 CMC, vol.58, no.1, pp.135-151, 2019

Blockchain-Based Trusted Electronic Records Preservation in Cloud Storage

Zhiliang Deng [1, 2], Yongjun Ren[3, 4], Yepeng Liu[3, 4], Xiang Yin[5], Zixuan Shen[3, 4] and Hye-Jin Kim[6, *]

Abstract: Cloud storage represents the trend of intensive, scale and specialization of information technology, which has changed the technical architecture and implementation method of electronic records management. Moreover, it will provide a convenient way to generate more advanced and efficient management of the electronic data records. However, in cloud storage environment, it is difficult to guarantee the trustworthiness of electronic records, which results in a series of severe challenges to electronic records management. Starting from the definition and specification of electronic records, this paper firstly analyzes the requirements of the trustworthiness in cloud storage during their long-term preservation according to the information security theory and subdivides the trustworthiness into the authenticity, integrity, usability, and reliability of electronic records in cloud storage. Moreover, this paper proposes the technology framework of preservation for trusted electronic records. Also, the technology of blockchain, proofs of retrievability, the open archival information system model and erasure code are adopted to protect these four security attributes, to guarantee the credibility of the electronic record.

Keywords: Trusted electronic record, blockchain, cloud storage.

1 Introduction

Electronic records, produced in production and living activities, have shown rapid growth all over the world, and have gradually become the primary source of recording human activities and an essential source of information resources. Electronic records as a memory of social activities not only has the function of saving data but also has the evidence value [Xie, Wang, Ma et al. (2017a, 2017b); Wang, Xu and Pei (2017)]. However, due to the

[1] School of Information and Control, Nanjing University of Information Science & Technology, Nanjing, 210044, China.

[2] Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, Nanjing, 210044, China.

[3] Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science & Technology, Nanjing, 210044, China.

[4] School of Computer and Software, Nanjing University of Information Science & Technology, Nanjing, 210044, China.

[5] College of Information Engineering, Yangzhou University, Yangzhou, 225127, China.

[6] Business Administration Research Institute, Sungshin W. University, Seoul, Korea.

[*] Corresponding Author: Hye-Jin Kim. Email: hye-jinkim@hotmail.com.

CMC. doi:10.32604/cmc.2019.02967 www.techscience.com/cmc

separation of the carrier and the information of electronic records, and the characteristics of being easy to change, flowing and vulnerable to the environment, many users are questioned on their originality, integrity, and evidentiary value. Thus, how to ensure the credibility of electronic records has become the research focus of electronic records [Xie, Wang, Ma et al. (2017a, 2017b); Qian (2014)].

Cloud storage eliminates the need for electronic record managers to physically install physical storage devices, paying only for the actually used storage. Moreover, the day-to-day maintenance of electronic records storage is the responsibility of the cloud storage provider, which significantly reduces the storage cost of electronic record managers [Ren, Shen, Wang et al. (2015); He, Yu, Zhang et al. (2017); Jiang, Zeadally, Ma et al. (2017)]. Also, cloud storage provides an electronic record integration service that allows users to access data from any network device, significantly expanding the sharing of electronic records and facilitating the management of the full lifecycle of electronic records [Shen, Zhou, Chen et al. (2017); Wei, Zhang and Ma (2018); Li, Niu, Kumari et al. (2018)]. Compared with the traditional storage methods, formed a considerable advantage, cloud storage will become the mainstream storage. However, in the cloud storage environment, the administrator of the electronic record loses the physical control over the electronic record, making it easier to copy the electronic record and tampering with the contents. Also, cloud storage servers may be intentionally deleted electronic records to save costs, resulting in irreparable damage to the owners of electronic records [Ren, Shen, Liu et al. (2016); Ge, Susilo, Wang et al. (2017); Shi, Wei, Wang et al. (2017)].

The Blockchain is the core support technology of digital cryptographic currency system represented by bitcoin. The core advantage of blockchain technology is its decentralization. It can implement point-to-point transactions based on decentralized credit in distributed systems where nodes do not trust each other by means of data encryption, time-stamping, distributed consensus and economic incentives, coordination and collaboration, so as to provide solutions to the prevalent issues of centralized agencies such as high cost, low efficiency and data storage unsafety [Yuan and Wang (2016); Jiang, Wei, Fu et al. (2016)].

Because of the excellent nature of blockchain-based solutions and the urgency of using blockchain as a trusted preservation solution to electronic records, there is a need for professionals to be more confident about the credibility of the technology of electronic records. Therefore, this article focuses on whether blockchain technology can indeed achieve this goal.

2 Related work

2.1 Trusted record

The concept of trusted electronic records originated earlier; representative ideas are: Professor Jiazhen Liu from Wuhan University pointed out: Trusted electronic record described the facts and events refer to the content of the electronic record is reliable, dependable and accurate. Moreover, the accuracy and truth of the content depending on the completeness of the record, the procedural control over the record formation, and the author's reliability. His primary concern is the content's credibility. Ronald Jantz of Rutgers University presented a conceptual framework of trustworthiness and pointed out

that trustworthiness should be maintained throughout the life of the electronic record.

InterPARES (International Cooperation Project to Ensure the Permanent Authenticity of Electronic Records), a large-scale international cooperation project, aims to study how to ensure the authenticity of electronic records for a long time. InterPARES has now entered the fourth phase of "InterPARES Trust" to explore the feasibility of electronic records under the network environment Permanent and trusted the phase as the theme [Xie, Wang, Ma et al. (2017a, 2017b); Wang, Xu and Pei (2017); Zeng, Dai, Li et al.(2018)].

International Standard ISO 16175 [ISO (2010)] (Information and documentation-Principles and functional requirements for records in electronic office environments) is a required standard devised by the international standard ISO 15489 for trusted electronic records. The standard is consists of three parts. The first part is an overview of ISO 16175. The second part consists of digital record management system guidelines and functional requirements of the ISO 16175. The third part is the ISO16175-3, which is a business system that provides record management guidelines and functional requirements. ISO 16175 sets forth the guiding principles and functional requirements for record management in an electronic office environment, as well as the specifications for designing electronic record management systems and business systems. ISO 16175 provides a stable demand for trusted electronic record management.

International Standard ISO/TR 26122 [ISO (2008)] (Work Process Analysis for Records) is a front-end control standard based on ISO 15489 for the establishment of a classification scheme for trusted electronic records. The standard provides two dimensions of the analytical work processed from the perspective of record creation, capture, and control, namely the job function analysis dimension and the workflow order analysis dimension. In the analysis of job functions, the standard provides the necessary steps of functional analysis, defines the analysis of job functions, business activities, business disposal methods and the establishment of functional analysis based on record classification scheme requirements. For the workflow sequence analysis, the standard provides the overall requirements of workflow analysis, defines the order of business processes in the workflow. Moreover, the results are analyzed, and the rules on how to identify and analyze variables provide controls in the workflow sequence.

Trusted electronic records are trustworthy and evidence-based. To realize the historical mission entrusted to the preservation of human memory, we need to manage and protect electronic records in the right way as emphasized in UNESCO's Joint declaration on archives. The correct way is to build management technology of trusted electronic record that is based on the standard practice of humanity and in conformity with the objective laws governing the management of records so that the evidence value of the electronic record can be maintained permanently. The system project is complicated and arduous and bound to become a historical research topic in the field of international records. Therefore, it is vital to pay full attention to explore and study the management of trusted electronic records.

2.2 Blockchain

The blockchain has the following five characteristics [Qian, Shao, Zhu et al. (2018); Wang, Xu and Pei (2017); He, Kumar, Wang et al. (2017); He, Yu and Zhang (2017)]. Firstly, it

is decentralized. The validation, bookkeeping, storage, maintenance, and transmission of data in blockchain are based on a distributed system structure. It uses pure mathematical methods instead of central institutions to build trust relationships between distributed nodes, thus forming a decentralized and trustworthy distributed system. Secondly, the blockchain is the time series data. Blockchain stores data in a chained block structure with time-stamping (add a time dimension to the data) with strong verifiability and traceability. The third is standard maintenance. The blockchain system uses a specific economic incentive mechanism to ensure that all nodes in the distributed system can participate in the data block verification process (such as mining in Bitcoin). The system picks up specific nodes by consensus algorithms to add new blocks to the blockchain. The fourth is programmability. Blockchain technology provides a flexible scripting code system that allows users to create advanced smart contracts, currencies, or other off-center. Finally, it is safe and trusted. Blockchain technology uses the principle of asymmetric cryptography to encrypt the data. At the same time, with the aid of the workload of each node in the distributed system, it can prove the mighty computing power formed by consensus algorithms to resist the external attacks. Moreover, it ensures the data in blockchain cannot be tampered with and cannot be forged. Thus, it has a higher safety. The infrastructure model of blockchain technology is shown in Fig. 1.

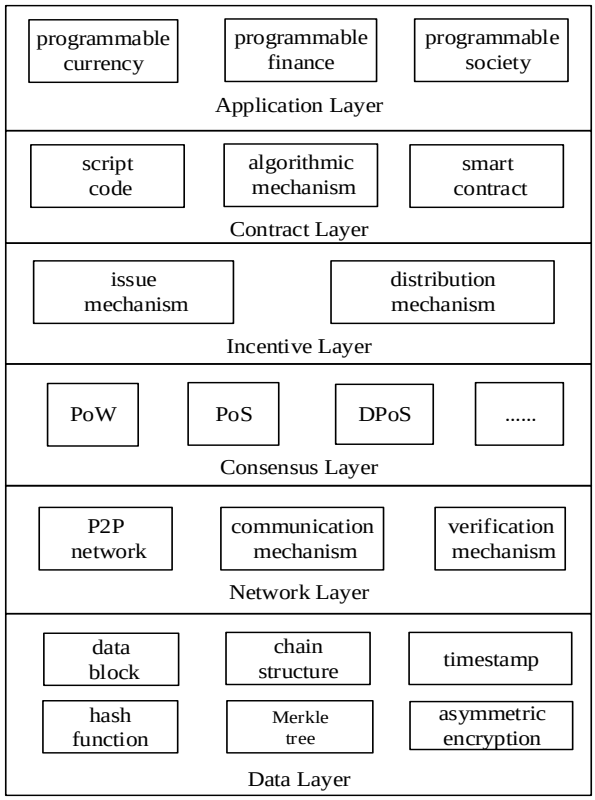

Figure 1: Architecture of blockchain

In general, the blockchain system consists of the data layer, network layer, consensus layer, incentive layer, contract layer, and application layer [Zhu, Gao, Shen et al. (2017); Wu, Li and Jiang (2017)]. In blockchain, the data layer encapsulates the underlying data blocks and related data encryption and time stamping techniques. Network layer includes distributed networking mechanism, data propagation mechanism, and data verification mechanism. Consensus layer encapsulates network nodes of various types of consensus algorithms. Incentive layer integrates economic factors into the blockchain technology system, mainly including the issuance mechanism and distribution mechanism of economic incentives. The contract layer encapsulates all kinds of scripts, algorithms, and smart contracts, which are the foundations for the programmable features of the blockchain. The application layer encapsulates various application scenarios and cases of the blockchain. In the model, the timestamp-based chain block structure, the consensus mechanism of distributed nodes, the economic incentive based on consensus power and the flexible and programmable smart contract are the most representative technology innovations of blockchain [Wang, Xu and Pei (2017)].

3 Problem statement

In the article, we define trusted electronic records as electronic records that are authentic, integrity, usable, and reliable according to the information security theory. They have evidence and trustworthiness. The four characteristics are the essential nature that electronic records must possess. Moreover, they are also the basis of trusted electronic records. The conceptual framework is shown in Fig. 2.

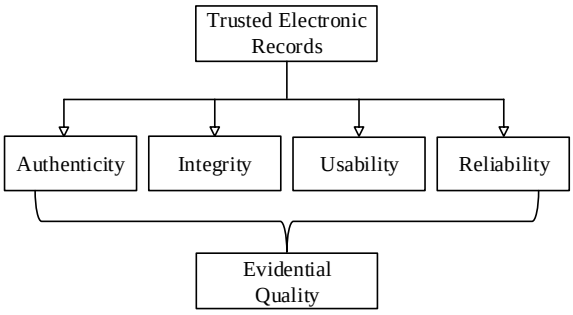

Figure 2: Essential attributes of trusted electronic records

3.1 Authenticity

Authenticity shall satisfy the correspondence of the record with the purpose of its text. The formation and transmission of the record coincide with its established formation and sender. Moreover, the formations or transmissions of the record are consistent with the fixed time. To ensure the authenticity of electronic records, corresponding policies and procedures need to be formulated to control the formation, collection, circulation, custody and disposal of electronic records so as to ensure that the creators, participants, and custodians of the electronic records are trustworthy, and prevent unauthorized tampering with the record, delete, use and hide, to ensure that the record is also accurate and reliable at the same time.

3.2 Integrity

Integrity refers to the record is complete and unaltered. The integrity of electronic records requires that the content, logical structure and background information of each electronic record be complete and must protect the circulation from being tampered. Metadata is the best solution for protecting the integrity of records. Also, records should be established throughout the life cycle of scientific and rational procedures and guidelines. The content should include user authentication, monitoring, authorization destruction, record annotation, tracking audit, system disaster recovery and other aspects to ensure that electronic records are entirely trusted.

3.3 Reliability

Reliability refers to the content of the record is trusted, can accurately reflect the affairs, activities or facts, and the follow-up affairs or activities can be based on its process. Reliability requires that electronic records, as a manifestation of factual activity, are trustworthy, based on the completeness of the record's form, and the control over the record formation process. The source of the recording environment is the crucial fact to ensure record reliable, including reliable formation, reliable generation system, reliable formation process, and reliable application software.

3.4 Usability

Usability refers to the record can be found, retrieved, displayed and understood. The availability of electronic records indicates that the records are directly related to the business activities and business processes that formed them. Therefore, trusted electronic records require that the electronic records contain the background information on the formation and utilization of the records in terms of "usability", so as to restore the authenticity of the business activities record, and at the same time, the proficiency of the function and role of this record in the context of significant business activities allows users to determine their trustworthy value.

3.5 Evidence value

Archives, the original record of decision-making, action, and memory, are the reliable evidence of administrative, cultural and intellectual activities. The user thinks the record is the only reliable evidence, which shows that the user makes sure that the evidentiary nature of the electronic record. Based on ensuring the authenticity, completeness, reliability, and usability of electronic records, the ultimate value of accurate records is precisely the goal of trusted electronic records.

To sum up, the authenticity of electronic records makes the electronic records become the recording form and original evidence of transactional activity worthy of users' reliance. The integrity and reliability of electronic records ensure that the users only trust the contents of electronic records. The availability of the electronic records ensures that the user can understand the use of the trust. It can be said that trusted electronic records with integrated quality based on the four physical attributes of authenticity, completeness,

reliability, and usability of electronic records. The evidence integration goals are the activities of pursuing the intrinsic value of electronic records.

4 Preliminaries

4.1 Open archival information system

4.1.1 OAIS model

The Open Archival Information System (OAIS) model is the most widely accepted international standard for the long-term preservation of digital information. It provides the necessary framework for the construction of trusted data storage and provides the most reference for the long-term preservation of electronic records [Wu, Li and Jiang (2017)]. The Standard guarantee essential contributions of OAIS are as follows. According to the characteristics of digital information and laws of life movement, six basic functional models are provided, and the information objects that need to be stored for a long time are classified in detail according to their functions. The proposed information model has the following significant reference value.

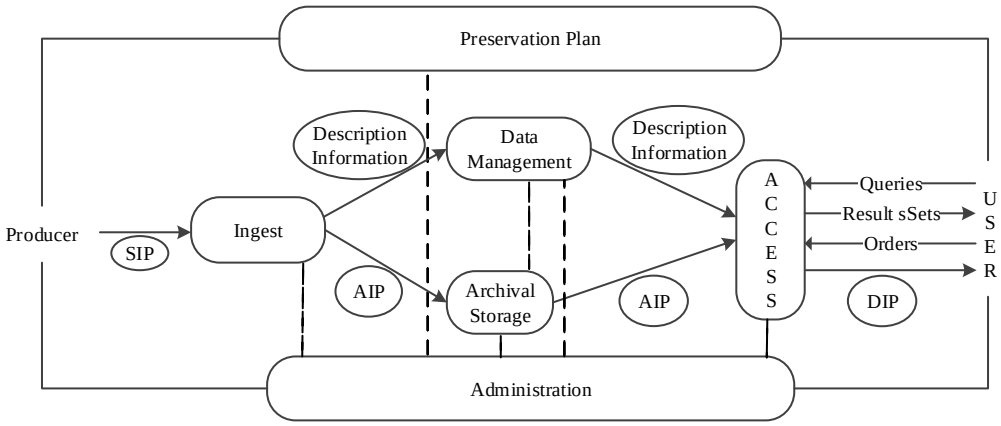

Figure 3: OAIS model

Receive.

It means that OAIS receives submittal information package (SIP) from the information producer and prepares the content so that the submitted information can be stored and managed in the file system. The function includes: receiving and submitting the information package for quality confirmation, generating an archive information package (AIP) conforming to the file system data format and file standard, extracting the description information from the archive information package for storing in the file database, and coordinating the storage of the file and data management which changed.

File storage.

It mainly to store, maintain and retrieve archive information package. This includes receiving archived packets from the ingest function, saving them to the storage system, managing the storage system's organization, updating the media that holds the file information, performing routine checks and maintenance, and providing disaster recovery

ability and provide archive information package to achieve file extraction.

Data management.

The purpose of data management is to implant, maintain, and access management information that identifies and records descriptive information about archival holdings and the archival system. This functionality includes managing the archive database (maintaining definitions of conceptual models and views in the database, maintaining referential integrity of the system, etc.), performing database updates (loading new descriptive information or archive management data), providing query capabilities to management data to produce results Set, and generate query reports from these search result sets.

Administration.

For the overall file system to provide management. The primary functions include negotiating with the information producer to decide on the submission of the agreement, the submission of the information to be checked to ensure that the submission is in accordance with the file standard, the system software environment is configured and managed, as well as monitoring and improving the file system health and record, report the contents of the file to transplant and modify, establish and maintain file standards, provide customer support, stimulate storage requests and other functions.

Preservation planning.

It mainly to monitor the environment of OAIS, provide recommendations to ensure long-term preservation of information stored in OAIS program, even after the original computer environment degradation, it can still be accessed by the corresponding target user (DC). The primary functions include assessing what is stored in the file system, periodically providing archival information advice and porting current file holdings, advising on file system standards and policies, monitoring service environment and target user (DC) service needs and knowledge Background changes. It also includes templates for designing information packages (IP), providing design assistance and detailed evaluations, and professionalizing these templates to create a specific commit and archive information packages, developing migration plans, developing migration software prototypes and corresponding test plans achieve management functions such as transplantation goals.

Access.

The goal is to support consumers so that consumers can determine, understand, locate, and access information stored in OAIS so that consumers can request and receive information products. The function includes communicating with consumers and receiving requests, managing and controlling specific specially protected information to restrict their access, coordinating the execution of the request to achieve the complete response of the request, and sending the result set of the distributed information package to the consumers.

4.1.2 Information packages in OAIS

OAIS proposed the concept of an Information Package to illustrate the filing process of archivists to OAIS and the distribution of OAIS to archivists. A packet is a container containing two kinds of information objects called Content Information and Preservation Description Information (PDI). The content information and the save description

information are encapsulated and identified by the packaged information, and the information packet needs to be described by the description information so that the packaged information can be understood and discovered.

OAIS makes a distinction between the packets submitted to it, the packets it stores, and the packets it distributes to other consumers. These packets are called the Submission Information Package (SIP), the archive packet Archival Information Package (AIP) and Dissemination Information Package (DIP).

(1) The submission packet (SIP) is a packet provided by the information producer to OAIS. Its format and details are usually the results of negotiations between producers and the OAIS system. Most submission packets usually consist of some content information and related save description information, but it may require multiple submission packets to form complete content information and corresponding save description information to form further an archive information package, one submission packet may also include information that needs to be included in multiple archive packets. The packaging information (PI) corresponding to the submitted package usually exists in some form.

(2) In OAIS, one or more commit packages need to be converted into one or more archive packages for saving. Archive information package has a complete set of saving description information and related content information. A submission packet may also consist of several other submission packets. Archiving package information (PI) requires compliance with OAIS internal standards and subject to OAIS management.

(3) At the request of the consumer, OAIS needs to provide all or a portion of the content of an archived package to the consumer in the form of a distribution package (DIP). A distribution package may also include multiple archive packages, but does not necessarily have a complete save description. The packaging information (PI) to distribute a package needs to somehow appear in a way that allows the consumer to discern the desired information. Depending on the distribution media and the needs of consumers, packaging information can exist in many ways.

4.2 Erasure code

Erasure code technology is used more and more in large-scale storage system design due to its strong fault tolerance and high space utilization. For example, RAID (redundant array of independent disks), Robustore and OceanStoreetc are all fault-tolerant storage systems based on erasure codes.

Erasure code [Wang, Xu and Pei (2017); Wang, Zhao and Hou (2012); Liu, Feng and Li (2014)] can be represented by triplets (n, k, k'), where $n>k' \geqslant k$. An (n, k, k')-erasure code divides a data object O of size M into k data blocks o_1, o_2, ..., o_k of size M/k, then these data blocks are operated through the corresponding coding algorithm to obtain n coding blocks c_1, c_2, ..., c_n, and ensure that any k' bits of the n code blocks can be decoded to recover the original data object O.

At present, the erasure code adopted by the storage system is linear, and the article only focuses on the linear erasure code. As shown in Eq. 1, in the linear erasure code, each coding block c_i can be expressed as a linear combination of all the data blocks, where $g_{ij} \in F_q$, F_q is a Galois Field containing q elements.

$$(g_{i1}, g_{i2}, \ldots, g_{ik}) \times \begin{pmatrix} o_1 \\ o_2 \\ \cdots \\ o_k \end{pmatrix} = c_i, i = 1, 2, \ldots, n \quad (1)$$

As with the encoding process, in the case of linear erasure codes, the repair of a failed block is also done by linearly combining the available blocks. The specific error-correcting code design dictates the choice of available blocks and the calculation of the corresponding combination coefficients. For example, for a standard RS code, *k* blocks need to be downloaded to repair any one block. For instance, most recently introduced improved error correction codes, such as locally repairable code [Fu, Wen, Ma et al. (2018)]. The number of blocks required for repair in most failure cases is less than *k*, which is higher than *k* in the case of multiple failures.

4.3 Proofs of retrievability

For the first time, Juels and Kaliski [Juels and Kaliski (2017)] proposed proofs of retrievability (POR) in a cloud storage environment to ensure the recoverability of cloud storage data. Compared to the provable data possession (PDP) [Ren, Shen, Wang et al. (2015); Juels and Kaliski Jr (2007)] only to ensure the integrity of the data, POR data recovery is more stringent. To tolerate a certain percentage of lost or corrupted data, POR uses error-correcting code technology to encode the data and recover the data by adding redundancy. The (*n, k, d*)-erasure code is chosen, where the minimum distance *d* is an even number, meaning that its error correction capability is *d*/2. The data processing algorithm includes the following steps.

(1) Coding and Encryption: The data *M* is divided into *b* blocks (*b* is a multiple of *k*), and the erasure code is calculated for every *k* data blocks. The redundant data generated here is also divided to the same length as the original data block. The resulting encoded data is *M'*, which contains *bn*/*k* data blocks m_i'. Then encrypt each data block m_i' using symmetric encryption technique to obtain m_i''.

(2) Generate Sentinels: Generate η random "sentinels" using a one-way function that has the same length as the data block and can, therefore, be considered a dummy data block.

(3) Replacement: Replace the data blocks generated in Steps 1 and 2 with the following permutation algorithm and send the permuted complete data to the cloud storage server.

When performing an integrity query, the verifier checks that the random sentinels, which are well-preserved, rather than directly querying and validating the real data block to determine the integrity (recoverability) of the cloud data. Since there are a total of η random sentinels, the scheme can only allow a limited number of (η/*k* times) non-intersecting integrity verifications, if each integrity query checks *k* sentinels. The following is a brief review of the POR scheme based on computational Diffie-Hellman problems and then explores its advantages over previous schemes. The data processing algorithm first encodes *M* to obtain *M*', and then every block of *M*' is divided into *s* pieces, that is, m_i = ($m_{i,1}$, $m_{i,2}$, ..., $m_{i,s}$). Calculate the metadata σ_i for each data block m_i as follows:

$$\sigma_i = \left(H(name||i) \times \prod_{j=1}^{s} u_j^{m_{i,j}}\right)^{\alpha} \quad (2)$$

Where $\mu = (\mu_1, \mu_2, \ldots, \mu_s)$ and metadata σ is as response *P*:

$$\mu_j = \sum_{(i,v_i)\in C} v_i m_{i,j}\ (1 \le j \le s), \sigma = \prod_{(i,v_i)\in C} \sigma_i^{v_i} \quad (3)$$

Validating the response *P* requires bilinear pairings to be performed twice:

$$e(\sigma,g) = e(\prod_{(i,v_i)\in C} H(name||i)^{v_i} \times \prod_{j=1}^{s} u_j^{\mu_j}, v) \quad (4)$$

Here, v is the user's public key corresponding to α.

Where α is the private key of the user and u_j ($1 \le j \le s$) is a parameter randomly selected from the bilinear group *G*. The factor contained $(\prod_{j=1}^{s} u_j^{m_{i,j}})^{\alpha}$ in metadata σ_i here also supports aggregation operations, so the cloud storage server can generate partially aggregated responses during the integrity verification phase. The algorithm also signs the data name, the number of data blocks and the parameter u_j to obtain the data label τ.

To verify the integrity of the data, the verifier submits a query challenge $C=\{(i, v_i)\}$, containing the randomly chosen data block number i and the corresponding coefficient v_i. The cloud server calculates aggregated data blocks.

5 Trusted record preservation based on blockchain

5.1 Overall of our mechanism

Blockchain-based solutions work by comparing the hash of the original digitally signed hash with the digitally signed hash stored on the blockchain. Any change in the protocol of the original digital record or protocol in the log or a small change in the implementation of the various components will cause the corresponding entries in the blockchain to change. Since digitally signed versions on the blockchain cannot be reverse-engineered to produce copies of the original record, the original must always be retained so that the original can be re-hashed and digitally signed for comparison. According to ISO 14721, the organization and level of investment in retaining originals should not be overlooked. Trustworthy digital repositories and additional elements such as technical, policy and institutional capabilities should be established for record keeping, record storage, data management, access, disseminate and migrate to new media and forms. However, there is not much research on how to deal with the preservation of original records. To solve the problem, the OAIS model which is based on cloud storage architecture to store the original record is proposed in the paper. To ensure the integrity of electronic records in cloud storage, we use the traditional PDP or POR to periodically check the integrity of the original records stored on the cloud servers. To further improve the reliability of the original records, we adopt erasure code technology. When the original records in cloud storage are defected or even lost, the redundant record blocks in erasure code are utilized to recover the missing record. The use of erasure code technology will ensure the reliability of the original records.

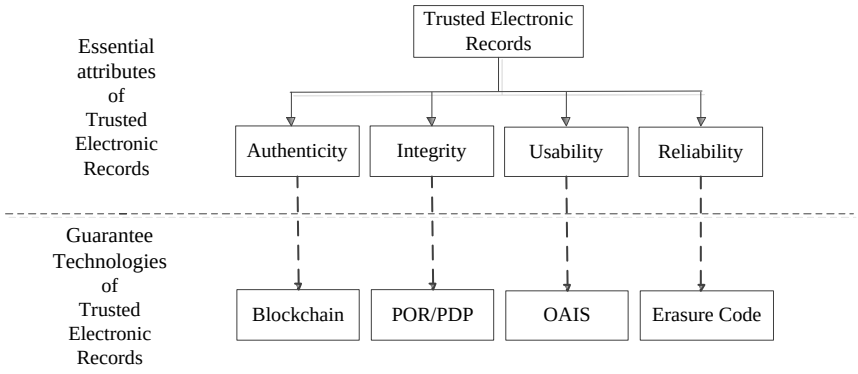

Figure 4: Technical guarantee of trusted electronic records

5.2 Electronic records based on regeneration code

The reproduction code is used to encode the original electronic record in the article. Storing record encoding to *n* nodes, each node storing an encoded record of size α. When one node fails, *d* nodes ($k \le d \le n$-1) of the remaining *n*-1 nodes are connected, and record size β ($\beta \le \alpha$) is downloaded from each node to be repaired. The *d* nodes that help to fix are called Help Nodes, and the total data download $d\beta$ for repair is called Repair Bandwidth. Achieve the minimum bandwidth *d* by increasing the amount of storage per node, and the minimum bandwidth regeneration point (MBR) can be obtained from the following expression:

$$\alpha = \frac{2Bd}{k(2d-k+1)}, \beta = \frac{2Bd}{k(2d-k+1)} \tag{5}$$

Here, *B* refers to the size of the source record, α refers to the storage overhead of nodes, β is the communication overhead of failover, *d* refers to the number of surviving nodes, and *k* refers to the number of nodes needed for record reconstruction.

Accordingly, the regenerative code generation matrix *M* has $n\alpha$ rows and *B* columns, and $n\alpha$ row vectors correspond to the $n\alpha$ redundant data blocks. If *n* nodes corresponding to the generated evidence M are divided into *n* sub-matrices M_i(*i*=1,..., *n*) according to rows, each sub-matrix corresponds to one cloud storage service and $M=[M_1, M_2, ..., M_n]^T$. M_i is a generator matrix of redundant data blocks in the *i*th cloud storage service, that is, α redundant data blocks ($E_{(i-1)\alpha+1}, \ldots, E_{i\alpha}$) in the cloud storage service *i* are equal to the product of the submatrix M_i and the ($D_1, D_2, ..., D_B$) of the original record block, i.e.

$$\begin{bmatrix} E_{(i-1)\alpha+1} \\ ... \\ E_{i\alpha} \end{bmatrix} = M_i \times \begin{bmatrix} D_1 \\ ... \\ D_B \end{bmatrix} \tag{6}$$

The regenerative code is also the MDS code and must also meet the MDS nature. Then all the redundant record blocks in any *k* of *n* cloud storage can recover the original record. Similarly, the matrix rank formed by combining any *k* sub-matrixes in *n* sub-matrices M_i is equal to *B*, and then the MDS property is satisfied.

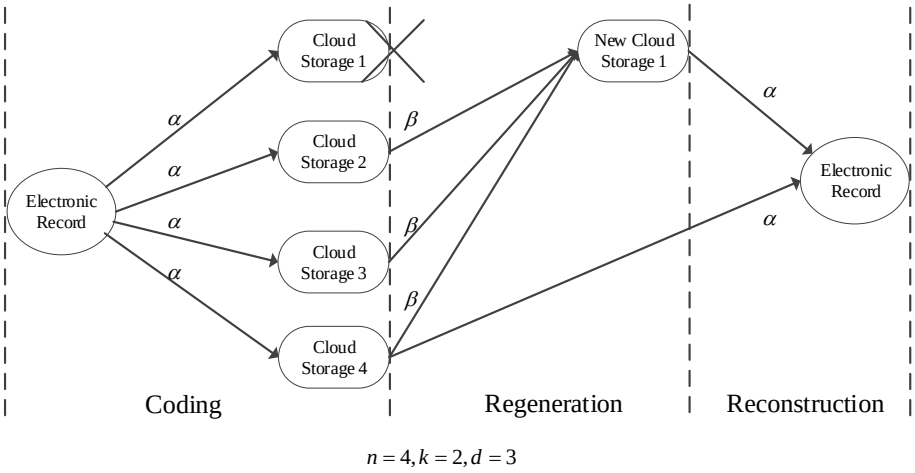

Figure 5: Coding, regeneration and reconstruction of electronic record based on regeneration code

5.3 Repair of electronic record

Encoded record with MDS properties allow for up to *n-k* cloud storage services to simultaneously lose data, but only one case of a cloud storage service losing its stored redundant data block ($n=d+1$) is discussed here. Supposing that the cloud storage service number of the redundant record block lost is *i*, the redundant record blocks ($E_{(i-1)\alpha+1}, \cdots, E_{i\alpha}$) in the cloud storage are lost and a generator matrix M_i is no longer available. The record restoration process includes: generating a repair of the sub-matrix M_i and repairing the redundant record block. Generating a submatrix repair will generate a new M_i*'* instead of an unavailable M_i so that the MDS properties of the generator matrix will continue to be maintained. The pre-restoration generator matrix will be denoted as *M* and the post generator will be denoted as *M'*. The repair of the redundant record blocks will really repair the redundant record blocks, and the successful repair of the previous generation sub-matrix ensures that the redundant record blocks to be generated also meet the MDS nature.

The idea of repairing M_i is to find β row vectors from *d* participating sub-matrices (M_1, ..., M_{i-1}, M_{i+1}, ..., M_n), combine these $d\beta$ row vectors linearly into α. The new sub-matrix M_i' of the row vector is used to replace the originally failed M_i so that the new generator matrix M_i' satisfies the MDS property. The specific steps are as follows:

(1) At first, downloads the metadata file F_{meta} of the record from any cloud storage service to obtain the *M* and *n, k, α, β, d, B* and other parameters for encoding;

(2) Sub-matrix M_i to randomly select $d\beta$ row vectors from the remaining $d=n-1$ sub-matrices to form a *CM* (candidate matrix), and each sub-matrix randomly selects β row vectors;

(3) Randomly generate a $\alpha \times d\beta$ repair matrix is used to combine candidate matrices CM into new sub-matrices M_i*'*;

(4) The new sub-matrix *Mi'* is the product of the repair matrix and the candidate matrix,

that is, $M'_{i\alpha\times B}=RM_{\alpha\times d\beta}\times CM_{d\beta\times B}$;

(5) Replacing the original sub-matrix M_i with M_i *'*, if the MDS property of the new generated matrix M_i after the replacement is maintained, the sub-matrix restoration is successful; if the MDS property is not maintained, returning to Step (2). The next round of repair attempts and reselects the candidate matrix and the repair matrix to generate a new sub-matrix M_i *'* until it satisfies the MDS properties of the newly generated M'.

5.4 Trusted records preservation based on OAIS in blockchain under cloud storage

The maximum amount of information that can be stored by each node in the current blockchain is 2 MB, which cannot meet the storage requirements of massive original electronic records. So, we propose storing the hash value of the original record address in each node of the blockchain instead of the original record itself; while the original record is stored using the existing cloud storage mechanism.

According to the OAIS model, the owner of the record generates SIP, AIP and DIP, and the corresponding metadata, from the encoded electronic record obtained in the previous section. The article proposes storing these parts in a cloud storage server, storing the hash value of the storage address in the recording layer of each chunk and building Merkle hash trees.

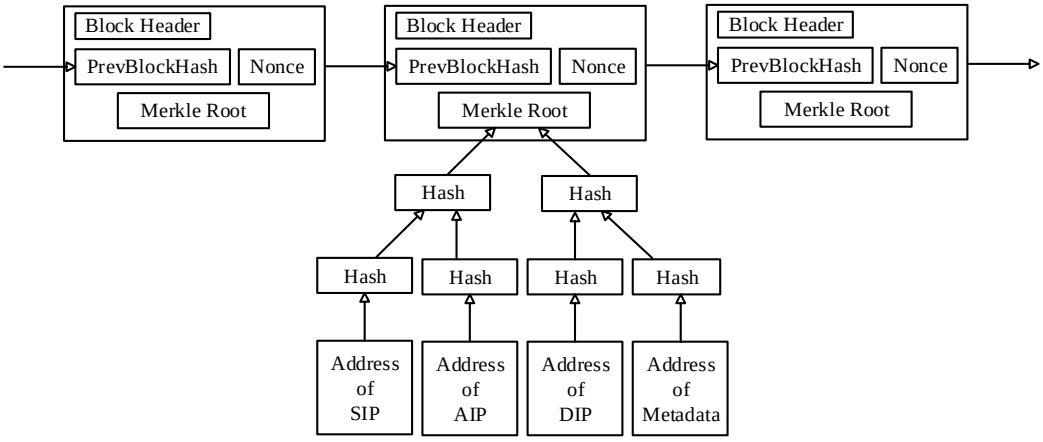

Figure 6: Trusted records preservation in blockchain

5.5 Record storage as transaction

Store transactions allow participants to store blocks of records. Stores need to enter chunks c, their certifiers $\sigma_1,\ldots,\sigma_n$, the client's public key pk_u. Storage is included in-store deals, disappear from the network and can no longer be spent. This payment is necessary to avoid a denial of service attack because an attacker can upload an arbitrary amount of data blocks free of charge, surpassing the available storage space on the network. Moreover, repayments are the form of protection against inflation. As the number of available decreases, the value of the remaining increases as only limited exist at any one time.

The miner can choose to store the chunk together with its authenticator so that a PoR can

be created on that chunk to generate a new chunk. Moreover, the miner needs to store the public pk_u of the transaction issuer for verification. The miners do not need to store all the files and may not even be able to do that. The motivation for storing files is economical, as storage allows you to tap into new blocks in the blockchain and collect mining rewards. Of course, the storage guarantee can be increased arbitrarily by applying the appropriate delete code on the file to be uploaded. In addition to these financial incentives, there is no further mechanism to increase storage assurance.

Chunk storage time is linearly dependent on the amount spent on the issue of the store's transaction. After the storage time has elapsed, the block containing the PoR for that particular block is no longer considered valid. Assuming that most miners do not accept such blocks, there is no incentive to store the blocks. Blockchain has provided loose time synchronization so that all miners can agree on when the required shelf life will be.

6 Conclusion

The paper analyzes the requirements of the trustworthiness in cloud storage during their long-term preservation according to the information security theory and subdivides the trustworthiness into the authenticity, integrity, usability, and reliability of electronic records in cloud storage. Moreover, the technology of blockchain, proofs of retrievability, the open archival information system model and erasure code are adopted to protect these four security attributes, to guarantee the credibility of the electronic record.

Acknowledgment: This work is supported by the NSFC (No. 61772280, 61772454, 6171101570, and 61702236), Natural Science Foundation of Jiangsu Province under grant No. BK20150460, the Changzhou Sci&Tech Program (No. CJ20179027), and the PAPD fund from NUIST. Prof. Hye-Jin Kim is the corresponding author.

References

Fu, Y.; Wen, S.; Ma, L.; Shu, J. (2018): Survey on single disk failure recovery methods for erasure coded storage systems. *Journal of Computer Research and Development*, vol. 55, no. 1, pp. 1-13.

Ge, C.; Susilo, W.; Wang, J.; Fang, L. (2017): Identity-based conditional proxy re-encryption with fine grain policy. *Computer Standards & Interfaces*, vol. 52, pp. 1-9.

He, D.; Kumar, N.; Wang, H.; Wang, L.; Choo, K. (2017): Privacy-preserving certificateless provable data possession scheme for big data storage on cloud. *Applied Mathematics and Computation*, vol. 314, no. 12, pp. 31-43.

He, P.; Yu, G.; Zhang, Y.; Bao, Y. (2017): Survey on blockchain technology and its application prospect. *Computer Science*, vol. 44, no. 4, pp. 1-7.

ISO (2008): Information and documentation-work process analysis for records (ISO/TR 26122). https://www.iso.org/standard/43391.html.

ISO (2010): Information and documentation-principles and functional requirements for records in electronic office environments (ISO 16175). https://www.iso.org/standard/55790.html.

Juels, A.; Kaliski Jr, B. S. (2007): PORs: Proofs of retrievability for large files. *Proceedings of the 2007 ACM Conference on Computer and Communications Security*, vol. 2007, pp. 598-609.

Jiang, Q.; Wei, F.; Fu, S.; Ma, J.; Li, G. et al. (2016): Robust extended chaotic maps-based three-factor authentication scheme preserving biometric template privacy. *Nonlinear Dynamics*, vol. 83, no. 4, pp. 2085-2101.

Jiang, Q.; Zeadally, S.; Ma, J.; He, D. (2017): Lightweight three-factor authentication and key agreement protocol for internet-integrated wireless sensor networks. *IEEE Access*, vol. 5, pp. 3376-3392.

Li, X.; Niu, J.; Kumari, S.; Wu, F.; Sangaiah, A. K. et al. (2018): A three-factor anonymous authentication scheme for wireless sensor networks in internet of things environments. *Network and Computer Applications*, vol. 103, pp. 194-204.

Liu, Q.; Feng, D.; Li, B. (2014): Ustor: Cloud storage system based on regenerating codes. *Journal of Communications*, vol. 35, no. 4, pp. 166-173.

Qian, W.; Shao, Q.; Zhu, Y.; Jin, C.; Zhou, A. (2018): Research problems and methods in blockchain and trusted data management. *Journal of Software*, vol. 29, no. 1, pp. 150-159.

Qian, Y. (2014): Study on the long-term preservation standard of trusted electronic records in China. *Archives Science Bulletin*, no. 3, pp. 75-79.

Ren, Y.; Shen, J.; Wang, J.; Han, J.; Lee, S. (2015): Mutual verifiable provable data auditing in public cloud storage. *Journal of Internet Technology*, vol. 16, no. 2, pp. 317-323.

Ren, Y.; Shen, J.; Liu, D.; Wang, J.; Kim, J. (2016): Evidential quality preserving of electronic record in cloud storage. *Journal of Internet Technology*, vol. 17, no. 6, pp. 1125-1132.

Shen, J.; Zhou, T.; Chen, X.; Li, J.; Susilo, W. (2017): Anonymous and traceable group data sharing in cloud computing. *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 4, pp. 912-925.

Shi, W.; Wei, W.; Wang, J.; Zhao, Q.; Lin, Z. et al. (2017): A verifiable sealed-bid multi-qualitative-attribute based auction scheme in the semi-honest model. *IEEE Access*, vol. 5, pp. 12380-12388.

Wang, J.; Cao, J.; Ji, S.; Park, J. H. (2017): Energy efficient cluster-based dynamic routes adjustment approach for wireless sensor networks with mobile sinks. *Journal of Supercomputing*, vol. 73, no. 7, pp. 3277-3290.

Wang, J.; Cao, Y.; Li, B.; Kim, H.; Lee, S. (2017): Particle swarm optimization based clustering algorithm with mobile sink for WSNs. *Future Generation Computer Systems*, vol. 76, pp. 452-457.

Wang, Y.; Zhao, Y. L.; Hou, F. (2012): Minimum bandwidth regeneration code of distributed storage system. *Journal of Chinese Computer Systems*, vol. 33, no. 8, pp. 1710-1714.

Wang, Y.; Xu, F.; Pei, X. (2017): Research on erasure code-based fault-tolerant technology for distributed storage. *Chinese Journal of Computers*, vol. 40, no. 1, pp. 236-255.

Wei, F.; Zhang, R.; Ma, C. (2018): A provably secure anonymous two-factor authenticated key exchange protocol for cloud computing. *Fundamenta Informaticae*, vol. 157, no. 1-2, pp. 201-220.

Wu, Z.; Li, W.; Jiang, S. (2017): Research and application of the provenance information packaging strategy based on the OAIS information model. *Library and Information Service*, vol. 61, no. 18, pp. 113-118.

Xie, L.; Wang, J.; Ma, L. (2017a): Trusting records: findings of team Asia InterPARES. *Archives Science Study*, vol. 4, pp. 8-13.

Xie, L.; Wang, J.; Ma, L. (2017b): The project of InterPARES: where it has been and where it is going. *Archives Science Study*, vol. 4, pp. 14-20.

Yuan, Y.; Wang, F. (2016): Blockchain: the state of the art and future trends. *Acta Automatica Sinica*, vol. 42, no. 4, pp. 481-494.

Zeng, D.; Dai, Y.; Li, F.; Sherratt, R.S.; Wang, J. (2018): Adversarial learning for distant supervised relation extraction. *Computers, Materials & Continua*, vol. 55, no. 1, pp. 121-136.

Zhu, L.; Gao, F.; Shen, M.; Li, Y.; Zheng, B. et al. (2017): Survey on privacy preserving techniques for blockchain technology. *Journal of Computer Research and Development*, vol. 54, no. 10, pp. 2170-2186.