Research and Modeling Electronic Data Tracing Scheme Based on Blockchain

Zhengtao Jiang¹², Yi Chen³, Zheng Liu¹⁻, Hao Chen³, Jianhong Zhang⁴ and Yanchao Zhu⁵

¹State Key Laboratory of Communication University of China, Beijing 100024, China
²Guangxi Key Laboratory of Cryptography and Information Security, Guilin, China
³Ele-Cloud Information Technology Co. Ltd., Beijing, China
⁴School of Information Sciences and Technology, North China University of Technology, Beijing, China
⁵Corresponding author email: liuzhengkeep@foxmail.com

Abstract. From the point of view that the traditional electronic data (documents in image, PDF, etc) is difficult to track and trace, taking image copyright as an example, this paper studies a new scheme to solve the problem of replication detection and traceability based on blockchain. With the decentralization and trust removing of blockchain, both copyright storage and evidence of infringement are stored in a tamper proof and transparent way. Based on the judgment of image similarity, a "resource reverse tracing strategy" is designed to control the infringement timely. The infringement evidence can reach the creator quickly, which provides convenience for the creator's rights protection. This method changes traditional tracing methods of image copyright. Also, it can be applied to the tracing problem of reuse of electronic invoice, short video, music, etc, and brings new ideas for the related problem.

Keywords: Blockchain, image copyright, P2P, cloud storage, image recognition, electronic invoice.

1. Introduction
In the era of “we media”, all kinds of self-claimed original works are full of every corner of the Internet, but the original source of works is difficult to guarantee. As a centralized management mechanism, the traditional digital copyright protection technology is authorized by an authoritative management organization.
Since different copyright management platforms do not support interoperability, it’s difficult to find the original copyright owner, which may harm the author’s interests. For massive image infringement, it is impossible to realize accurate and rapid tracing of image copyright by manual means[1]. For providing better digital rights protection, the decentralization and traceability of blockchain are concerned by researchers[2]. It realizes digital copyright registration, transaction, and tracing through its special mechanism which has no need of third party. Compared with the traditional way of image copyright protection, the image copyright tracking and tracing system based on blockchain improves the efficiency of image copyright protection, greatly saves the cost of image copyright protection, and reduces the risk of being infringed.

2. Related Work
Some researchers have studied copyright protection based on blockchain. In 2015, Kubesch et al analyzed the method to realize digital copyright protection by applying user access and the related code
protocols. They also try to explore solutions to balance the interests and relationships among users. Wang et al designed a mechanism based on blockchain and the technique of watermark to reduce the need of a trusted third party[3]. In their paper, they proposed a blockchain-based image copyright protection framework and solved the blockchain data expansion problem by interplanetary file system. In addition, the image owner can authenticate the image and trace the copyright.

Zihao Lu et al introduced a blockchain-based digital rights management scheme to balance the problems of content protection, copyright protection and trading of works. According to the scheme, the off-chain design work and its copyright record are binded in the blockchain[4]. By using smart contracts and public key cryptography, their scheme has no need for active operation from participants in the trade.

D. Roman and K. Vu use smart contracts to solve the problem of single point failure for trusted third parties in data markets[5]. In their method, they proposed an architecture and then analysed the feasibility of the architecture for data markets which based on smart contracts and multi-party computation.

In [6], R. A. Dobre et al designed a system based on blockchain which aimed at detecting and avoiding copyright infringement. Photographers can register their photos with it. Photographers can use it to register their photos and the system can check the copyright condition of the image to be used. Their system also has the function to buy the right of photo for using. It is likely that recompressed versions of the same photo is stored in different places, since businesses can buy images on many sites. We design an image copyright tracking and tracing system based blockchain technique. It mainly consists of two parts: the public chain composed of copyright owners and the alliance chain composed of “we media” platform. The copyright information of copyright owners and the image usage of we media can be traced through the system.

3. System Overall Design

The system uses virtual Ubuntu, and installs the open source application container engine Docker. It deploys the public chain hyperledger fabric1.2, builds a super ledger network (fabric), and compiles the smart contract of the blockchain using solidity. The copyright tracking and tracing system adopts B/S system architecture and utilizes the following techniques: public chain technology, private chain technology, MySQL database, massive image storage technology based on HBase, zookeeper, Kafka distributed cluster, image similarity recognition technology. It records user copyright data and infringement data based on blockchain technology.

Fabric is a distributed ledger based on blockchain. Each ledger has its own block, and the transaction data will be stored in the ledger block. The first block, genesis block, needs to be generated manually. Then the blockchain operation can be realized. Smart contract runs on a replicable and shared ledger program. It can process, receive, store and send information.

Massive images are stored by HBase's massive image storage technology. Image similarity is compared for infringement judgment. It realizes dynamic tracking of image copyright by returning copyright infringement warning to the copyright owner. The design framework of the system is illustrated in Fig. 1.

![Figure 1. The framework of image copyright tracking and tracing system.](image-url)
The recommendation strategy of the system in Fig. 1 is based on machine learning and token voting mechanism. Fig. 2 illustrates function composition of the image copyright tracking and tracing system based on blockchain. It is divided into two platforms: copyright user platform and media platform. The function of the copyright user platform consist of real name authentication, copyright registration, copyright warning, infringement complaints, infringement information revocation, copyright inquiry, et al. The function of the media platform is consist of image usage record, copyright information query, infringement warning, infringement information management, infringement complaint and so on. Functions of the image copyright tracking and tracing system are constructed from blockchain.

Figure 2. Functions of the image copyright tracking and tracing system based on blockchain.

The technological process of the blockchain based image copyright tracking and tracing system is shown in Fig. 3.

1) Images are uploaded to media platforms.
2) The uploaded date is deduplicated then is integrated into consortium blockchain. Also, it is encrypted and stored in the cloud.

On the other hand, 1') The copyright owner uploads copyright information. 2') This data is integrated into public blockchain. Also, it is stored in the cloud.

The system automatically compares the above two groups of data, and it identifies and selects the infringement data. Such data is stored in the infringement blockchain platform for evidence collection. This system helps copyright owners to track copyright in real time.

All copyright information is stored in the public chain and it is open and transparent.

When there is a need for image use, users can directly view the copyright notice of the images they use. This helps to prevent infringement.

Figure 3. Process of the image copyright tracking and tracing system based on blockchain.
change the data of the whole blockchain, which ensures the security of the blockchain. The picture copyright information tracking and traceability system model is shown in Fig. 4.

![Image copyright information tracking and traceability system model.](image)

**Figure 4.** Image copyright information tracking and traceability system model.

### 4. System Implementation

The development is based on zookeeper, Kafka distributed cluster and PHP + MySQL database. MVC design pattern is used to realize the separation of view, control and model code, simplify the workload, and accelerate the development speed.

#### 4.1 Database Design of the System

The main entities of the system include: copyright user information table, "we media" platform information table, node information table, operation log table, image copyright information table, image usage information table, Merkletree leaf node information table, infringement information table. The main steps of image copyright registration are as follows:

1) Copyright users register first, and after completing the information registration, users can log in.
2) Users with real name authentication can upload their own image copyright information.
3) The system checks the duplicate of the copyright information.
4) Calculate hash value of copyright information.
5) The copyright owner signs the above hash value and broadcast to the network. Finally it returns the upload result to the copyright owner.

Brief process is illustrated in Fig. 5.

![Copyright registration.](image)

**Figure 5.** Copyright registration.

#### 4.2 Image Similarity Comparison

Image similarity calculation is usually used to evaluate the similarity of content between two images. The common methods include pixel point contrast, center of gravity contrast, projection contrast, block comparison.
It cuts the image into several blocks, and then matches each block to calculate the similarity. And then the vector distance is calculated to get the similarity. The method of block comparison can improve the accuracy of the results significantly.

Euclidean distance algorithm is the most common algorithm for vector distance calculation:

\[ d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + \cdots + (y_{n_1} - y_{n_2})^2} \]

The above formula is used to match the image of the existing copyright database to get multiple similarity metrics. We multiply each similarity by a weight and add them to get the total similarity. The highest total similarity is the recognition result.

4.3 Image Copyright Warning

The data in the public chain is composed of image copyright, and the data in the alliance chain is composed of image usage information. Infringement can be detected by comparison of these two types of data. Unauthorized image information is fed back to the corresponding copyright owners and “we media” platform.

The system mainly links the picture copyright information, picture usage and other information to generate blocks through dynamic storage. In order to ensure the accuracy of the information on the chain, the module verification is added, and the Merkle tree structure is used to store the dynamic data. The block storage technology framework system is shown in Fig. 6.

![Figure 6. The block storage technology framework system.](image)

When there is a query task in the image use database, PHP controllers (control layer) will determine the image similarity. If the calculated similarity less than 60%, it is determined that there is no data. If more than 60%, then it judges whether there is authorization information: if there is authorization information, it will be judged that there is no infringement; if there is no authorization information, a hash value of the infringement information will be calculated. Then a signature will be generated based on the user's private key and time stamp.

The signature will be added to the infringement blockchain, and is broadcasted to the whole network, which is the infringement data. Finally, the copyright owners and “we media” platform are informed to give copyright warning and realize timely tracking of the image infringement. The main steps are illustrated in Fig. 7.
5. Conclusion

Traditional image infringement is difficult to trace. In this paper, the following solutions are proposed:

1) Image copyright protection is integrated into blockchain, which enables the copyright information to be stored decentralized. It improves the efficiency of copyright authorization and tracing, and provides convenience for the creators and for users of images.

2) The system innovatively realized a recommendation strategy based on machine learning and token voting mechanism, namely "resource reverse tracing strategy". Copyright use information is timely fed back to copyright owners and "we media" platform to realize early warning of infringement. This idea can also be used in other fields, such as electronic invoice (e-invoice) system, for detecting falsification of e-invoice, issuing invoice in time, and preventing the reuse of e-invoice.

Acknowledgments

Authors wishing to acknowledge financial support from National Key R&D Program of China(2018YFB0803900), Beijing Municipal Natural Science Foundation(4212019), Guangxi Key Laboratory of Cryptography and Information Security (GCIS201808).

References

[1] L. J. Xiao, W. H. Huang, Y. Xie, et al. A blockchain-based traceable IP copyright protection algorithm. *IEEE Access*, 2020, 49532-49542.

[2] G. C. Yuan, L. P. Fengng, J. Ning, et al. Survey on the application of blockchain in Digital Rights Protection. *International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI)*, 2020, 183-187.

[3] B. W. Wang, J. W. Shi, W. S. Wang, et al. A blockchain-based system for secure image protection using zero-watermark. *IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems*, 2020, 62-70.

[4] Z. H. Lu, Y. Q. Shi, R. Tao, et al. Blockchain for digital rights management of design works. *IEEE 10th International Conference on Software Engineering and Service Science (ICSESS)*, 2019, 596-603.

[5] D. Roman, K. Vu. Enabling data markets using smart contracts and multi-party computation. *International Conference on Business Information Systems*, Berlin, Germany, 2018, 258-263.

[6] R. A. Dobre, R. O. Preda, R. A. Badea, et al. Blockchain-based image copyright protection system using JPEG resistant digital signature. *IEEE the 26th International Symposium for Design and Technology in Electronic Packaging (SIITME)*, 2020, 206-210.