Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Research on Copyright Protection and Data Security Based on Blockchain Technology in Digital Library

Peng Xu[1, ✉]

[1] *Library, South-Central Minzu University, Wuhan, Hubei, 430074, China*

ABSTRACT

Currently, digital libraries face challenges in piracy and illegal distribution, data and privacy security, digital content identification and traceability. In this paper, we design a blockchain-based copyright protection system for digital libraries to provide true and reliable digital copyright information for libraries and users, and to ensure the security of data information stored in the digital copyright registration system. Firstly, we classify blockchain and analyze in detail the three core technical principles of consensus mechanism, cryptography principle, and hash algorithm. Then design the copyright registration protection system that contains the functions of unique authentication of digital work copyright, IPFS distributed storage, and privacy data encryption. The designed algorithm is tested for performance and the service performance of this paper's scheme is analyzed in real applications, and it is found that the throughput performance of this paper's algorithm when the number of nodes ranges from 4 to 20 is on average 36.19% more than that of the PBFT consensus algorithm, and 55.92% more than that of the RBFT consensus algorithm. When there are 5000 digital resource feature vectors in the system database, the time required for similarity retrieval is only 0.523s, which meets the requirements of the system's non-functional needs for similarity retrieval runtime, and realizes a good balance between the operational efficiency of digital libraries and security. The research has practical reference significance for the application of blockchain technology in the field of digital copyright protection.

Keywords: digital copyright, blockchain, cryptography principle, hash algorithm, data security

1. Introduction

Against the background of the growing maturity of digital technology and the strong demand for social resources, digital libraries and the development of large-scale digital utilization of books have encountered institutional bottlenecks. The copyright law system constructed on the basis of printing technology is difficult to apply to the new digital way of accessing and utilizing works, and the

✉ Corresponding author.
E-mail address: 18571751983@163.com (P. Xu).
Received 20 July 2024; Revised 10 October 2024; Accepted 05 February 2025; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-188

traditional copyright authorization system has already hindered the development of digital libraries and the realization of the value of copyright, and the existing interest pattern of digital libraries and copyright holders has been broken [1-3]. There is an urgent need to introduce relevant technologies and methods to build a harmonious digital copyright ecosystem, help the construction of informationization, and do a good job of balancing the interests of libraries and copyright holders [4-5].

Library digital copyright management work mainly covers two aspects. The first is to carry out the use management of digital copyright, that is, to deal with the relationship between copyright production and consumption, libraries should clearly define the specific areas of digital copyright use and related licensing matters, and improve the integrated digital copyright use mechanism [6-8]. Secondly, it is to carry out the rights and interests management of digital copyrights, that is, to deal with the relationship between copyright revenue and supervision, the library should clearly define the conditions of use of digital copyrights, the scope of attribution, and supervise the improper use of digital copyrights in an all-round way and without any dead angle [9-11]. However, with the continuous development and maturity of blockchain technology, its application on digital library copyright protection has become more and more widespread [12]. Blockchain technology, as a distributed database system based on the principle of cryptography, has the characteristics of tampering, traceability, and trustworthiness, and is considered to be one of the most influential disruptive technologies after the Internet [13-16]. The application of blockchain technology in the whole chain management of digital copyright in libraries is conducive to improving the service economy and legal fairness of digital resources, and safeguarding the economic gains of copyright holders [17-18]. In addition, for the current digital library copyright protection method in which different users own the same content copyright at the same time, the blockchain technology system can make use of its decentralized and tamper-proof features to build a decentralized digital library copyright protection platform, so as to solve the problem of privilege management between users and users [19-22]. It can be seen that the integration of the use of blockchain technology in the digital copyright management work of the library is conducive to the realization of copyright protection, reduces the high cost and complexity of copyright management, makes the digital copyright transaction more diversified, and is conducive to the innovation of copyright management mode [23-26].

With the continuous development of digital libraries, they face the problem of copyright protection, which not only affects the development of digital libraries, but also has a negative impact on social culture, in view of this, a large number of scholars have carried out research on the protection of digital copyright in libraries based on blockchain technology. Literature [27] analyzes the obstacles encountered by the traditional digital copyright management system in managing library-related resources, and shows that the unique advantages of blockchain technology can improve the security, accessibility and preservation of the copyright of digital resources in libraries, which provides a perfect solution to the proposed problem. Literature [28] explores the potential of blockchain technology in library management, which possesses transparent and tamper-proof features that can effectively optimize operations in areas such as digital rights management and enhance data security in the library ecosystem. Literature [29] shows that the data security of digital library systems is closely related to copyright information, and the application of blockchain technology will be beneficial in solving the data security and privacy security challenges that exist in the digitalization of libraries and provide accessibility for authenticated users. Literature [30] used blockchain technology to construct a digital copyright management model for university libraries, which effectively protects the legitimate rights and interests of copyright holders through the whole process of digital copyright creation monitoring and timely updating of information. Literature [31] pointed out that the digital ledger supported by blockchain technology can effectively track the intellectual property rights and ownership of relevant digital resources, and provide native protection to prevent data tampering,

which plays an important role in digital library copyright management. Literature [32] designed a digital copyright transaction system based on blockchain technology, which, by virtue of the decentralization, security, and traceability features of the technology, guarantees the reliability of the whole chain of digital copyright transactions in digital libraries and safeguards the rights and interests of digital copyright subjects. Literature [33] introduces the application value embodied in the application of blockchain and semantic web to library management, which effectively improves the management efficiency of intellectual property and digital rights, and also analyzes the numerous upcoming restrictions and challenges. Literature [34] believes that intellectual property is a key component of today's social and economic development, and the proposal of blockchain technology deepens the protection of library-related intellectual property information, which has a positive impact on the development of related industries. Although blockchain technology can effectively solve the current problems in copyright protection of digital libraries, it is still necessary to further optimize and improve blockchain technology in order to ensure that good results can be achieved after blockchain technology is applied to digital libraries.

In this paper, based on blockchain and smart contract, IPFS is used to realize the digital library copyright protection system of work copyright storage. Firstly, feature extraction and similarity detection are carried out on digital works to give the unique authentication of digital work copyright. Then the AES algorithm is used to encrypt the voluminous data, and the SM2 algorithm is used to manage the AES key, combining the advantages of both to solve the balance between efficiency and security. The designed algorithm is compared with PBFT and RBFT algorithms to explore the security of the data and analyze its request error rate and copyright similarity retrieval time in practical applications.

2. Digital libraries based on blockchain technology

2.1. Digital Copyright and Blockchain Technology

2.1.1. Digital Rights Characteristics of Libraries. Library digital copyright refers to the personal and property rights granted by the copyright law to the copyright owner and copyright-related right holders for digital works. The difference between digital copyright and traditional copyright is that its creation and editing, storage mainly rely on technologies such as computers, the Internet, big data and blockchain, and the dissemination methods are diverse and are not limited to traditional paper media [35]. Compared with traditional copyright, digital copyright has the following characteristics:

First of all, the content of digital copyright is unique and diverse. It is a new form of copyright produced by the development of computer, Internet and new technology of information dissemination, such as user-generated content, multimedia data, etc., which is very different from the traditional printed or visual art works, and brings great difficulties to the right holders' control of copyright.

Secondly, digital copyright can be used multiple times and flexibly. A fundamental requirement of the intellectual property system is that license agreements are fairly and effectively exchanged between multiple contracting parties.

Third, digital copyrights are highly technical and easier to transmit and use. Digital technology allows for accurate counting of the number of uses and value of copyrights, making it easier for parties to enter into licenses.

In addition, digital copyright is also characterized by searchability and linking, which makes it easy for users to find and use copyrights, facilitates the deposit of evidence by right holders, and allows people to exercise fuller rights and interests in digital products than in non-digital products.

2.1.2. Blockchain technology features. Blockchain technology is a distributed ledger technology that records and verifies transactions and data through a decentralized approach. It consists of one or more blocks, each containing a series of transactions and related data. These blocks are linked together in

chronological order to form a tamper-proof, transparent chain of data. Blockchain technology mainly consists of several key features including distributed and decentralized, traceability, transparency, tamperability and anonymity, and smart contracts.

Blockchain technology is distributed and decentralized, blockchain technology is built based on a distributed network, there is no centralized control body, all participants work together to maintain the ledger and share information and rights. The data and control of the blockchain are dispersed across multiple nodes of the network rather than centralized in a single centralized body, which makes the blockchain more reliable and resistant to attack [36].

Blockchain technology then uses timestamps and traceability features to ensure the traceability of copyright information. Each block contains a time cutoff that records when the block was created. Copyright enforcement and protection requires a clear point in time and evidence, and the sequence and timestamps of the blockchain allow the creation and modification history of copyright information to be pinpointed. This provides evidence for copyright owners to use to prove that they owned the copyright at a specific point in time.

A smart contract is an automated contract executed on the blockchain. Smart contracts allow rules and conditions to be set to automatically execute and validate transactions when set requirements are met, reducing the need for intermediaries and increasing the efficiency and security of transactions. The authorization and licensing of digital copyrights in libraries requires clear contracts and agreements. When the copyright owner authorizes the copyright to others, the smart contract automatically executes the authorization contract and ensures the legality and compliance of the copyright usage.

2.1.3. Digital Copyright Protection for Libraries Applying Blockchain Technology:

1) Copyright registration and proof. Blockchain technology can be used for copyright registration and proof to ensure the authenticity and immutability of copyrights by storing copyright information on the blockchain. Each copyright information is recorded in the immutable distributed ledger of the blockchain, providing reliable evidence for proving the existence and ownership of the copyright.

2) Copyright transaction and license management. Blockchain technology can be used for copyright transaction and authorization management to automate the copyright transaction and authorization process through smart contracts. The copyright owner can authorize the copyright to other people, and when the pre-set conditions of the smart contract are met, the authorization contract will be automatically executed, and the legality and compliance of the copyright use will be ensured.

3) Tracking and Punishment of Infringement. The tracking function of blockchain technology can ensure that copyright owners can accurately understand the occurrence and dissemination of infringement. By recording information about copyright usage and transactions, blockchain technology can establish a clear chain of information linking infringements to their sources and distribution paths. In this way, copyright owners will be able to trace back the specific details of the infringing behavior, providing strong evidence for the pursuit of infringement liability.

The application of blockchain technology can also provide copyright owners with an effective means of legal recourse. By recording information on copyright use and transactions, blockchain technology can ensure that copyright owners have reliable evidence for suing infringers and defending their rights and interests. The application of this technology provides a more efficient and reliable means for copyright protection, making it difficult for infringers to escape legal sanctions.

4) Decentralized copyright protection platform. Blockchain technology can support the establishment of a decentralized copyright protection platform, which can realize decentralized management and supervision of copyright protection through the distributed characteristics of blockchain. Such a

platform can provide a more fair and transparent copyright protection environment and reduce the involvement of intermediaries and potential misbehavior.

5) Blockchain-based digital watermarking technology. Blockchain technology can be combined with digital watermarking technology to achieve more secure and traceable digital copyright protection. By associating digital watermarking information with copyright information and recording it on the blockchain, piracy and tampering of digital content can be effectively prevented.

2.2. Principles of Blockchain Core Technology

2.2.1. Blockchain Fundamentals. Blockchain technology has several unique advantages, the most important of which are decentralization and autonomy. It consists of one block arranged in the order of occurrence and is a log record of the entire state change. It achieves data exchange by establishing an open and transparent algorithm based on a consensus mechanism, which ensures that the data exchange between nodes is secure and has autonomy and anonymity. The data structure of the block is shown in Figure 1.

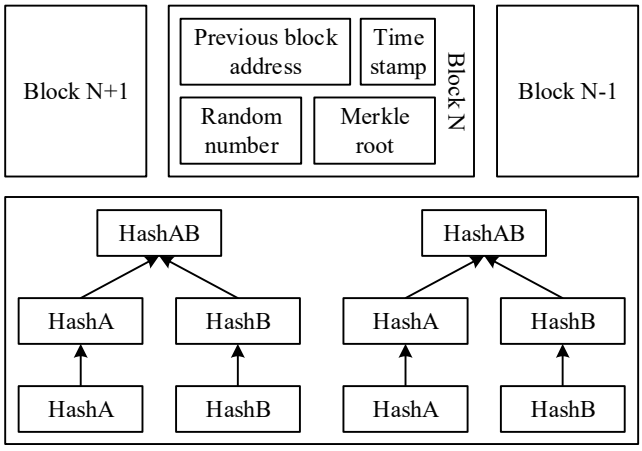


Fig. 1. Blockchain data structure diagram

2.2.2. Blockchain classification:

1) Public Blockchain. A public blockchain is a technology that can be shared by users around the world, allowing users to send transactions and validate them effectively, as well as participate in a cognitive consensus process to determine which blocks can be added to the blockchain and specify the current state.

2) Coalition Blockchain. The consensus process of coalition blockchain is not carried out by all nodes to participate, but some nodes are selected to participate in the consensus process. It is only for a specific group or a limited number of third-party organizations, and selects some nodes to carry out consensus and bookkeeping, and finally these selected nodes decide to generate each block. Other access nodes can only perform transactions and cannot participate in consensus bookkeeping.

3) Completely private blockchain. Completely private, signifies the separation of write and read permissions, thus better protecting data security and reliability. This technology can be used in a variety of areas such as database management, auditing, and corporations, but in many cases it does not require public auditability or public readability.

2.2.3. Blockchain core technology:

1) Consensus mechanism. The consensus mechanism of blockchain technology is an algorithmic mechanism that allows multiple participating nodes to obtain the same conclusion under predetermined rules using communication interactions.

Currently, Bitcoin and Ether are two of the most successful projects in the blockchain community, and they both use the POW consensus mechanism to ensure consistency among distributed nodes. Nodes compete for bookkeeping rights by computing hash puzzles. In general, the probability of a node successfully solving a puzzle is proportional to the computational resources it has.

The bookkeeping right in the POS proof algorithm is awarded to the node with the highest stakes. In other words, the arithmetic power in POW becomes the coin age of that node. Compared to POW, the POS algorithm saves arithmetic resources and improves the efficiency of consensus. However, in this case, it may lead to monopolization of bookkeeping rights by a few nodes.

The DPOS proxy proof-of-interest algorithm is used by the users of the coin-holding nodes to vote for a certain number of node representatives to maintain the blockchain network. The selected node witnesses get rewarded according to a certain percentage. If a node representative fails to generate a bookkeeping block within the corresponding time limit, the bookkeeping rights will be given to the next witness candidate to regenerate the block. This algorithm has lower arithmetic resource consumption, faster consensus, and lower network operation and maintenance costs.

Practical Byzantine Fault Tolerance (PBFT) is a state-machine replication algorithm that ensures that the system still achieves distributed consensus when at most (n-1)/3 nodes in the network fail or operate maliciously, assuming that n is the number of nodes participating in the consensus.

The consensus process of the PBFT algorithm uses a three-stage protocol with the following steps:

(1) First is the client sends a request message to the master node

(2)When the master node receives the request messages and confirms that they are correct, it saves them and generates a pre-prepared message based on these request messages and broadcasts it to the other nodes.

(3)Each consensus node receives the pre-prepared message day verifies successfully, it saves it and generates a prepared message and broadcasts it to other nodes.

(4)Each node receives the ready message and verifies it successfully, saves it and generates a commit message to broadcast to other nodes.

(5) Each node receives (2n+1)/3 commit messages and verifies that they are correct, and then performs the operation in the request operation message from the client.

2) Principles of Cryptography. Cryptographic encryption techniques are divided into two categories: symmetric encryption, and asymmetric encryption. Among them, symmetric encryption uses the same key for encryption and decryption. While asymmetric encryption uses a different key for encryption and decryption, it has two keys: public key, private key. Asymmetric encryption is also called public key cryptography, which demonstrates the application of different keys for encryption and decryption [37]. The process of asymmetric encryption and decryption for message transmission is shown in Figure 2.

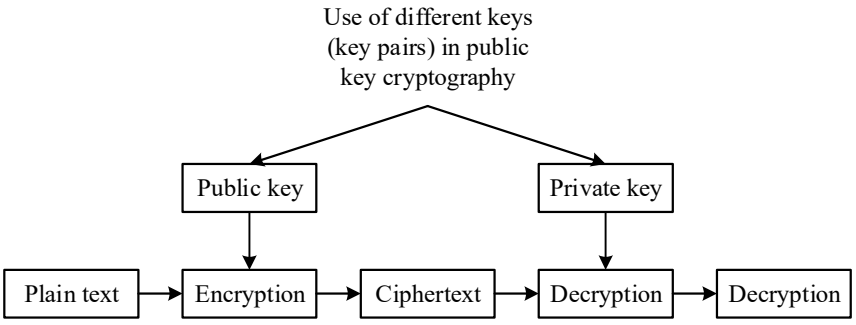


Fig. 2. Asymmetric encryption and decryption process diagram

The process is as follows:

(1) The receiver of the message prepares the public and private keys.

(2)The private key is kept by the receiver and the public key is released to the sender.

(3)The sender of the message can encrypt the message using the public key of the receiver.

(4)The receiver of the message decrypts the message using its own private key.

The algorithm execution process of digital signature is shown in Fig. 3. It is an electronic signature generated based on PKI system, which is based on the application of digital digest technology and asymmetric key encryption technology, and contains the electronic document and sender's identity information, which is able to recognize whether the sender's identity information has been tampered with or not.

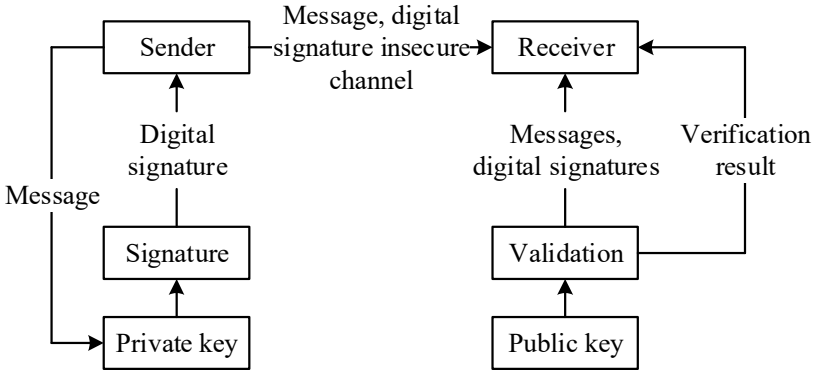


Fig. 3. Algorithm execution process of digital signature

3) Hashing Algorithm. The hash algorithm maps a data message to a shorter fixed-length value called a hash or message digest. It is a one-way encryption system that enables irreversible mapping from plaintext to ciphertext. Hashing algorithms are capable of generating a hash value corresponding to the input value quickly and within a limited time and resources. Even if the hashing algorithm and the input and output values are known, it is difficult to accurately deduce their input values within a limited time and resources [38].

4) Merkle Tree. Merkle Tree is a data structure that represents the hash values of data in the form of a tree, which is very similar to a binary tree. The values on its leaf nodes are usually hashes of data blocks. The use of Merkle Tree in blockchain enables proof of assets or proof of liabilities.

2.3. Design of blockchain-based copyright protection system for digital libraries

2.3.1. Design of copyright registration programs for digital works. Upload the IPFS system through a smart contract and return the hash value to be saved in the Fabric blockchain. The specific process is shown in Figure 4.

1) The user enters the system through the Web client, uploads the digital work and enters the copyright information of the digital work, and clicks upload to proceed to the next step.

2) The copyright protection system automatically performs feature extraction on the uploaded digital work and compares the similarity with the digital work whose copyright has been registered in the copyright protection system to ensure that the work is not infringed.

3) Digital works that pass the automatic similarity detection function of the copyright protection system are uploaded to the IPFS system by invoking a smart contract to generate a hash value.

4) After the work is successfully uploaded, the smart contract is invoked to generate a new block to store the copyright information of the work, and the work is stored in the IPFS system.

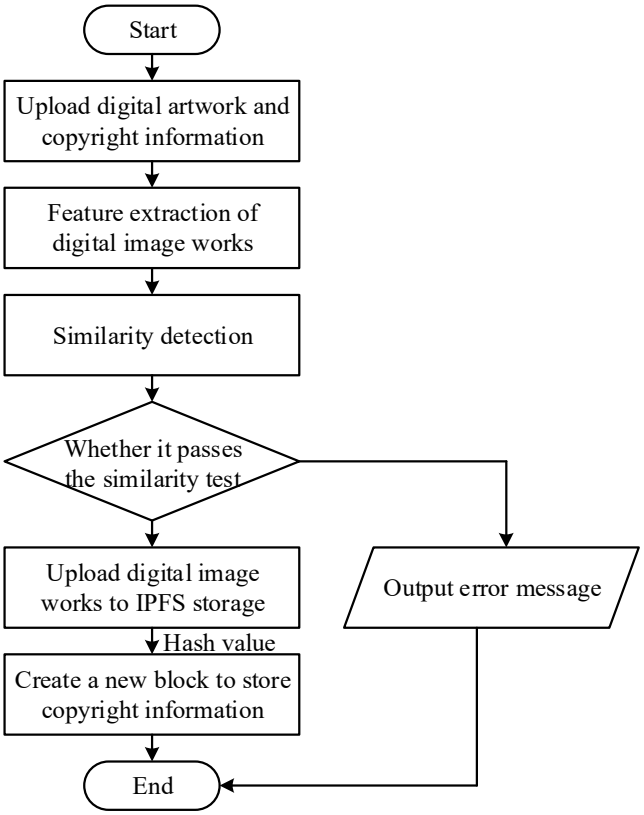


Fig. 4. IPFS storage process

2.3.2. Unique authentication of the copyright of a digital work. When users register the copyright of new digital works, the copyright protection system needs to judge and identify the digital works with high similarity in a dynamic way, so as to prevent the occurrence of the phenomenon of repeated registration of the copyright of the same work. The object of copyright protection is digital libraries, mainly for digital works, which need to fully consider the uniqueness of digital libraries when the system automatically performs feature extraction and similarity detection. The following digital cultural works of libraries are analyzed by algorithmic features such as color and similarity algorithm of multi-feature fusion, so as to identify the uniqueness of the copyright of digital works uploaded to the copyright protection system and the originality of the works.

1) CLD Algorithm. The full name of CLD is Color Layout Descriptor, and there is an extraction process that includes four stages: image segmentation, representative color selection, DCT transformation, and Zigzag scanning.

(1) In the image segmentation stage, the input image (in RGB color space) is segmented into 64 blocks.

The size of each block is M/8 x N/8, where M and N denote as the width and height of the input image;

(2) In the representative color selection process, YCbCr, which is recommended by the MPEG-7 standard, is used for the CLD color space. First, after the image segmentation is finished, the MPEG-7 standard recommends using the average pixel color value of the region block as the representative color because it is simple to select and the description accuracy is generally sufficient.After obtaining the representative color image, the image is further converted from RGB color space to YCbCr color space;

(3) In the DCT transformation stage, the luminance (Y) and the blue and red chromaticity (Cb and Cr) are transformed by an 8×8 DCT transformation, which results in three sets of DCT matrices containing 64 coefficients, and the DCT transformations are performed using the following equations:

$$B_{pq} = a_P a_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A_{mn} \cos \frac{\pi(2m+1)p}{2M} \cos \frac{\pi(2n+1)q}{2N} \quad (1)$$
$$0 \le p \le M-1, 0 \le p \le N-1$$

$$a_p = \begin{cases} \frac{1}{\sqrt{M}}, P = 0 \\ \sqrt{\frac{2}{M}}, 1 \le P \le M-1 \end{cases} \quad a_q = \begin{cases} \frac{1}{\sqrt{N}}, q = 0 \\ \sqrt{\frac{2}{N}}, 1 \le q \le N-1 \end{cases} \quad (2)$$

(4) Zigzag scanning of the three matrices containing 64 DCT coefficients obtained in the previous step is performed in the order shown in the figure, the purpose of such scanning is to group the low-frequency coefficients of the 8×8 matrix, the basic process is shown below:

$$D = \sqrt{\sum_i w_{yi} (DY_i - DY_i^{'})^2} + \sqrt{\sum_i w_{bi} (DCb_i - DCb_i^{'})^2} + \sqrt{\sum_i w_{ri} (DCr_i - DCr_i^{'})^2} \quad (3)$$

After completing the above CLD feature extraction, CLD feature matching is performed, which is accomplished by calculating the distance between two CLD features, if two CLD feature values {DY, DCb, DCr} and {DY', DCb', DCr'} are given, the distance between the features is calculated as:

Where Yi, Cbi and Cri denote the ith coefficient of Y, Cb, Cr color components, W_{1i}, W_{2i} and W_{3i} represent the weighted value of the ith set of coefficients, respectively, and the weighted value should be reduced according to the Zigzag scanning order because the low-frequency part contains more information.

2) Phash algorithm. Perceptual hash algorithm, abbreviated as pHash, is a kind of hash algorithm, which is mainly similar to the detection work of pictures. The specific steps are as follows:

(1) Reduction of size. In order to simplify the calculation of the DCT, pHash starts with a small picture (it is recommended that the picture is larger than 8×8, 32×32);

(2) Simplify color. As with aHash, the picture needs to be converted to a grayscale image to further simplify the computation.

(3) Calculate the DCT. The DCT is to decompose the image into frequency aggregates and trapezoidal shapes. The full name of the DCT transform is Discrete Cosine Transform, which is mainly used for the compression of the data or image, and is able to convert signals from the null domain to the frequency domain, with a good performance of de-correlation.

The positive transform equation of one-dimensional discrete cosine transform is:

$$F(u) = c(u) \sum_{i=0}^{N-1} f(i) \cos \left[\frac{(i+0.5)\pi}{N} u \right] \quad (4)$$

$$c(u) = \begin{cases} \sqrt{\frac{1}{N}} & u = 0 \\ \sqrt{\frac{2}{N}} & u \ne 0 \end{cases} \quad (5)$$

where, $f(i)$ is the original signal, $F(u)$ is the DCT transformed coefficients, N is the number of points of the original signal $c(u)$ can be considered as a compensation coefficient that allows the DCT transform matrix to be orthogonal.

The orthogonal transform formula for the 2D discrete cosine transform is:

$$F(u,v) = c(u)c(v) \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i,j) \cos \left[\frac{(i+0.5r)\pi}{N} ug \right] \cos \left[\frac{(j+0.5r)\pi}{N} v \right] \quad (6)$$

$$c(u)=\begin{cases}\sqrt{\frac{1}{N}} & u=0\\ \sqrt{\frac{2}{N}} & u\neq 0\end{cases} \tag{7}$$

(4) Shrinking the DCT. The DCT results in a matrix of size 32 × 32, but only the 8 × 8 matrix in the upper left corner, which presents the lowest frequency in the picture, should be retained;

(5) Calculate the mean value. Calculate the mean value of the DCT;

(6) Further reduce the DCT. according to the 8 × 8 DCT matrix for comparison, greater than or equal to the average value of the DCT is set to "1", less than the average value of the DCT is set to "0". If the overall structure of the image remains unchanged, the hash result value remains unchanged;

(7) Construct the hash value. Combination of 64 bits to generate the hash value, the order is arbitrary, but before and after to maintain consistency.

(8) Compare fingerprints; compute the fingerprints of the two images and calculate the Hamming distance.

3) SIFT algorithm. SIFT, Scale Invariant Feature Transformation, is an image local feature extraction algorithm, which extracts features by searching for the precise location and main direction of the extreme points (feature points, keypoints) in different scale spaces, and constructing keypoint descriptors. The specific steps are as follows:

(1) Scale-space extreme value detection. The DoG scale space is constructed and the "scale" can be understood as the degree of image blurring. The image is convolved using Gaussian kernels with different sigma values to scale the image.

$$L(x,y,\sigma)=G(x,y,\sigma)*I(x,y) \tag{8}$$

$$G(x,y,\sigma)=\frac{1}{(2\pi\sigma^2)}e^{\left((x^2+y^2)\big/(2\sigma^2)\right)} \tag{9}$$

where $G(x,y,\sigma)$ is the Gaussian kernel function, $L(x,y,\sigma)$ is the scaled image under the corresponding σ-valued scale processing, and $I(x,y)$ is the original image.

(2) Key point search and localization. The extreme points found in the discrete space are not necessarily the real extreme points, we use hash value interpolation to transform the hash space into a continuous space to get more accurate extreme points. The low-contrast keypoints and unstable edge response points are also removed;

(3) Key point direction assignment. For the direction gradient of the key point, the gradient magnitude and direction are calculated by the formula:

$$G(x)=L(x+1,y)-L(x-1,y) \tag{10}$$

$$G(y)=L(x,y+1)-L(x,y-1) \tag{11}$$

$$G(x,y)=\sqrt{G(x)^2+G(y)^2} \tag{12}$$

$$\theta=\tan^{-1}\frac{G_y}{G_x} \tag{13}$$

(4) Feature point description. Take 8*8 pixels around the feature point for gradient direction statistics and Gaussian weighting. Every 4*4 window generates 8 directions so that 2*2*8 vectors are generated as the mathematical description of the feature point.The flow of SIFT algorithm is shown in Fig. 5.

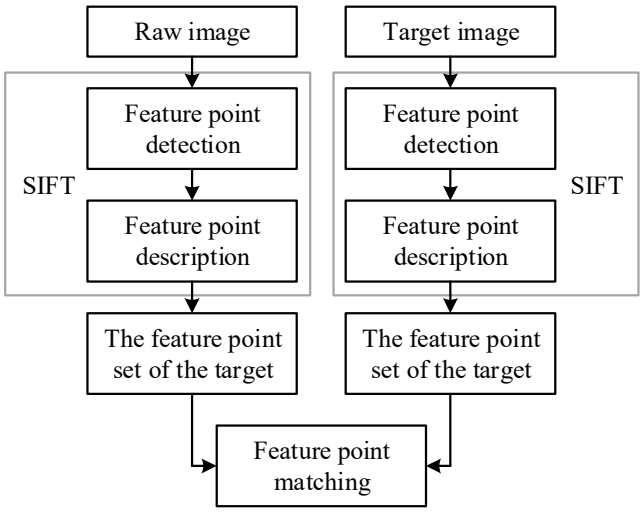


Fig. 5. SIFT algorithm flowchart

2.3.3. IPFS distributed storage. IPFS (Interplanetary File System), a peer-to-peer hypermedia distribution protocol for content addressability, can also be understood as a website that supports distributed storage.

The IPFS protocol makes innovations mainly in two aspects: data storage and file transfer. When saving a file in the IPFS system, the system breaks the file into a number of fragments of the same size. Then each fragment is hashed to get a hash value, and then the hash value of all the fragments and related data are integrated and hashed to get a final hash value. This hash value is transferred to the IPFS system. The fragments of the uploaded file may be stored in any computer with the same file system. This computer neither knows what the contents of these fragments are nor for whom the file is stored. As long as there is no hash value corresponding to the file, no individual or organization can view the contents of the stored file, and there is no need to worry about the uploaded file or data being utilized by someone without permission, which ensures data security.

2.3.4. Privacy data encryption. This paper encrypts the private user data involved.

The key generation process of SM2 algorithm: Assuming that the setup parameters are a, b, p, an elliptic curve EP(a, b) is selected; a base point G(x, y) is selected on the elliptic curve EP(a, b); the base point G(x, y) is multiplied by the private key d, and the public key Q=dQ is obtained through computation to get the key pair (k1, k2);

AES key encryption process: private data is encrypted by AES algorithm, the key of AES algorithm is encrypted using SM2 encryption algorithm to generate the ciphertext of the key, the sender chooses a random number r(r<n), embeds the AES key in a point M(m1, m2) on the elliptic curve Ep(a, b), calculates Cl=rG, C2=M+rk2, and transmits data to the receiver information C1, C2;

AES key decryption process: when the receiver receives the data message C, Cz, according to the formula calculated M, M = C2-K1Cl = M + rK-KrG = M + rK, G-K1rG = M; decoding M can get the key of the AES algorithm, the AES algorithm will be decrypted to get the corresponding encrypted information of the original text.

3. Algorithm performance comparison and application analysis

3.1. Algorithm Performance Analysis

3.1.1. Push-up testing. In order to verify the performance of the blockchain copyright protection algorithm designed in this paper, the simulated experiment is designed for completing the election

process of the group nudging algorithm, and letting the nudging nodes use the batch consensus to complete the batch bookkeeping node election consensus and block data production consensus.

In the experiment, the chain network is divided into four organizations, each organization has four nodes, and the total number of nodes is 16. The performance index of all nodes needs to be initialized at the beginning of the test, and then the values of each node are evaluated by the end of the first round of push bookkeeping, and the list of performance indexes of the nodes in the group is recorded. In order to simulate the environment, different groups of nodes with serial numbers 3, 11, and 16 were made to have a high probability of various low performance evaluation phenomena such as network timeout, block out speed delay, and so on through intentional control of the network environment, which will ultimately lead to the reduction of their performance index evaluation within the organization. The number of preferred bookkeeping records the number of times the pushed node is elected as the first bookkeeping node in each round of bookkeeping process, and the statistics of node pushed and preferred bookkeeping times in the simulation environment are shown in Table 1.

The experimental process can be seen that the algorithm in this paper can ensure that the chain network nodes participating in the push-up bookkeeping process can obtain the reliability of fair participation in the system based on the performance of the voting also allows nodes with consistently high performance to obtain a higher number of bookkeeping participation. The higher performance group 1 node 1, group 2 node 1, group 3 node 2, and group 4 node 2 are all useful for the highest number of bookkeeping participation, totaling 49% of the total number of bookkeeping participations, and the corresponding lower performance group 1 node 3, group 3 node 3, and group 4 node 4 receive fewer bookkeeping participations, but with a certain degree of assurance, receiving 7% of the total number of bookkeeping participations.

Table 1. Each node participates in the promotion and the preferred number of records

Serial number	Grouping	Numbering	Performance index	Candidates/times	Lift times/times	Preferred record number/times
1	1	1	22	500	317	77
2		2	13	500	134	38
3		3	4	500	13	4
4		4	6	500	45	12
5	2	1	19	500	216	54
6		2	6	500	97	25
7		3	3	500	85	22
8		4	7	500	119	27
9	3	1	9	500	106	28
10		2	25	500	258	68
11		3	5	500	56	15
12		4	9	500	97	24
13	4	1	10	500	176	46
14		2	12	500	88	49
15		3	5	500	85	21
16		4	3	500	59	16

Figure 6 shows the comparison of the number of times nodes in each group get pushed. High-performance metrics nodes have a higher likelihood of being elected as a nudge node in the group. The number of times a node is elected as an envisioned node directly affects the number of times a node receives a bookkeeping node.

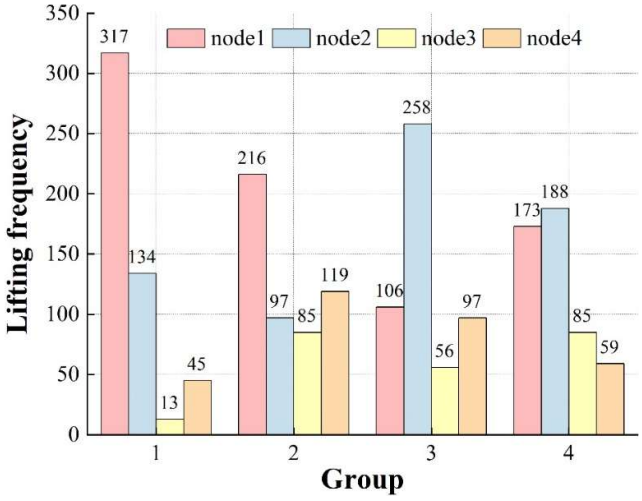


Fig. 6. The nodes were compared with the number of promotions

Fig. 7 shows the number of times each node gets preferred bookkeeping. In Fig. 7 it can be found that group 1 node 1, group 2 node 1, group 3 node 2 and group 4 node 2 have 58.78%, 42.19%, 50.37% and 37.12% percentage of preferred number of bookkeeping times in each group respectively.

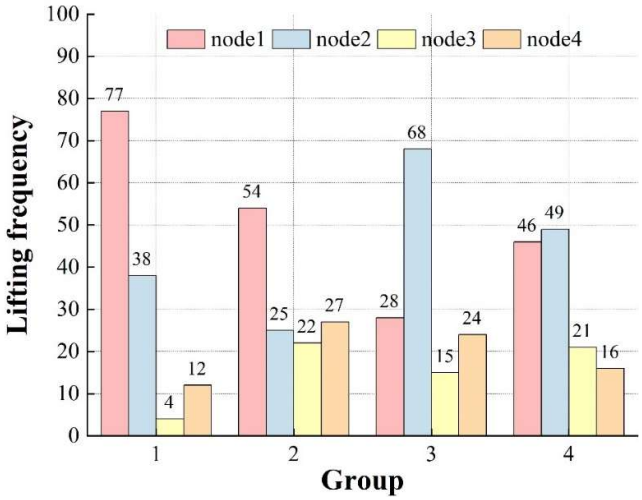


Fig. 7. Each node gets the preferred number of entries

3.1.2. Performance testing. The performance test is set up to include the PBFT consensus algorithm, RBFT consensus algorithm simulation experimental data flow, and compared with the algorithm designed in this paper in terms of the network performance indicators, the performance indicators are divided into two categories, one for the space resources in the network to occupy the bandwidth, and the other for the data throughput of the network.

Define B as the required network bandwidth using the unit of Mbps, define the total number of nodes in the chain network as n, using the unit of one, define the size of block data as Size, using the unit of Mb/s, because the size of the block data used in the practice in the blockchain network is 1.1MB, so in the simulation experiments also set the Size to 1.1MB. The experiment compares the bandwidth usage of the three consensus algorithms under different number of nodes, as shown in the results in Figure 8.

It can be seen that the bandwidth usage of this paper's algorithm under different number of nodes is lower than the baseline model, and the bandwidth usage when the number of nodes is 20 is more than 200 Mbps less than that of RBFT, which occupies less space resources in the blockchain network and outperforms the PBFT and RBFT algorithms.

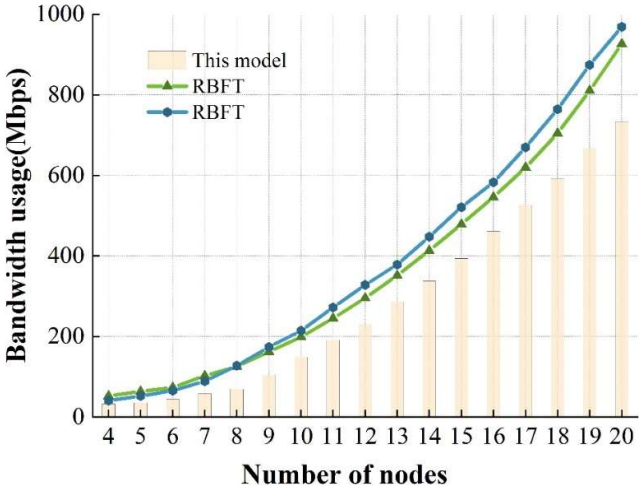


Fig. 8. Bandwidth occupancy contrast diagram

The throughput of an alliance chain network system is obtained by counting the total number of transactions that can be processed by the network in a unit of time. After obtaining the throughput of

the network system, it is only necessary to compare the size of the throughput values of different consensus mechanisms under the same network topology, so as to understand the difference in the ability of different consensus mechanisms in concurrent task processing. According to the flow of transaction data processing in the chain system, TPS is used to represent the number of transactions that can be successfully booked on the chain by the bookkeeping nodes to complete consensus per second as the throughput of the system.

The comparison of the algorithm transaction throughput obtained according to the simulation experiment is shown in Figure 9. The throughput performance of this paper's algorithm is 36.19% more than the PBFT consensus algorithm on average when the number of nodes is from 4 to 20, and 55.92% more than the RBFT consensus algorithm on average. Therefore it can be concluded that the throughput performance of this paper's algorithm is better than PBFT algorithm and RBFT algorithm.

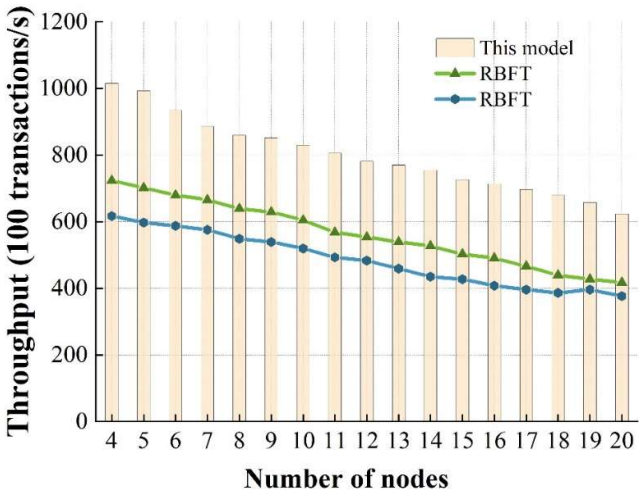


Fig. 9. Throughput contrast diagram

3.1.3. Data security analysis. In real application scenarios, not all members of the network are trustworthy, i.e., there are malicious/semi-malicious vehicles (malicious/semi-malicious workers, malicious/semi-malicious miners) as well as compromised RSUs. The performance of the proposed method is compared with vanilla FL and FL chain by setting different proportions of malicious users and testing on MNIST dataset. Figure 10 shows the variation of model accuracy under different methods. The model accuracy of vanilla FL is very low when there are a certain number of malicious members in the network, and the model accuracy is not improved during the training process afterwards. At the very beginning, its model accuracy fluctuates at 20.38%, while after the 50th round of training, the overall accuracy of the model falls and stays around 11.42%. Comparing the FL chain and the method in this paper, both of them have some resistance to malicious members due to the introduction of blockchain technology. Specifically, FL chain achieves 79.36% accuracy after enough training, while DBL has a final accuracy of 86.19% and reaches a maximum accuracy of 89.23% in round 90. Compared to the baseline method vanilla FL, the method in this paper can to some extent resist the influence of wrong parameter uploads on the global model caused by various reasons, such as itself as well as the outside world, to ensure the robustness of the system, and is able to prevent low-quality nodes from influencing the learning results.

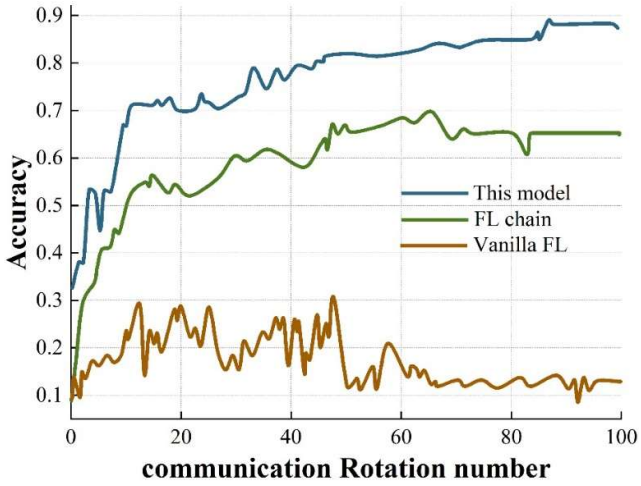
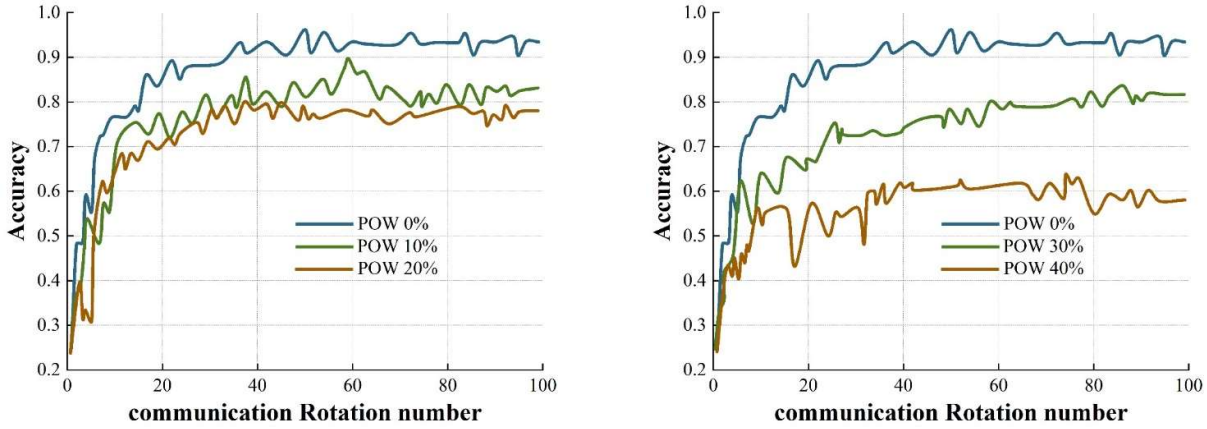


Fig. 10. Changes in the precision of the model under different methods

Assuming that the proportion of malicious users is {0%, 10%, 20%, 30%, 40%}, the malicious users will launch poisoning attacks to interfere with the training of the model. The variation of the model's accuracy with the communication rounds is shown in Fig. 11, where (a)(b) shows the accuracy statistics at 10% and 20% and 30% and 40% of the malicious user proportion, respectively.

It can be seen from the figure that the average accuracy of the final model is maintained around 87.25%. This is because the blockchain network consisting of RSUs acting as validation nodes can determine how good or bad the model updates uploaded by each vehicle are based on local data, and the miner vehicles can filter out the malicious models before aggregating them. The experimental results show that the proposed DBL method is resistant to poisoning attacks.



(a) The proportion of malicious users is 0%, 10%, 20%

(b) The proportion of malicious users is 0%, 30%, 40%

Fig. 11. The proportion of malicious users is 0%,10%,20%

Specifically, as the number of malicious users increases, the ability of the model to resist malicious user attacks becomes worse and worse, and the final accuracy of the model becomes lower and lower, and the specific accuracy changes are shown in Table 2. With different proportions of the number of malicious devices, the method in this paper can achieve about 88.43% model accuracy.

Table 2. Accuracy comparison under different proportions of malicious nodes

The proportion of malicious user equipment	Accuracy of the first 100 rounds	Maximum accuracy	The communication rotation of the maximum precision
0.00%	95.37%	95.96%	94
10.00%	90.62%	93.71%	88
20.00%	87.38%	88.47%	86
30.00%	81.68%	84.27%	87
40.00%	79.45%	79.75%	89

3.2. Application Analysis of Copyright Protection in Digital Libraries

Analyze the security, stability and scalability of this paper's solution in a real-world application of copyright protection in libraries.The Jmeter testing tool can be well implemented to test the performance of the system, including presenting the results of the performance metrics as above. This section tests the performance by varying the system load as well as the number of concurrencies.Jmeter version is 5.4.3.

3.2.1. Library WEB service performance testing. In the performance test under the WEB module, according to the actual situation users in the system mainly to browse the digital resources list of the home page of the digital library, so to get the list of works as an example to carry out performance testing. Jmeter set the number of threads to 4000, a total of 10 tests, set the HTTP request path to get the list of works interface path and add the aggregation report. The test results show that in every 4000 request data, the error rate of each request is 0, the average response time is 29ms, and the system throughput is 490/s, which meets the user requirements.

3.2.2. Copyright similarity search performance test. In the performance test of copyright similarity retrieval, the feature vector similarity retrieval function of digital resources is selected, and the computation time of vector similarity search for a single video, the test results are shown in Fig. 12. In the line graph between the number of feature vectors and the similarity retrieval time, the horizontal coordinate is the number of video copyright feature vectors stored in the system, and the vertical coordinate is the computation time to be carried out for the similarity search of a single video.

The efficiency of copyright similarity retrieval is extremely important in the case of a huge amount of copyrights stored in the work copyright protection system, and as the number of work feature vectors stored in the system grows, the longer the similarity retrieval of a new work feature vector takes. From the figure, it can be seen that when there are 5000 digital resource feature vectors in the system database, the time required for similarity retrieval is only 0.523s, which meets the requirements of the system's non-functional needs for similarity retrieval runtime.

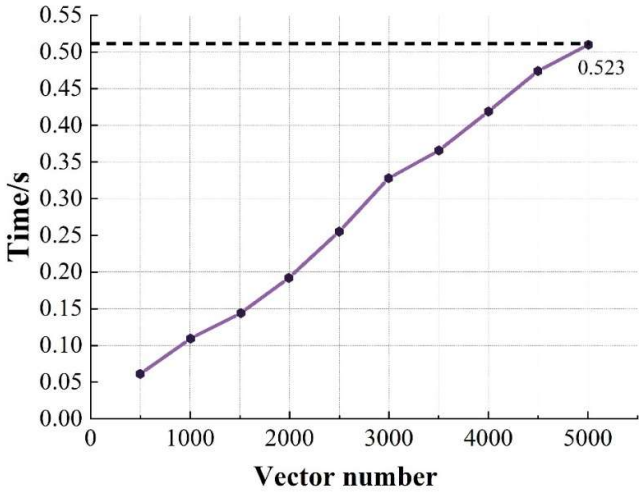


Fig. 12. Vector number and similarity retrieval time line diagram

3.2.3. Average user response time. In this subsection, the average response time metric will be tested to show the relationship between the number of users on the system at the same time and the response time of the system by viewing the results of the metric test under different load conditions. Figure 13 shows the user-response time line graph. The horizontal coordinate of the graph is the number of online users of the system, and the vertical coordinate is the average response. It can be seen that as the number of simultaneous users of the system increases from 10 to 100, the average response time of the system is longer, and when the number of simultaneous users of the system is 100, the average response time of the system is 76.4ms, which meets the daily performance requirements.

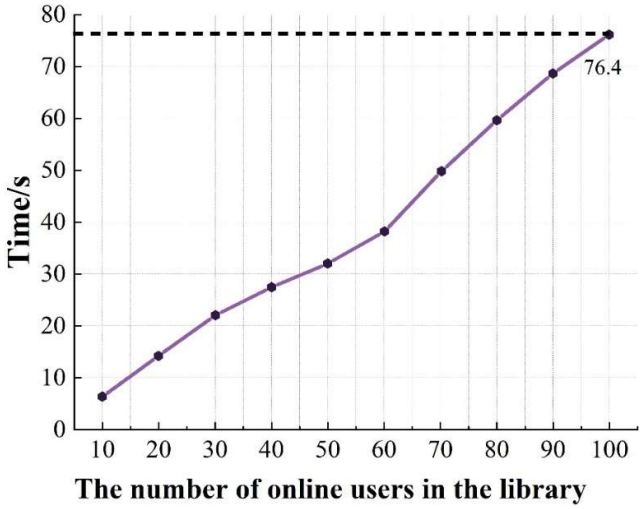


Fig. 13. User-response time line diagram

4. Conclusion

In this paper, a digital library copyright protection system is designed based on blockchain technology and data security, which realizes the functions of copyright confirmation, infringement monitoring, digital authentication and privacy protection. The performance of the algorithms is compared and the service performance of this paper's scheme is analyzed in practical applications. It is found that the bandwidth usage of this paper's algorithm is lower than that of the baseline model under different node numbers, and the bandwidth usage is more than 200Mbps less than that of RBFT when the node number is 20, and the space resources occupied in the blockchain network are lower, and the

performance is better than that of PBFT and RBFT algorithms. In every 4000 request data, the error rate of each request is 0, the average response time is 29ms, and the system throughput is 490/s, which meets the user's usage requirements. Comprehensive analysis of the results of the dimensions, it can be considered that the scheme designed in this paper can solve the problems of copyright infringement and lack of data security in the digital copyright industry.

References

[1] Vij, R. (2019). Copyright and libraries: Balancing the interests of stakeholders in the digital age. Asian Journal of Multidimensional Research (AJMR), 8(5), 208-216.

[2] Roncevic, M. (2020). Digital Rights Management: How libraries deal with publisher restrictions. American libraries, 51(3-4), 47-48.

[3] Iroaganachi, M. A. (2018). Trends and issues in digital libraries. In Library Science and Administration: Concepts, Methodologies, Tools, and Applications (pp. 1648-1673). IGI Global.

[4] Chhabra, S., & Natu, K. (2024). Blockchain Empowered: A Comprehensive Study on the Role of Libraries in Safeguarding and Managing Electronic Theses. Recent Advancements in Science and Technology, 1.

[5] Sabeena, S., Sathviga, C., & Jaishankar, R. (2025). Secure and Transparent Transactions in Libraries Using Blockchain. In Enhancing Security and Regulations in Libraries With Blockchain Technology (pp. 255-278). IGI Global.

[6] Mwanzu, A. (2021). Perceptions of librarians on the usefulness of DRM Technology in Protecting against copyright violation. Library Philosophy and Practice (e-journal), 5517.

[7] Abu Sirhan, A., Abdrabbo, K. M., Ahmed Ali Al Tawalbeh, S., Hamdi Ahmed, M., & Ali Helalat, M. (2019). Digital rights management (DRM) in libraries of public universities in Jordan. Library management, 40(8/9), 496-502.

[8] Panda, S. (2021). Digital Rights Management (DRM) in the Libraries of Digital-era: Concepts, IPR Issues & Concerns of LIS Community. Library Philosophy and Practice (e-journal), 6645, 1-20.

[9] Singh, M., & Sharma, A. (2022). Digital Rights Management and Their Implications in the Library and Information Centres. The Changing Landscape for the Libraries and Librarians: Issues & Challenges, 163.

[10] Meesad, P., & Mingkhwan, A. (2024). Emerging Technologies in Smart Digital Libraries. Libraries in Transformation: Navigating to AI-Powered Libraries, 211-270.

[11] Khan, S. A., & Shahzad, K. (2024). Key features of digital library management system (DLMS) for developing digital libraries: An investigation from LIS practitioners in Pakistan. Journal of Librarianship and Information Science, 56(1), 29-42.

[12] Eiriemiokhale, K. A. (2018). Copyright issues in a digital library environment. In Handbook of Research on Managing Intellectual Property in Digital Libraries (pp. 142-164). IGI Global.

[13] Safdar, M., Qutab, S., Ullah, F. S., Siddique, N., & Khan, M. A. (2023). A mapping review of literature on Blockchain usage by libraries: Challenges and opportunities. Journal of Librarianship and Information Science, 55(3), 848-858.

[14] Wang, Y. C., Chen, C. L., & Deng, Y. Y. (2021). Authorization mechanism based on blockchain technology for protecting museum-digital property rights. Applied Sciences, 11(3), 1085.

[15] Ashtagi, R., Rajput, V., Mohite, S., Musale, V., Jaybhaye, S. M., & Bidwe, R. V. (2024). Utilizing Blockchain for Enhanced Security and Transparency in Library Transactions. Library of Progress-Library Science, Information Technology & Computer, 44(1).

[16] Negi, D. S., Sharma, P., & Sharma, A. (2024). Private Blockchain, Hybrid Blockchain, Public Blockchain Technologies in Digital Library Service. ACADEMIC LIBRARIES, 424.

[17] Tamilselvan, N. (2024). Blockchain-Based Digital Rights Management for Enhanced Content Security in Digital Libraries. Technology (IJBT), 2(1), 1-8.

[18] Jha, S. K. (2023). Application of blockchain technology in libraries and information centers services. Library Hi Tech News.

[19] Emmanuel, V. O., Efemini, M., Yahaya, D. O., & Oladokun, B. D. (2023). Application of blockchain technology to 21st century library services: Benefits and best practices. Data and Metadata, 2, 59-59.

[20] Nachiappan, B., Viji, C., Mohanraj, A., Moorthi, I., Mir, M. H., & Rajkumar, N. (2025). Enhancing Data Security and Accessibility in Libraries Through Blockchain Technology. In Enhancing Security and Regulations in Libraries With Blockchain Technology (pp. 87-116). IGI Global.

[21] Omame, I. M., & Alex-Nmecha, J. C. (2021). Application of blockchain in libraries and information centers. In Handbook of Research on Knowledge and Organization Systems in Library and Information Science (pp. 384-397). IGI Global.

[22] Ahmad, M. I. (2024). Use of Blockchain Technology in Library. Synergy: International Journal of Multidisciplinary Studies, 1(2), 9-14.

[23] Elihaki Kanyika, M., Sadykova, R., Tuenbaeva, K., & Chuwa, L. (2025). Adapting to new technologies: A systematic literature review of blockchain implementation in academic libraries. Information Development, 02666669241312153.

[24] Fasola, O. S., Oyadeyi, A. E., & Iyoro, A. O. (2024). Awareness, Acceptance and Readiness to Use Blockchain Technology for Library Services in Academic Libraries in Nigeria. Communicate: Journal of Library and Information Science, 26(1), 270-288.

[25] Abid, H. (2021). Uses of blockchain technologies in library services. Library Hi Tech News, 38(8), 9-11.

[26] Hasan, N. (2020). Blockchain technology and its application in libraries. Library Herald, 58(4), 118-125.

[27] Kishore, H. S., Steffi, R., & Senthilkumar, K. R. (2025). Blockchain-Based Digital Rights Management and Its Impact on English Language Resources in Libraries. In Enhancing Security and Regulations in Libraries With Blockchain Technology (pp. 337-350). IGI Global.

[28] Karthick, N., Nithya, P., & Rajkumar, R. (2025). Blockchain Technology in Libraries: Paving the Way for Transparent and Reliable Services. In Leveraging Blockchain for Future-Ready Libraries (pp. 77-92). IGI Global Scientific Publishing.

[29] Hota, P. K., Hota, L., & Dash, P. K. (2022). Blockchain in Digital Libraries: State of the Art, Trends, and Challenges. Machine Learning Adoption in Blockchain-Based Intelligent Manufacturing, 17-32.

[30] Qian, C. (2021, April). Digital copyright management model of university library based on blockchain technology. In 2021 International Conference on Computer, Blockchain and Financial Development (CBFD) (pp. 508-513). IEEE.

[31] Oyelude, A. A. (2022). Trending issues in advancing blockchain technology in libraries, archives and museums. Library hi tech news, 39(6), 6-7.

[32] Xiao, Y., Xiao, R., Zhu, C., & Wei, S. (2024). Digital Copyright Management Strategy of Digital Library Based on Blockchain Technology. Journal of Combinatorial Mathematics and Combinatorial Computing, 119, 153-162.

[33] Bashir, F., & Warraich, N. F. (2022). Prospects of semantic web and blockchain technologies in libraries. In Blockchain and Deep Learning: Future Trends and Enabling Technologies (pp. 31-45). Cham: Springer International Publishing.

[34] Longshak, J. E., Oyeboade, S. A., Abdullahi, M. S., & Chanai, K. M. (2023). Intellectual Property Rights (IPR) in the Blockchain Era. In Global Perspectives on Sustainable Library Practices (pp. 263-296). IGI Global.

[35] Shani Friedman.(2025).The Law of the Sea Goes Digital—Indigenous Peoples' "Right to Exclude" Their Traditional Knowledge from the Digital Sphere.Ocean Development & International Law(1),88-99.

[36] Lobarkhon Ruzmurodova Mirzabek Qizi & Bakhshillo Khodjaev Kamalovich.(2024).Blockchain Technology Usage on Intellectual Property Rights.International Journal for the Semiotics of Law - Revue internationale de Sémiotique juridique(2),1-18.

[37] Ankita Sharma,Nayancy & Rajat Verma.(2025).The Confluence of Cryptography, Blockchain and Artificial Intelligence.

[38] Yang Donghui,Wang Yan,Shi Zhaoyang & Wang Huimin.(2025).Toward topic diversity in recommender systems: integrating topic modeling with a hashing algorithm.Aslib Journal of Information Management(1),47-69.