Future Generation Computer Systems 102 (2020) 902–911

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Blockchain data-based cloud data integrity protection mechanism

PengCheng Wei [a], Dahu Wang [b,*], Yu Zhao [a], Sumarga Kumar Sah Tyagi [c], Neeraj Kumar [d]

[a] School of Mathematics and Information Engineering, Chongqing University of Education, Chongqing, China
[b] School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo, Henan, China
[c] School of Electronic and Information Engineering, Zhongyuan University of Technology, Zhengzhou, China
[d] Thapar Institute of Engineering and Technology, Patiala, Punjab, India

ARTICLE INFO

Article history:
Received 24 May 2019
Received in revised form 23 August 2019
Accepted 14 September 2019
Available online 18 September 2019

Keywords:
Blockchain
Cloud data
Integrity verification
Merkel hash tree

ABSTRACT

Despite the rapid development of cloud computing for many years, data security and trusted computing are still the main challenges in current cloud computing applications. In order to solve this problem, many scholars have carried out a lot of research on this, and proposed many models including data integrity test and secure multi-party calculation. However, most of these solutions face problems such as excessive computational complexity or lack of scalability. This paper studies the use of blockchain techniques to improve this situation. Blockchain is a decentralized new distributed computing paradigm. Applying blockchain technology to cloud computing, using the security mechanism of the former to improve the performance of the latter's secure storage and secure computing is a promising research topic. In this paper, the distributed virtual machine agent model is deployed in the cloud by using mobile agent technology. The virtual machine agent enables multi-tenants to cooperate with each other to ensure data trust verification. The tasks of reliable data storage, monitoring and verification are completed by virtual machine agent mechanism. This is also a necessary condition for building a blockchain integrity protection mechanism. The blockchain-based integrity protection framework is built by the virtual machine proxy model, and the unique hash value corresponding to the file generated by the Merkel hash tree is used to monitor the data change by means of the smart contract on the blockchain, and the data is owned in time. The user issues a warning message for data tampering; in addition, a "block-and-response" mode is used to construct a blockchain-based cloud data integrity verification scheme.

1. Introduction

In the new generation of information technology, blockchain technology will be the key to breaking the problem [1]. At present, blockchain technology is becoming a frontier field of high value with its unique technological advantages, innovative value concepts and wide application scenarios [2,3]. Many experts even believe that blockchain technology is expected to become the technology that has the potential to trigger the next wave of disruptive revolutions after steam engine, power, information and Internet technology [4]. Blockchain can solve the problem of trust mechanism. Trust is a key element of blockchain technology. It is more like a public account book that everyone can record, view, and maintain. Any record has a permanent time stamp and cannot be tampered with [5]. It is precisely because the blockchain technology has broken the centralization characteristics of the traditional Internet that the crisis of trust that plagues the modern economy has been solved to some extent. When the transaction is executed and resolved on the ledger, the parties themselves do not need to establish a trust relationship, but only need to trust the blockchain itself to achieve this goal. Blockchain can solve the problem of data authenticity. Blockchain can effectively promote data circulation and sharing [6]. Blockchain can effectively promote data production convergence. The blockchain led us to open the door to the "value Internet". The emergence of the Internet has made the means of information dissemination leap, and information can flow efficiently on a global scale without third-party and peer-to-peer implementation. The efficiency of value transfer has not been improved simultaneously. The birth of the blockchain is the beginning of human beings building the Internet worth equal to the information Internet. The value of the Internet will lay the foundation for the entire human society to enter a transparent and reliable credit society. Since the emergence of the concept of big data, data science and technology has developed rapidly. At the same time, the big data field is also facing certain problems [7–9]. Especially in the collaborative sharing of data, data transactions and data privacy protection. In the face of these problems, there are currently some solutions,

* Corresponding author.
E-mail addresses: wpc75@cque.edu.cn (P. Wei), dahuwang2008@126.com (D. Wang), zhaoyunccq@foxmail.com (Y. Zhao), sumarga@zut.edu.cn (S.K.S. Tyagi), Neeraj.kumar@thapar.edu (N. Kumar).

https://doi.org/10.1016/j.future.2019.09.028
0167-739X/

but these solutions are centralized, that is, through some trusted third-party organizations to deal with the problems faced in data processing. However, there are other problems. The so-called trusted third-party organization is really credible. Once a trusted third-party organization is doing evil, it will cause huge losses to users and data owners involved in data processing. The tripartite organization has enormous rights. Once the third-party trusted organization is controlled by the hacker, it will undoubtedly cause some losses to the user. The combination of blockchain and big data will provide solutions to the problems faced by the current big data field, and at the same time avoid the existing centralization problems [10].

The World Economic Forum report pointed out that there are currently more than 20 countries investing in blockchain-related technology areas, 80% of banks began to implement some blockchain distributed ledger-related projects in 2017, and the blockchain has become the Internet A technology that has received much attention around the world [11]. From the perspective of development trend, with the continuous maturity of blockchain technology and the increasing investment in blockchain technology research in hot industries around the world, people will explore the practical application of this technology in three stages [12]. Phase 1.0: Blockchain technology is mainly used to support digital currency represented by Bitcoin. By supporting transaction transactions such as transfer and payment between accounts, sellers and buyers can realize digital currency security without the help of third-party guarantees. Letter trading [13–15]. Phase 2.0: Combine digital currency with smart contracts, use blockchain technology to optimize a wider range of scenarios and processes in the financial sector, replace the contract with an algorithmic trading program, and trigger the network to automatically execute the contract through external conditions. In the financial industry, products such as bonds and derivatives are supported in asset trading, fund clearing, and intelligent agreements [16–19]. Stage 3.0: Blockchain technology extends from the economic field to social management, charity, culture and entertainment, medical health, science and culture and other social fields, challenging traditional centralized IT application systems, and may change our production in the future., life and social rules [20].

The block chain was first proposed by Nakamoto in 2008 and became known and familiar with the popularity of digital currencies such as Bit coin [21]. In recent years, blockchain technology and application have attracted extensive attention in academia and industry in China. Major technology companies represented by BAT and financial institutions such as banks have carried out related technical research and application research and development. In July 2016, Alibaba's Ant Financial Service developed a blockchain-based donation platform. In September 2016, Tencent's Weizhong Bank first launched the bank blockchain business in China. In July 2017, Baidu launched Commercial-grade blockchain cloud application platform and more. Blockchain has a very broad application prospect in many fields such as Internet of Things, financial technology, digital forensics, and e-government [22]. However, the blockchain technology was first proposed in 2008, and its theoretical research and application in various fields are still not mature and reliable. In the future, more new technology research and development is needed to further expand the usability and reliability of the blockchain. Many scholars have made efforts and contributions. Herlihy M et al. proposed a distributed computing platform based on blockchain, which uses external blockchain as a network controller to implement access control and privacy protection [23]. In the same year, Aniello L, Baldoni R, Gaetani E and others proposed a distributed personal data management system based on blockchain, which enables users to better control their own data and achieve privacy protection [24]. Kiayias et al. proposed the first blockchain protocol based on the verifiable security-based equity proof mechanism, and compared the corresponding security attributes in the Bitcoin blockchain protocol [25,26].

In this paper, the distributed agent model is deployed in the cloud by using the mobile agent technology. The virtual machine agent enables multi-tenants to cooperate with each other to ensure data trust verification, and complete the tasks of reliable data storage, monitoring and verification through the virtual machine agent mechanism. This is also a necessary condition for building a blockchain integrity protection mechanism. The blockchain-based integrity protection framework is built by the virtual machine proxy model, and the unique hash value corresponding to the file generated by the Merkel hash tree is used to monitor the data change by means of the smart contract on the blockchain, and the data is owned in time. The user issues a warning message for data tampering; in addition, a "block-and-response" mode is used to construct a blockchain-based cloud data integrity verification scheme.

2. Proposed method

2.1. Blockchain

(1) Introduction to blockchain

Block chain (Block chain) is an open decentralized to distributed databases, data block of the database generated by the use of cryptographic algorithms lot of time ordered in accordance generate links from. Blockchain is a new distributed computing paradigm. Its core design ideas and advantages are decentralized, and can be distributed without trust in nodes through encryption algorithms, time stamping techniques, consensus mechanisms, and incentive mechanisms. Network technology is used to implement peer-to-peer based point-to-point information transfer, coordination and collaboration. A blockchain consists of a sequence of chronological blocks that hold a complete list of all valid transactions in the network. Each data block in a blockchain generally includes a block header and a block body. In the block, the block body mainly contains a transaction counter and detailed transaction data. The block header contains information such as Merkle root hash, parent hash value, time stamp, calculation difficulty, and random number. The maximum number of transactions a block can contain depends on the size and block size of each transaction. The type of data record is determined according to the scene, such as asset transaction records, asset issuance records, clearing records, and IoT data records. The data is recorded in the stored procedure, usually organized into a tree logical structure, such as a Merkle tree. The underlying data is evaluated layer by a hash function such as SHA-256 until the hash value of the Merkle root node is finally determined. Root hash. The other information of the timestamp also includes the version information of the block, the random number, etc., and the random number is an important parameter in the workload proof mechanism.

(2) Key characteristics of the blockchain

In general, the blockchain has the following key features:

First, decentralization. In the center of the traditional trading systems, every transaction needs to be validated by the central mechanism credible, which inevitably leads to the cost and security center server. In contrast, the central mechanism is no longer needed in the blockchain system, and the participants in the blockchain network maintain the data consistency in the distributed network through a consensus algorithm.

Second, persistence. Effective transaction data can be quickly verified, and once a transaction is added to the blockchain, it is almost impossible to delete, modify, or roll back the transaction. Moreover, the block containing invalid transaction information

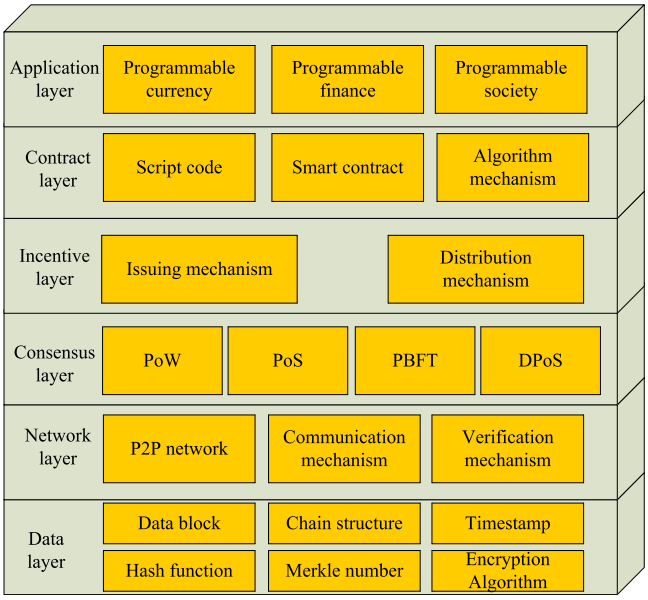

Fig. 1. Blockchain infrastructure.

during the verification process will be quickly checked out and discarded.

Third, anonymity. Each user interacts with the blockchain using the generated user address to hide the true identity information.

Fourth, auditability. Bitcoin uses the (UTO) model to store user balance information. Any transaction must reference some of the previous unsold transactions. Once the transaction is recorded in the blockchain, the status of the quoted transaction changes. Therefore, transaction records can be easily verified and tracked.

In recent years, many financial technology institutions, government departments, and technology companies have begun to research and apply blockchain technology in view of the important characteristics of blockchain technology such as safety, irreversibility, and irreversible modification. Similar to the development and application classification of cloud computing, the current blockchain technology can be roughly divided into three categories according to the application scenario. One type is called the Public Block chain, such as bitcoin, Ethereum and other virtual currency systems. The other category is the Private Block chain. Common application examples include the Linux Foundation's Hyper ledger project, the R3CEVCorda platform, and the Gem Health network. The last category is the Consortium Block chain, such as IBM's Fabric, MicroBank's BCOS project, and the Golden Chain Alliance.

(3) Infrastructure

A blockchain is an open, distributed decentralized digital ledger that acts as a data structure for recording transaction record information between users in digital currency such as bitcoin. The blockchain system is automatically managed based on peer-to-peer networks and distributed timestamp servers. The blockchain's infrastructure and underlying technology are characterized by decentralization, openness, autonomy, information modification and anonymity. These features make the blockchain faster and more traditional than the traditional key–value exchange system. Safe and less expensive. The seven-layer model of analog computer network divides the blockchain system model into data layer, network layer, consensus layer, incentive layer, contract layer and application layer, as shown in Fig. 1.

According to the above division model, at the bottom is the data layer, which is the basis for the implementation of various functions of the upper layer. This layer mainly implements two functions: one is the package construction of the block, which is the process of data storage, and the other is the realization and security of the account and transaction. Therefore, in the data layer, the algorithms related to the encryption and block construction of the underlying data and the block data are mainly included. This is followed by the network communication layer, which mainly implements network construction, verifies block data, and discovers connection and communication of new nodes and nodes in a decentralized distributed peer-to-peer network. The network layer mainly includes peer-to-peer (P2P) technology, unicast/multicast communication, and authentication technologies. The consensus layer is the core component of the blockchain system, and mainly includes the consensus algorithm of the blockchain system. Common consensus mechanisms include a workload proof mechanism (PoW), a rights proof mechanism (PoS), and a practical Byzantine fault tolerance mechanism (PBFT). The incentive layer integrates economic factors into the blockchain technology system, so that decentralized consensus nodes can be self-assembled into large-scale distributed computing systems. The contract layer mainly includes execution scripts, smart contracts, and other types of transaction logic implementation algorithms, which are the basis for the blockchain to implement programmable features. The business application layer is the specific application scenarios and cases of the blockchain. Based on blockchain technology, not only digital currency applications like Bitcoin and Ethereum can be constructed, but also various business applications such as intellectual property protection, forensic services, online voting, and online games.

2.2. Cloud data storage security related technology

By analyzing the security requirements of current cloud storage data, it can be concluded that the three aspects of anti-eavesdropping, tamper resistance, anti-discarding and anti-abuse are corresponding to the three elements of data security, namely confidentiality, integrity and availability (CIA). Correspondingly, data encryption, access control, data integrity verification, data recovery and other technologies are currently the key technologies to ensure data security in the cloud storage environment. The ciphertext access control technology and integrity verification technology will be briefly introduced.

(1) Cloud storage data security requirements

In cloud computing, data storage is not only the focus, but also the foundation. Data security is also the top priority of cloud computing security protection. Therefore, data information is an important part of the assets of enterprises and individual users. Whether data information stored in the cloud is safe is the most concerned issue for users. A general cloud storage service security model is shown in Fig. 2.

For all kinds of malicious operations that need to be guarded against the cloud server, user data security requirements include the following aspects:

1. Anti-eavesdropping. All data stored by the user on the cloud server should be encrypted ciphertext; users should use ciphertext as much as possible when communicating with the cloud server or other users to avoid leakage of data information.

2. Tamper-proof. All data stored by the user on the cloud server side should have a complete data integrity protection scheme to detect malicious tampering behavior of the cloud server.

3. Anti-discard. Because the amount of data stored on the cloud server is too large, cloud service providers may discard some of the less commonly used data for their own economic benefits. To this end, users should have their own integrity detection function when storing their data to ensure that data can be stored in the cloud at any time and anywhere.

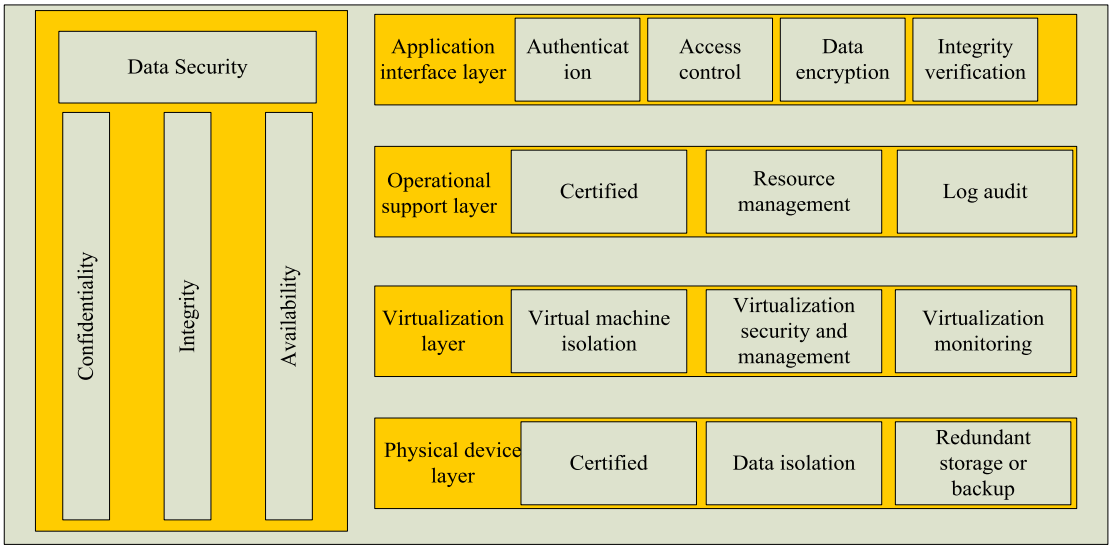

Fig. 2. Cloud storage service security model.

4. Anti-abuse. When users store their data, they should also be equipped with relevant measures to prevent data from being abused or processed by the cloud server.

(2) Ciphertext access control technology

Data confidentiality means that only data owners and authenticated users can access or obtain data expressly, and no other user can access or access data explicitly. The most common way to protect data confidentiality is encryption. Typically, users encrypt data before transferring their data to the cloud. Access control is one of the important strategies to ensure the legality of information use, network security, and system resource protection. In different access control models, the subject implements access to its resources based on different policies. Since the data stored in the cloud is in the ciphertext state, the user's access to it becomes a ciphertext access control problem. The ciphertext access control technology controls the user's access by encrypting the key information and controlling the access rights of the key information. It is an important means to ensure the confidentiality of user data in the cloud untrusted environment, not only can effectively improve the confidentiality and privacy of user data, but also greatly reduce the risk of user data from being illegally leaked. Attribute-based ciphertext access control (ABAC) is an attribute-based encryption algorithm (ABE) combined with access control technology that allows data to be accessed only if the user requesting access satisfies the attribute's decision rules. Because ABAC has better flexibility and finer access control granularity, it is more suitable for cloud storage environments with multiple tenants and frequent permission changes. In the KP-ABE and CP-ABE schemes, the user can decrypt only when the attribute set satisfies the access tree. Combined with the high resource sharing, high system openness and high data dynamics in the cloud storage environment, CP-ABE is more suitable for cloud storage access control systems than KP-ABE.

(3) integrity verification technology

Data integrity is an important basis to reflect whether the data is real and reliable, including the integrity of storage and use. Usually, data integrity in cloud storage environment means that cloud storage service providers store data completely on cloud servers according to users' needs, that is, storage integrity. Integrity verification of data in a cloud storage environment, also known as provable storage, is to allow users to retrieve a small amount of stored data with a certain knowledge protocol or to judge whether the data stored in the cloud is high or not. complete. In the traditional sense, there are two main ways to verify the integrity of data in a storage system: one is based on access; the other is based on challenge-response. In contrast, the latter is more suitable for use in distributed cloud storage environments. The challenge-response-based integrity verification model includes verifiers and responders. The certifier is usually the data owner or a trusted third party, and the responder is the cloud server. The working principle is: first, the verifier sends a challenge information to the cloud server; then the cloud server generates and returns corresponding response information according to the content of the challenge information; finally, the verifier performs integrity verification and judgment according to the received response information. The main technologies used include provable data holding (PDP) technology and recoverable proof (POR) technology.

The workflow of the challenge-response-based data integrity verification scheme includes the following three phases: 1. Setup phase: The data owner first preprocesses the data files (such as blocking data files, generating various tag information, etc.), and then transfer the preprocessed data information to the cloud storage server. 2. Challenge stage: The verifier generates corresponding challenge data according to its own needs and sends it to the cloud server. 3. Check Proof stage: The cloud server generates corresponding response data according to a specified protocol and returns it to the verifier. The verifier performs calculation based on the returned response data and determines whether the data is true.

2.3. The correctness of the verification process

(1) The correctness of the single-user verification process

Evidence validation phase, the user after receiving the evidence submitted by the storage service, with their public key and preset parameters to calculate the verification equation is established, to conclude that data integrity is damaged. In the case where the data is complete, that is, the parameters are not changed, and the left and right sides of the verification equation are respectively calculated.

Left side of the equation:

$$e(\sigma, g) = e\left(\Pi_{i=1}^{r_c}\sigma_i^{\omega_i}, g\right) = e\left(\Pi_{i=1}^{r_c}(h(f_i)\cdot u^{f_i})^{k\cdot\omega_i}, g\right)$$
$$= e\left(\prod_{i=1}^{r_c}(h(f_i)\cdot u^{f_i})^{\omega_i}, g\right) = e\left(\prod_{i=1}^{r_c} h(f_i)^{\omega_i} \bullet \prod_{i=1}^{r_c} u^{f_i\omega_i}, g\right) \quad (1)$$
$$= e\left(\prod_{i=1}^{r_c} h(f_i)^{\omega_i}\cdot u^{\sum_{i=1}^{r_c} f_i\omega_i}, v\right)$$

Right side of the equation:

$$e\left(\prod_{i=1}^{r_c} h(f_i)^{\omega_i}\cdot u^{\mu}, v\right) = e\left(\prod_{i=1}^{r_c} h(f_i)^{\omega_i}\cdot u^{\sum_{i=1}^{r_c} f_i\omega_i}, v\right) \quad (2)$$

It can be seen that the equations are equal on both sides. Since the parameters u and v are both publicly available, only when the parameters in the formula (3) are changed, there will be an imbalance between the two ends of the equation, and the data on the cloud server side may be changed. Therefore, by determining the equilibrium of the equation, it is determined whether the integrity of the data is intact.

$$proof = \{\mu, \sigma, (h(f_i))_{1\le i\le r_c}\} \quad (3)$$

(2) Extension of multi-user data integrity authentication

Because the cloud server can handle multiple data verification for multiple different users simultaneously. The specific improvement steps are as follows:

First, the initialization phase

During the initialization phase, the user completes the random generation of public–private key pairs. For each user $n \in (1, \ldots, N)$, apply for a certificate in the CA system, obtain a pair of signed public and private keys (ssk, spk), then select a random number to select $x_n \leftarrow Z_p$, and calculate $v_n = g^{x_n}$, user n will use v_n as the public key, at this time $v_n \in Z_p$.

Second, the data label phase

Suppose each user divides the outsourced file into k segments to get the formula (4). The user selects $u_n \leftarrow G$ and calculates the signature as in Eq. (5).

$$F_n = f_{n,1} \parallel f_{n,2} \parallel \ldots \parallel f_{n,k} \quad (4)$$

$$\sigma_{n,j} \leftarrow [h(f_{n,j})\cdot u_n^{f_{n,j}}]^{x_n} \quad (5)$$

Third, the challenge stage

The user randomly selects a subset $R = \{i_1, i_2, \ldots, i_c\}$ from the set $\{1, 2, \ldots, k\}$, and the subset contains c elements, assuming that the elements in R have been arranged in order from small to large. The message verifier sends an inquiry to the server, as in Eq. (6).

$$chal = \{(i, v_i)\}_{i_1\le i\le i_c} \quad (6)$$

Fourth, the evidence generation stage

The server receives (as in Eq. (6)), and for each user $n \in \{1, \ldots, N\}$, the server calculates as shown in Eqs. (7) and (8). The server replies to the verifier, as in Eq. (9).

$$\tau = \sum_{\{(i,v_i)\}i_1\le i\le i_c} v_i\cdot f_{n,j} \in Z_p \quad (7)$$

$$\sigma = \prod_{n-1}^{N}\left(\prod_{\{(i,v_i)\}_{i_1\le i\le i_c}} \sigma_{n,j}^{v_i}\right) = \prod_{n-1}^{N}\left(\prod_{\{(i,v_i)\}_{i_1\le i\le i_c}} [h(f_{n,j})\cdot u_k^{f_{n,j}}]^{x_n v_i}\right) \quad (8)$$

$$\{\sigma, \{\tau\} 1 \le n \le N, \{h(f_{n,j})\}\} \quad (9)$$

Fifth, the evidence verification phase

After receiving the evidence, each user uses his or her public key to verify whether the equation (Eq. (10)) is established. If it is established, the data integrity is not destroyed.

$$e(\sigma, g) = \prod_{n=1}^{N} e\left(\prod_{\{(i,v_i)\}_{r_1\le i\le r_c}} h(m_{n,i})^{v_i}\cdot u_n^{\tau_n}, v_n\right) \quad (10)$$

The correctness of data integrity verification is similar to that of a single user.

2.4. Cost calculation

(1) Storage cost calculation

The storage cost of applying the m-MHT directly on the Pivot-Universal storage mode to establish the corresponding auxiliary verification structure can be regarded as the sum of multiple m-MHT storage costs. If the tenant customizes k query attributes, it can be obtained:

$$C_s^{m-MHT} = \sum_{i=1}^{k}\left(C_s^P + C_s^U\right)$$
$$\approx 2k\cdot\left(\frac{Nf-1}{f-1}\cdot(|k|+|h|) + \frac{N-1}{f-1}\cdot|p| + |s|\right) \quad (11)$$

Similarly, the storage cost of MTAS is:

$$C_s^{MTAS} \approx \sum_{i=1}^{k} C_s^{RPUA} + |H - set|$$
$$\approx k\cdot\left(\frac{Nf-1}{f-1}\cdot(|k|+|h|) + \frac{N-1}{f-1}\cdot(|p|+|h|+N\cdot|k|+|s|)\right)$$
$$+N\cdot(2|k|+|h|) \quad (12)$$

The storage cost is the sum of the verification tree storage costs built by the tenant based on the query attributes after customizing the k query attributes. $(k-1)\cdot N\cdot|k| + k\cdot|s| - (N\cdot 2\cdot|k| + \frac{N-1}{f-1}(|h|-|p|))$ can be obtained by subtracting (11)–(12) from the above formula. Since the size of the digital signature and hash is larger than the query keyword and pointer size, it can be concluded that it is different from each query attribute. The m-MHT method is constructed for the data in the pivot table and the universal table respectively. The MTAS reduces the system storage cost, because the MTAS reduces the storage of the $(k-1)\cdot N/f$ hash value on the leaf node, and reduces the $k\cdot|s|$ root node signature storage on the root node.

(2) V0 construction cost

In the m-MHT method, the corresponding verification path needs to be constructed for the query results set(QI) and set(QS) respectively, so that the maximum value of the object V0 is verified by considering that the root node does not intersect the ql+ and qu+ verification paths. for:

$$C_s^{m-MHT} = 2\cdot[2\cdot(\log_f N - 1)\cdot(f-1)\cdot(|k|+|p|+|h|)+|s|]+|AI| \quad (13)$$

The maximum value of the V0 size built on the MTAS is:

$$C_s^{MTAS} = 2\cdot(f-1)\cdot[(\log_f N - 1)\cdot(|k|+|p|+2|h|)+(2\cdot|k|-|h|)] \quad (14)$$

It can be seen that, compared with the m-MHT method separately constructed for the data in the pivot table and the universal table according to each query attribute, the V0 constructed on the MTAS only needs to perform the traversal of the RPUA tree to obtain the query result set(QI) and Set (QS) validation object reduces the number of traversal of the tree and reduces the complexity of the validation object generation.

(3) verification cost

When applying data verification, the integrity protection module reconstructs the RPUA tree root node by using Set(QI), Set(QS), VO and tenant auxiliary information returned by the data engine. Assuming $d = (\log_f N - \log_f N_Q)$ and $d' = (\log_f N = \log_f \frac{N_Q}{t})$, the computational cost of reconstructing the root node in the m-MHT tree is:

$$C_C^{m-MHT} = ((N_Q \cdot f - 1)/(f-1) + d \cdot (f-1)) \cdot C_h + C_v \quad (15)$$

The computational cost of reconstructing the root node in the RPUA tree is:

$$C_c^{PUA} = ((N_Q \cdot (f+1) - 1)/(f-1) + 2 \cdot d \cdot (f-1)) \cdot C_h + C_v \quad (16)$$

2.5. Merkel hash tree

Merkle Tree can be seen as a generalization of Hash List (Hash List can be seen as a special Merkle Tree, a multi-forked Merkle Tree with a height of 2). At the bottom, like the hash list, we divide the data into small chunks of data, with corresponding hashes and corresponding ones. But going up, instead of directly computing the root hash, we merge the two adjacent hashes into a string, and then operate the hash of the string, so that every two hashes get married and have children. A "sub-hash". If the lowest number of hashes in the bottom layer is singular, then a single hash will inevitably appear in the end. In this case, it is hashed directly, so it can also get its child hash. So push up, still the same way, you can get a new number of new hashes, and eventually form an upside down tree. At this position in the root of the tree, this generation has a root hash. We call it Merkle Root. Before downloading the network from the p2p network, obtain the Merkle Tree root of the file from a trusted source. Once you have the root of the tree, you can get the Merkle tree from other sources that are not trusted. Check the received Merkle Tree with a trusted root. If the Merkle Tree is corrupt or fake, get another Merkle Tree from another source until you get a Merkle Tree that matches the root of the trusted tree. The main difference between Merkle Tree and Hash List is that you can download and immediately verify a branch of Merkle Tree. Because the file can be divided into small data blocks, so if there is a piece of data corruption, just re-download the data block. If the file is very large, then both the Merkle tree and the Hash list are there, but the Merkle tree can download one branch at a time, and then verify the branch immediately. If the branch is verified, the data can be downloaded. The Hash list can only be verified by downloading the entire hash list.

3. Experiments

3.1. Experimental environment

Cloud data integrity protection has become an important research hotspot at this stage, and the method of using network coding technology for storage and integrity protection has achieved certain results, but its computational overhead in the integrity protection process is too large., seriously affecting the operational efficiency of the storage node and reducing the availability of the cloud storage service. In order to verify the correctness of the above theoretical proof, this paper carries out simulation experiments. The experimental environment is as Table 1.

Experimental simulation implementation process (see Fig. 3):

In the experimental simulation process, a virtual cloud storage service environment is set up, including 1 cloud storage management server and 5 storage node servers, one of which serves as a backup node for the data integrity protection node and four as the running cloud. Storage service node. The cloud service management node stores the basic information of each node storing data,

Table 1
Experimental environment parameters.

Category	Parameter
CPU	Core(TM)2
Main frequency	2.67 GHz
RAM	4G
Operating system	Windows7
Simulation software	MATLAB7.0

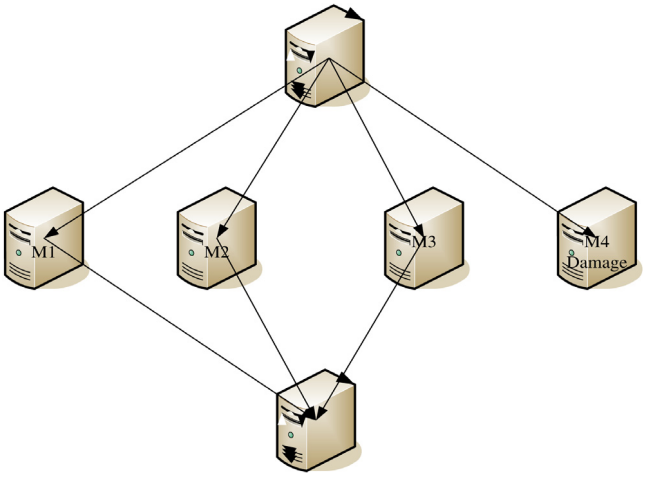

Fig. 3. Simulation experiment topology.

and its main function is to manage the corresponding storage node; the storage node stores the corresponding data and the connection between the storage servers, mainly for data storage. And data transfer during data integrity protection. The specific environment of the experimental simulation is as follows. The k data blocks are stored on n nodes for discussion of data integrity protection in case of damage to a single node. The calculated amount of data is the number of multiplication operations in the data integrity protection process to compare the computational overhead of the two data integrity protection algorithms.

3.2. Calculation in the program

By deploying distributed virtual machine agents in the cloud and using the multi-tenancy in the cloud to form a blockchain network, this paper. An integrity verification scheme BPDP for cloud data is considered, which consists of five algorithms:

(1) Generate a public–private key pair, as shown in Eq. (17):

$$KeyGen\ (1^k) \rightarrow (pk.sk) \quad (17)$$

(2) Generate a digital label, as shown in Eq. (18):

$$TagBlock\ (pk, sk, m) \rightarrow T_m \quad (18)$$

(3) Generate challenge information, as shown in Eq. (19):

$$GenChal\ (c, r) \rightarrow chal \quad (19)$$

(4) Generate evidence, as shown in Eq. (20):

$$Gen\ \mathrm{Pr}\ oof\ (pk, F, chal, \Sigma) \rightarrow V \quad (20)$$

(5) Test evidence, as shown in formula (21):

$$Check\ \mathrm{Pr}\ oof\ (sk, V) \rightarrow result \quad (21)$$

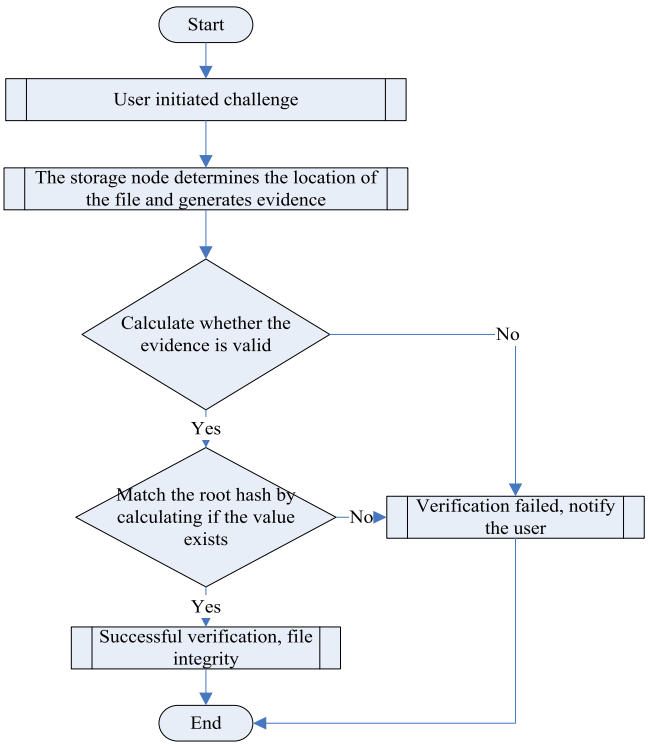

Fig. 4. Integrity verification flow chart.

3.3. Integrity verification process

The integrity verification phase is shown in Fig. 4. The user randomly extracts a data block, challenges the storage node of the CSP through the VMA node, obtains the file location according to the IPFS cluster challenge block, generates evidence, and returns it to the VMA. The VMA passes the verification to calculate the evidence. Whether it is valid, if it is valid, the second step is verified. The MHT is used to calculate whether the challenge block exists and is consistent with the root hash value. If they are consistent, the proof file is complete, otherwise the file is destroyed.

4. Discussion

4.1. Cloud data integrity analysis based on blockchain

(1) Integrity analysis

Firstly, this paper analyzes the accuracy of the sampling-based integrity verification protocol. It is assumed that the total number of data blocks on the cloud server is n. If the data is illegally tampering with e data blocks, the ratio of corrupted data blocks is: $p_b = \frac{e}{n}$, assuming t is per The number of data blocks of the secondary challenge accounts for the ratio of the total number n, so the probability of detecting illegal tampering each time is P, as in Eq. (22).

$$P = P\{X \geq 1\} = 1 - \frac{n-e}{n} \cdot \frac{n-1-e}{n-1} \cdots \frac{n-t\cdot n-e}{n-t\cdot n}$$
$$\geq 1 - \left(\frac{n-e}{n}\right)^{t\cdot n} = 1 - (1-P_b)^{t\cdot n} \quad (22)$$

The security parameter λ selected in this paper is 160bit, which means $|p| = 320$, the TagCost of the data tag is: $n * p/16$, and n is the number of data blocks. In order to realize the dynamic addition, deletion and change, the index table IHTCost is established, and the storage size is $n*(2*p2logn)/16$. When the number of data segments is constant, the larger the data segment, the smaller the number of data blocks formed, so the index table The storage overhead is reduced, as shown in Fig. 5.

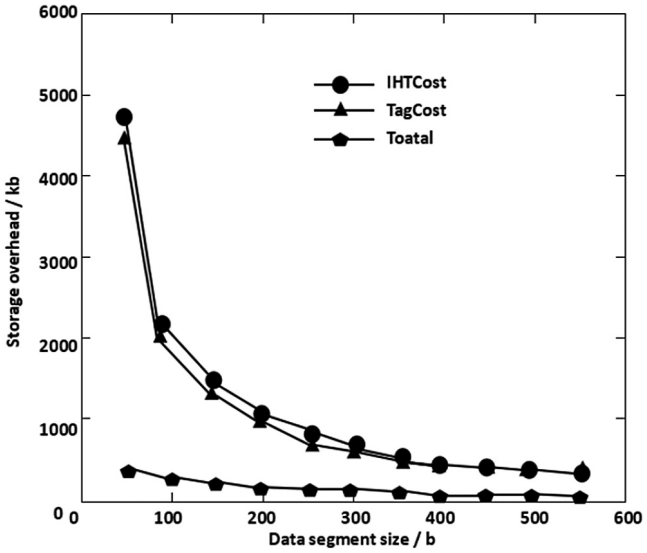

Fig. 5. Storage overhead map.

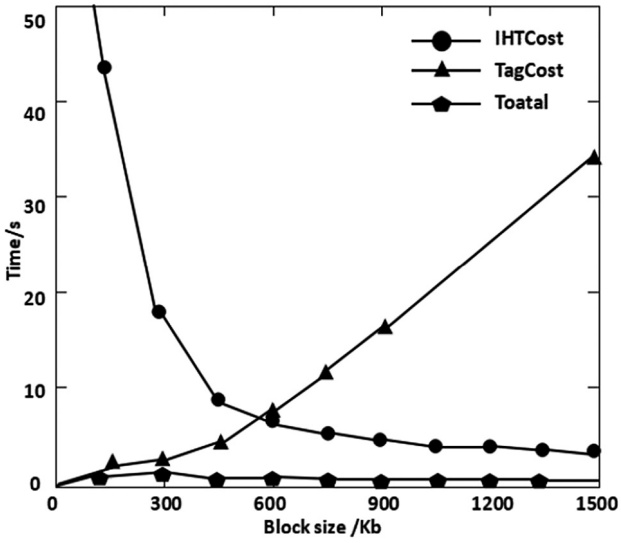

Fig. 6. Time overhead.

Time module analysis of key module cloud data integrity verification, as shown in Fig. 6, when the file block is too small, resulting in a sharp increase in the number of file blocks, the preprocessing time takes a long time, when the file block is large, resulting in data When the block is divided into data segments, the number of data segments increases sharply, which leads to an increase in the time for generating evidence. The setting file is 1 GB. In the case of reasonable setting of the file block size, a series of integrity verification is completed in 10 s.

(2) Cost analysis

Fig. 7 shows the average V0 size after 100 random queries based on individual query ranges on the tenant data tuple. As can be seen from the figure, the MTAS has a smaller V0 size than the m-MHT method, that is to say the MTAS has a smaller communication cost consumption. The main reason for this result is that, due to the use of the composite verification structure,

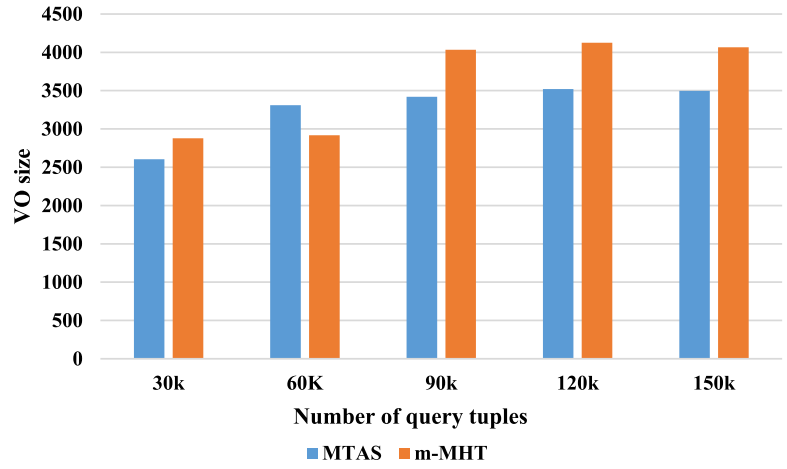

Fig. 7. Comparison of V0 size in m-MHT and MTAS.

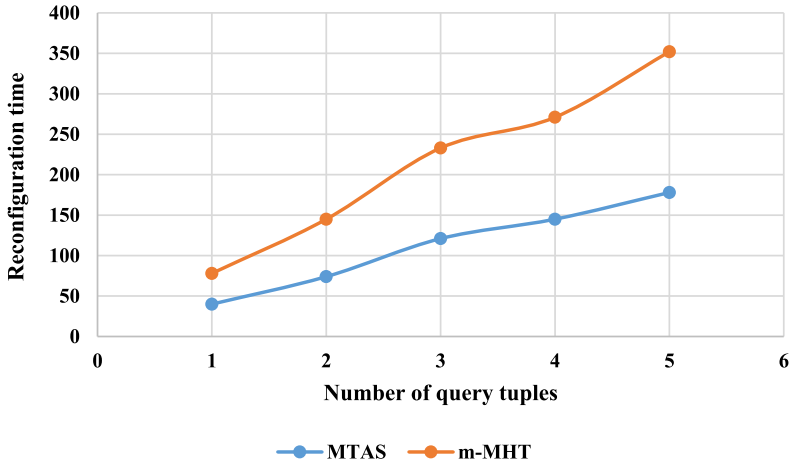

Fig. 8. Comparison of reconstruction time in m-MHT and MTAS.

MTAS does not need to return the hash value of the leaf node corresponding to the boundary data tuple in the pivot table under the same query condition, so the V0 size in the MTAS is smaller than The size of V0 in the m-MHT method.

Fig. 8 shows the comparison of reconstruction time in the m-MHT and MTAS methods. MTAS has more efficient verification efficiency. This is because when the root node hash value is reconstructed in the MTAS, it is not necessary to calculate the corresponding hash value for each data tuple in the pivot table, and only the corresponding multiple pivot table data elements in the AUL layer need to be calculated. A hash value of the group, so the larger the number of branches f of the PUA tree, the more the number of hash calculations saved by the MTAS, so the MTAS has better verification efficiency.

4.2. Implementation of some functions

In this paper, a fixed-size file is generated by randomly filling letters or numbers. The user sets the appropriate number of data segments, data segment size and data block size to preprocess the file. First, the file is divided into data blocks, and then the data blocks are divided. Divided into data segments, as shown in Fig. 9.

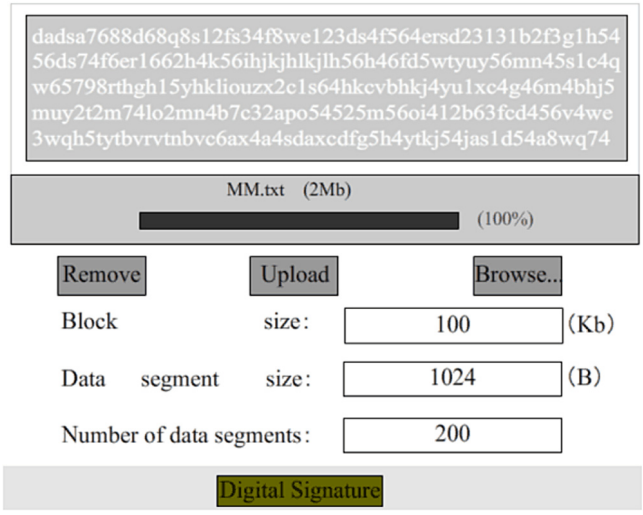

Fig. 9. Preprocessing function.

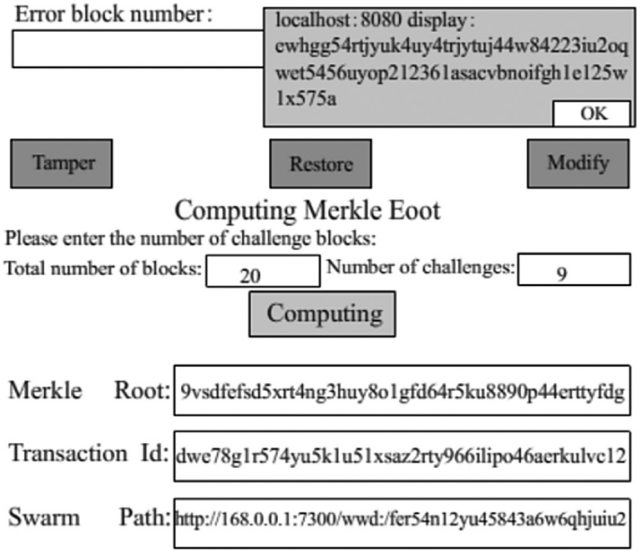

Fig. 10. Simulation data label destruction verification results.

Fig. 10 is a graph showing the results of simulation data tag destruction verification. If the data stored in the storage node is destroyed, the digital label of the data is also tampered with, and only the MHT root hash value is recalculated, and the information in the blockchain network is compared, and the file is destroyed, and the new calculation is performed. The MHT hash is inconsistent with the original MHT hash, indicating that the data tag was corrupted.

5. Conclusions

(1) This paper uses mobile agent technology to deploy distributed virtual machine proxy model in the cloud. Through virtual machine agent, multi-tenant can cooperate with each other to ensure data trust verification, and complete virtual data storage, monitoring and verification through virtual machine proxy mechanism. Waiting for tasks, this is also a necessary condition for building a blockchain integrity protection mechanism. The blockchain-based integrity protection framework is built by the virtual machine proxy model, and the unique hash value corresponding to the file generated by the Merkel hash tree is used to monitor the data change by means of the smart contract on the blockchain, and the data is owned in time. The user issues a warning message for data tampering; in addition, a "block-and-response" mode is used to construct a blockchain-based cloud data integrity verification scheme.

(2) Combining the advantages of blockchain and cloud computing, this paper constructs a key technology application scheme of cloud computing based on blockchain, which realizes the security protection and integrity check of cloud data on the one hand, and achieves wider security on the other hand. Multi-party calculations.

(3) Propose a secure multi-party computing scheme based on blockchain. Security mechanisms and algorithms such as encryption algorithm, consensus mechanism and incentive mechanism in blockchain are studied, and the general process of scalable multi-party computing, attack model and typical cryptographic tools including password sharing are analyzed and studied. Furthermore, the blockchain scheme based on computing node operation log information is studied and applied to secure multi-party computing modeling.

Declaration of competing interest

None.

Funding

This work was supported by Chongqing Big Data Engineering Laboratory for Children, China, Chongqing Electronics Engineering Technology Research Center for Interactive Learning, China, Project of Science and Technology Research Program of Chongqing Education Commission of China. (N0. KJZD-K201801601) and National Natural Science Foundation of China Youth Science Foundation (No. 61601172).

References

[1] Y. Yuan, F.Y. Wang, Blockchain: The state of the art and future trends, Acta Automat. Sinica 22 (03) (2016) 1882.
[2] M. Pilkington, Blockchain technology: Principles and applications, Soc. Sci. Electron. Publ. 51 (07) (2015) 121–122.
[3] R. Pass, L. Seeman, A. Shelat, Analysis of the blockchain protocol in asynchronous networks, in: International Conference on the Theory & Applications of Cryptographic Techniques, 2017.
[4] A. Azaria, A. Ekblaw, T. Vieira, et al., MedRec: Using blockchain for medical data access and permission management, in: International Conference on Open & Big Data, 2016.
[5] J. Yli-Huumo, D. Ko, S. Choi, et al., Where is current research on blockchain technology?-A systematic review, Plos One 11 (10) (2016) e0163477.
[6] N. Zhang, Y. Wang, C. Kang, et al., Blockchain technique in the energy internet: Preliminary research framework and typical applications, Proc. Csee 33 (01) (2016) 11–12.
[7] A. Kiayias, A. Russell, B. David, et al., Ouroboros: A provably secure proof-of-stake blockchain protocol, in: International Cryptology Conference, 2017.
[8] I. Weber, X. Xu, R. Riveret, et al., Untrusted business process monitoring and execution using blockchain, in: International Conference on Business Process Management, 2016.
[9] X. Yue, H. Wang, D. Jin, et al., Healthcare data gateways: Found healthcare intelligence on blockchain with novel privacy risk control, J. Med. Syst. 40 (10) (2016) 218.
[10] G.W. Peters, E. Panayi, Understanding modern banking ledgers through blockchain technologies: Future of transaction processing and smart contracts on the internet of money, Soc. Sci. Electron. Publ. 71 (01) (2016) 1113.
[11] M. Conoscenti, A. Vetrò, J.C.D. Martin, Blockchain for the internet of things: A systematic literature review, Comput. Syst. Appl. (2017).
[12] N.Z. Aitzhan, D. Svetinovic, Security and privacy in decentralized energy trading through multi-signatures, blockchain and anonymous messaging streams, IEEE Trans. Dependable Secure Comput. PP (99) (2016) 1.
[13] M. Mettler, Blockchain technology in healthcare: The revolution starts here, in: IEEE International Conference on E-Health Networking, 2016.
[14] V.L. Lemieux, Trusting records: is Blockchain technology the answer? Rec. Manag. J. (2) (2015).
[15] S. Huckle, R. Bhattacharya, M. White, et al., Internet of things, blockchain and shared economy applications, Procedia Comput. Sci. 98 (C) (2016) 461–466.
[16] M. Swan, Blockchain thinking : The brain as a decentralized autonomous corporation [commentary], IEEE Technol. Soc. Mag. 34 (4) (2015) 41–52.
[17] H. Watanabe, S. Fujimura, A. Nakadaira, et al., Blockchain contract: Securing a blockchain applied to smart contracts, in: IEEE International Conference on Consumer Electronics, 2016.
[18] M. Atzori, Blockchain technology and decentralized governance: Is the state still necessary? Soc. Sci. Electron. Publ. 6 (1) (2016).
[19] Z. Yu, J. Wen, The iot electric business model: Using blockchain technology for the internet of things, Peer-to-Peer Netw. Appl.sss 10 (4) (2017) 983–994.
[20] F. Glaser, Pervasive decentralisation of digital infrastructures: A framework for blockchain enabled system and use case analysis, Soc. Sci. Electron. Publ. 11 (02) (2017) 67–70.
[21] Z.K. Tang, Nakamoto satoshi: He invented bitcoin? East West North (9) (2014) 58–59.
[22] H.J. Jin, K.H. Kim, J.H. Kim, Block chain based data security enhanced IoT server platform, in: International Conference on Information Networking, 2018.

[23] M. Herlihy, Blockchains and the future of distributed computing, in: Acm Symposium on Principles of Distributed Computing, 2017.
[24] L. Aniello, R. Baldoni, E. Gaetani, et al., A prototype evaluation of a tamper-resistant high performance blockchain-based transaction log for a distributed database, in: Dependable Computing Conference, 2017.
[25] A. Kiayias, A. Russell, B. David, et al., Ouroboros: A provably secure proof-of-stake blockchain protocol, in: International Cryptology Conference, 2017.
[26] X. Mao, Application of drug knowledge base based on big data technology in pediatric clinical drug use, Investig. Clín. 60 (5) (2019) 1268–1277.

Pengcheng Wei was born in Hechi, Guangxi, P.R. China in 1975. Wei earned his Ph.D in Engineering in 2008. Now he is the professor at Chongqing University of Education known for his work on computational intelligence, information security and big data analysis.

Wang Dahu was born in Jiangsu, Xuzhou, P.R. China, in 1969. He received the master's degree from Henan Polytechnic University, P.R. China. Now, he works in College of Electrical Engineering and Automation,Henan Polytechnic University. His research interests include virtual reality,machine vision, computer graphics and face recognition.

Yu Zhao was born in Chongqing, P.R. China in 1991. He received the Master degree from Shanghai Jiao Tong University, P.R. China. Now, he works in School of mathematics and Information Engineering, Chongqing University of Education. His research interests include computer vision, deep learning and machine leaning.

Sumarga Kumar Sah Tyagi currently works at Zhongyuan University of Technology, China. Sumarga does research in Green Cloud Communication Networks, Cloud RAN, 5G, Deep Learning, and IoT.

Dr. Neeraj Kumar received his Ph.D. in CSE from SMVD University, Katra (J & K), India, and was a postdoctoral research fellow in Coventry University, Coventry, UK. He is working as an Associate Professor in the Department of Computer Science and Engineering, Thapar Institute of Engineering & Technology, Patiala (Pb.), India since 2014.