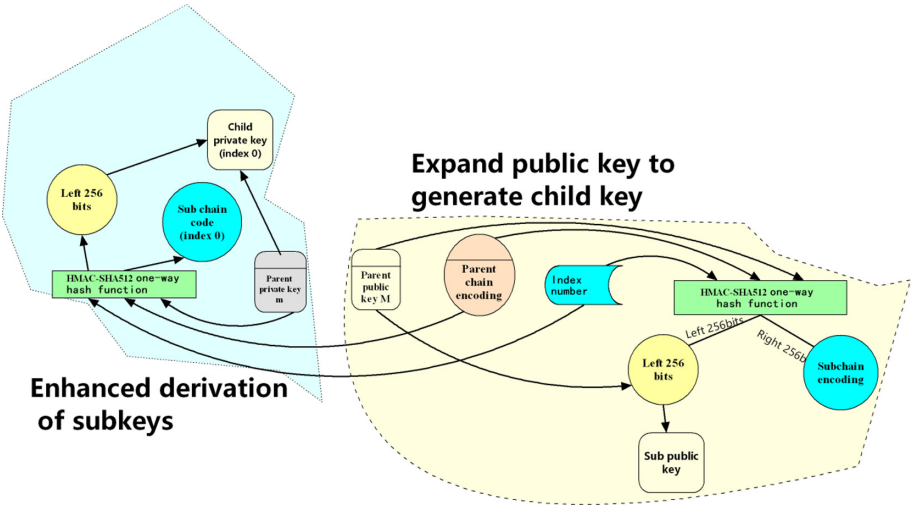

Fig. 2. Derivation of child key.

Path of Tree-Like Structure of HD Wallet. The tree structure of the HD wallet can be very large, each parent key can be composed of 4 billion subkeys: 2 billion regular subkeys, 2 billion enhanced derived subkeys. The same is true for each subkey, and so on. The tree structure of the HD wallet is huge. It is difficult to identify and manage a particular branch. A solution is provided through the bitcoin protocol BIP0044: the protocol specifies a structure consisting of five predefined tree hierarchies [6]:

$$M/purpose'/coin\_type'/'account'/change/address\_index$$

As shown in Fig. 3, purpose of the first layer is set to $44'$; The "$cion\_type''$ of the second layer specifies the currency, such as: Bitcoin uses $m/44'/0'$, Bitcoin Testnet uses $m/44'/1'$, Litecoin uses $m/44'/2'$; The third layer is "account". Users can create multiple sub-accounts to facilitate management and statistics. For example, $m/44'/0'/0'$and $m/44'/0'/1'$ are two sub-accounts of bitcoin. The fourth layer is "change". Each HD wallet has two subtrees here, one for creating a collection address and one for creating a change address. The fifth layer is the child of the fourth layer, and "$address\_index$" is the available address for the HD wallet.

4 Empirical Analysis of Trading Scripts

P2KH ($pay-to-public-key-hash$) is the most widely used bitcoin transaction. When trading, by typing the private key of digital signature and public key that can unlock by the output of the locking P2KH script will unlock script and locking in end-to-end form combination script can be verify, structure as

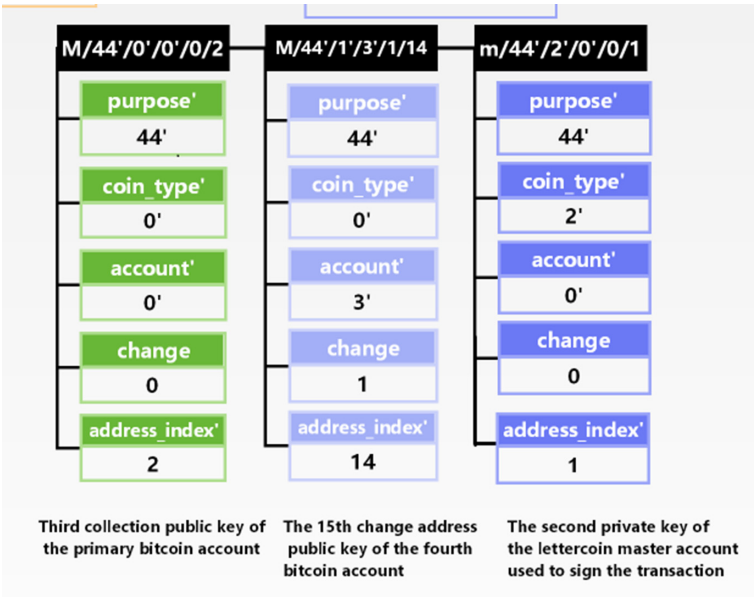

Fig. 3. BIP0044 HD wallet structure.

shown in Fig. 4, apply combination script combining stack implementation of transaction security authentication, is the realization of the programmability of the currency of important step.

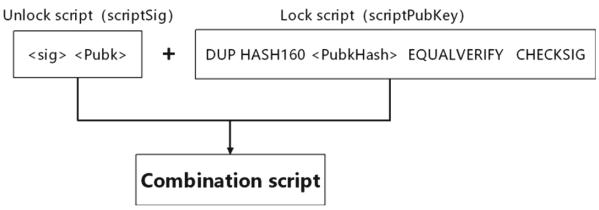

Fig. 4. Combination script.

A locking script is a constraint placed on an expense that can be unlocked and used when certain conditions are met, the process is shown in Fig. 5. Unlocking script: a script that, as the name implies, satisfies the constraints of the locking script so that it can be expensed. The implementation of the script depends on the data structure of the stack, which is a last in, first out queue. Numbers are pushed onto the stack, opcodes push or pop one or more parameters from the stack, manipulate them, and push the results onto the stack.

The p2sh script code is as follows:

```
def is_pay_to_script_hash(class_, script_public_key):
    return (len(script_public_key)==23 and
    byte2int(script_public_key)==OP_HASH160 and
    indexbytes(script_public_key,-1)==OP_EQUAL)
```

The code implements a double-level hash encoding of HASH160 for the length of the public key script, compares the result with the public key hash in the source code, and finally returns the type value of a bool.

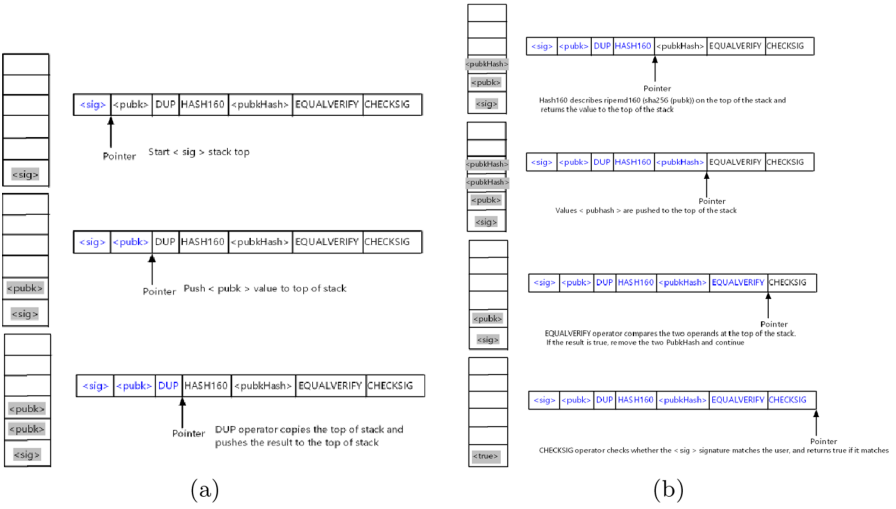

Fig. 5. P2KH combined script verification

When faced with wallets, blockchain browsers, and other applications, we see information like addresses and balances, and the transaction itself does not contain the address of bitcoin, but locks and unlocks the UTXO corresponding to the face value. When part of the UTXO is locked, only the unlocking script that meets the unlocking body condition can control the corresponding UTXO. The locking script and the unlocking script are constructed by a series of constraints, which are guaranteed by the security of cryptographic algorithms, which greatly improves the spontaneity and security of transactions.

5 Software Architecture Approach of Data Privacy Protection

Blockchain based application services may have sensitive data, which should only be available to some blockchain participants. However, the information on the

blockchain is designed to be accessible to all participants, and there is no privileged user in the blockchain network, no matter whether the blockchain is public, federated or private [16]. At the same time, the storage capacity of the blockchain network is limited, because it contains the complete history of all transactions of all participants in the blockchain network (once written, it cannot be tampered with). As a result, sensitive data privacy protection of blockchain data services is a problem, data ownership affects data transaction effect, and high redundant data storage results in huge growth of storage space. Therefore, not all data are put on one chain, only sensitive and small data are stored on the main chain. For example, food quality traceability system can store traceability information (such as traceability number and results) required by traceability laws and regulations on the main chain, and put data such as factory production process photos on the secondary chain. The advantage of storing data in multi chain is to make better use of the attributes of blockchain and avoid the limitations of blockchain data service. Blockchain services can guarantee the integrity and invariance of key data on the chain. Original big data is stored in the secondary chain or outside the chain, so the data storage size on the main chain of the blockchain service will not grow so fast, and the hash of files outside the storage chain can further ensure the integrity of files outside the chain.

In addition to the traditional choice of appropriate cryptography technology to achieve privacy protection of transaction data, we adopt the strategy of divide and rule under the condition of distributed strong load, and adopt the transaction algorithm to replace the data security sharing and business cooperation scheme of direct data interaction from the direction of software architecture (Fig. 6). The core of public service lies in the ability of data regulation. We have invented countless rules for data exchange, data opening and data maintenance. When faced with multi-party data and business collaboration, we are often unable to do our best. Establish the main technical concept of trusted, transparent and traceable blockchain data exchange and business collaboration services - data three rights separation, that is, to realize data three rights separation by technical means, to solve the most core trust problem of the digital economy.

Data owner (seller): with local data (big data), data is increasingly becoming a core asset; data executor: a trusted execution service environment for data exchange to fill the gap between data owner and user; data user (buyer): it is conducive to data analysis and processing, and can only get analysis results without necessarily obtaining source data.

From Fig. 6, under the traditional centralized data management mode, the distributed data processing process is as follows:

1. Obtain data (transactions) from all participants (sellers);
2. Perform relevant data analysis algorithm for all aggregate data (big data). Data is essentially copied from the data owner. As a result, the network traffic is large, the data privacy risk of data owners is increased, and the data transaction efficiency is low.

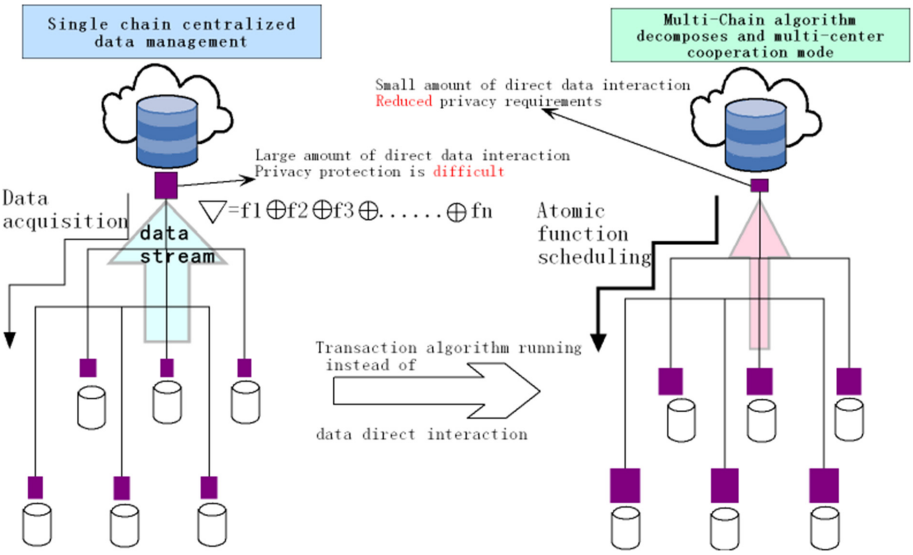

Fig. 6. Algorithm decomposition multi-center collaboration

Thus, the rules of data service usage are established, which are "do not require all, but use, call and go". By confirming the right of data and clarifying the relationship between the responsibility and right of data, the compliance exchange and capitalization of data under the compliance scenario are realized. The specific data service operation process adopts the scheme of privacy protection data sharing and collaborative implementation, establishes the privacy isolation method of the primary and secondary chain data on the chain, provides the connection function of the directory chain (main chain or on the chain) small data to the data provider (sub chain or off chain) big data, sets the directory chain pre drive contract to access the data provider to complete the data transaction, and calculates the pre drive Sub algorithm of data processing (transaction) for machine distribution.

6 Conclusions and Further Works

In this paper, we focus on the cryptography mechanism in blockchain and how to ensure the security of blockchain. Through the analysis of specific examples of bitcoin, From the perspective of security, transaction generation, broadcast, verification, block and other links are closely linked, and these links are inseparable from the calculation of multiple types and high frequency password hash functions. Cryptographic hash function plays an important role in digital signature, password protection, integrity verification, information compression and proof of work. It is precisely because of the anti-collision, antigenic attack, antigenic second attack, high sensitivity and other characteristics of cryptographic hash

function that block chain can provide people with a secure and credible trust mechanism. On the basis of this underlying architecture of equality and mutual trust, richer and more secure digital behaviors can be realized. The block chain architecture is based on distributed computing, with cryptography as the guarantee, smart contract as the entry point, and more secure value transfer as the overall goal, which is a disruptive innovation for the future development of the digital era.

Starting from the software architecture, the paper constructs data privacy protection in system, puts forward the realization method of data security sharing and business collaboration that algorithm runs instead of the direct interaction of source data, and establishes the data service usage rules that "do not require all, but use, call and go". We have designed a complete architecture and implementation method of data efficient transaction, such as data three weight separation, data privacy isolation on the chain, cross network and cross cloud deployment algorithm decomposition multi center collaboration, as well as collaborative services on the chain and off the chain, focusing on the feasibility and simple and effective implementation.

From the perspective of software architecture, further work on blockchain privacy protection includes:

1. Research and application of security-oriented block chain platform abnormal attack defense deployment.
2. Research and implementation of block chain security audit auxiliary tools.
3. Analysis and response of block chain security vulnerabilities.
4. Research on the implementation mechanism of block chain application threat intelligence (BTI) collection and feedback.
5. Optimize and quantify the resource sharing capability of data transaction services through the cooperative implementation mechanism of privacy protection data sharing.

Acknowledgments. This research has been supported by the Natural Science Foundation of China (No. 61562024, No. 61463012).

References

1. Zheng, Z., Xie, S., Dai, H., Chen, X., Wang, H.: An overview of blockchain technology: architecture, consensus, and future trends. In: IEEE International Congress on Big Data (BigData Congress), pp. 557–564. IEEE (2017)
2. Aitzhan, N.Z., Svetinovic, D.: Security and privacy in decentralized energy trading through multi-signatures, blockchain and anonymous messaging streams. IEEE Trans. Dependable Secure Comput. **15**(5), 840–852 (2018)
3. Zheng, Z., Xie, S., Dai, H.N., Chen, X., Wang, H.: Blockchain challenges and opportunities: a survey. Int. J. Web Grid Serv. **14**(4), 352–375 (2018)
4. Hu, L., Song, L.: Review and development of cryptographic hash function. Commun. China Comput. Assoc. **15**(7), 23–28 (2019)
5. Zhu, L., et al.: Survey on privacy preserving techniques for blockchain technology. J. Comput. Res. Dev. **54**(10), 2170–2186 (2017)

6. Andreas, M.: Antonopoulos. Mastering Bitcoin. The United States of America (2010)
7. Luu, L., Chu, D.-H., Olickel, H., Saxena, P., Hobor, A.: Making smart contracts smarter. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, pp. 254–269. ACM (2016)
8. Reyna, A., Martin, C., Chen, J., Soler, E., Diaz, M.: On blockchain and its integration with loT: challenges and opportunities. Future Gener. Comput. Syst. Int. J. Escience **88**, 173–190 (2018)
9. Tosh, D.K., Shetty, S., Liang, X., Kamhoua, C.A., Kwiat, K.A., Njilla, L.: Security implications of blockchain cloud with analysis of block withholding attack. In: 2017 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID), p. 458 (2017)
10. Dorri, A., Steger, M., Kanhere, S.S., Jurdak, R.: BlockChain: A distributed solution to automotive security and privacy. IEEE Commun. Mag. **55**(12), 119–125 (2017)
11. Barnas, N.: Blockchains in national defense: trustworthy systems in a trustless world. Blue Horizons Fellowship, Air University, Maxwell Air Force Base, Alabama (2016)
12. Hurich, P.: The virtual is real: an argument for characterizing bitcoins as private property. Bank. Finance Law Rev. **31**, 573 (2016)
13. He, Z., Cai, Z., Yu, J.: Latent-data privacy preserving with customized data utility for social network data. IEEE Trans. Veh. Technol. **67**, 665–673 (2018). https://doi.org/10.1109/TVT.2017.2738018.CrossRef
14. Solat, S., Potop Butucaru, M.: Zeroblock: preventing selfish mining in bitcoin. Ph.D. thesis, University of Paris (2016)
15. Zheng, X., Cai, Z., Li, Y.: Data linkage in smart IoT systems: a consideration from privacy perspective. IEEE Commun. Mag. **56**(9), 55–61 (2018)
16. Zheng, W., Zheng, Z., Chen, X., Dai, K., Li, P., Chen, R.: NutBaaS: a blockchain-as-a-service platform. IEEE Access **7**, 134422–134433 (2019)