MaSRChain: A Trusted Manuscript Submission and Review System Based on Blockchain

Fengqi Li(✉), Kemeng Liu, Haoyu Wu, and Xu Zhang

School of Software Technology, Dalian University of Technology, Dalian, China
lifengqi@dlut.edu.cn

Abstract. Manuscript submission and review (MaSR) systems play an important role in scholarly publishing. However, there are some problems to be solved. Authors cannot gain an authoritative copyright certificate of manuscripts. Journals and conferences cannot achieve effective detection of multiple contributions with one manuscript. Reviewers may intentionally submit malicious evaluations due to competition. In this paper, we propose a trusted decentralized manuscript submission and review system based on blockchain (MaSRChain) to solve problems above. At first, we use blockchain and Attribute-Based Encryption (ABE) to protect manuscript copyright and realize access control of manuscripts for authors. Secondly, we utilize blockchain to realize manuscript sharing that encrypted by Locality Sensitive Hash (LSH), which can achieve multiple contributions detection among different institutions. Thirdly, we apply Ring Signature to realize authentication of review evaluations, while providing some anonymity to reviewers. Finally, we conduct experiments based on Hyperledger Fabric and experimental results demonstrate the effectiveness and efficiency of the system.

Keywords: Manuscript submission and review system · Blockchain · Copyright protection · Multiple contributions detection · Anonymous peer review

1 Introduction

MaSR systems have become a central fixture in scholarly publishing. However, current centralized systems bring lots of risks for all parties. At first, after the submission, authors lose the control of manuscripts. When dishonest program committee members or peer reviewers plagiarize unpublished work [1], it is difficult for authors to prove their ownership of the work. Moreover, some dishonest authors submit one manuscript to multiple institutions [2]. However, it is impossible to detect multiple contributions because they cannot share unpublished manuscripts. Finally, reviewers may deliberately submit negative review evaluations due to the lack of effective constraints [3].

To solve these problems, we propose MaSRChain. Firstly, we design a manuscript copyright protection protocol based on the tamper-proof infrastructure of blockchain and ABE. The manuscript encrypted by ABE is recorded into the blockchain, which

Supported by The Fundamental Research Funds for the Central Universities (NO. DUT19ZD209).

X. Wang et al. (Eds.): APWeb-WAIM 2020, LNCS 12318, pp. 18–26, 2020.
https://doi.org/10.1007/978-3-030-60290-1_2

realizes copyright protection and access control to manuscripts for authors. Secondly, we propose a detection method of multiple contributions based on the distributed features of blockchain and LSH. The manuscript hashed by LSH can be shared among different institutions. So, the multiple contributions can be detected on the premise of protecting the confidentiality of manuscript. Finally, we realize accountable and anonymous review protocol based on blockchain and Ring Signature. The tamper-proof storage of review evaluations restricts reviewers to review manuscripts more fairly, while Ring Signature helps authors verify the authenticity of evaluations anonymously.

The paper is organized as follows. The preliminary is introduced in Sect. 2. Section 3 describes the system model. The system analysis and experiment are described in Sect. 4 and Sect. 5. We list related work and conclude this paper in Sect. 6 and Sect. 7.

2 Preliminary

2.1 Distributed Ledger Technology

MaSRChain is built on the Hyperledger Fabric [4]. In Fabric, *Peer* interfaces with applications, executes smart contracts. *Orderer* sorts transactions, and ensures data consistency of the whole network. Inter Planetary File System (IPFS) is a peer-to-peer distributed file system [5]. it is high-capacity, content-addressable and allocates a unique identifier for stored file, which makes up for the limited storage space of blockchain.

2.2 Locality Sensitive Hash

we utilize Simhash [6] and Perceptual Hash (PHash) [7] to detect the similarity of text and figures in different manuscripts respectively. After tokenizing, hashing, weighting, summation and dimensionality reduction, we get the Simhash value of the text with N bits length. In addition, after reducing size, simplifying color, getting lowest frequency matrix and calculating average, we get the PHash value of figures. Finally, the algorithm uses Hamming distance to judge the similarity of text and figures in manuscripts.

2.3 Attribute-Based Encryption

we use DPUPH-CP-ABE [8] to realize protection and access control of manuscripts for authors. The DPUPH-CP-ABE consists of five algorithms.

Set(1^λ, U)$\rightarrow$*pk*, *msk*. Inputs are security parameter λ and attribute universe description U. It outputs public key pk and master key msk.

Encr(*pk*, *M*, $\mathbb{A}$)$\rightarrow$*ct*. Inputs are pk, a message M and access structure $\mathbb{A}$. It outputs ciphertext ct and E_{info} which is encrypted information about message M.

KeyG(*pk*, *msk*, *S*)$\rightarrow$*sk*. Inputs are pk, msk and attributes S. It outputs secret key sk.

Decr(*pk*, *sk*, *ct*)$\rightarrow$*M*. Inputs are pk, sk and ct. It outputs the message M.

Update(E_{info}, $C_i^{(2)'}$)$\rightarrow$*ct'*. Inputs are E_{info} and $C_i^{(2)'}$ which is calculated from $C^{(2)}$ in ct. It outputs a new ciphertext ct'.

2.4 Ring Signature

Ring Signature [9] allows the authorization of a collection of identities to perform an action, while maintaining the privacy of the specific identity that performed the action. The Ring signature consists of three algorithms.

RKeyGen(P)→x_i, y_i. Inputs is a big prime number P. It outputs the public key y_i and secret key x_i for reviewers.

RSign(m, L, x_i)→σ. Inputs are the message m, a public key set L of n no-signers and the secret key x_i of the actual signer. It outputs the signature result σ.

RVeri(m, σ)→{1|0}. Inputs are m and σ. It outputs the verification result.

3 System Overview

3.1 Overview

MaSRChain aims to provide a distributed, tamper-proof and verifiable solution for manuscript copyright protection, multiple contributions detection, and accountable and anonymous peer review. As shown in Fig. 1, there are five entities in system: *authors*, *editors*, *reviewers*, *permissioned blockchain* and *external storage*. The permissioned blockchain consists of publishers, journals and conference agents based on Fabric. The publishers are *orderer*, while journals and conference agents are *peer* in different *organizations*. The solution can be divided into five phases.

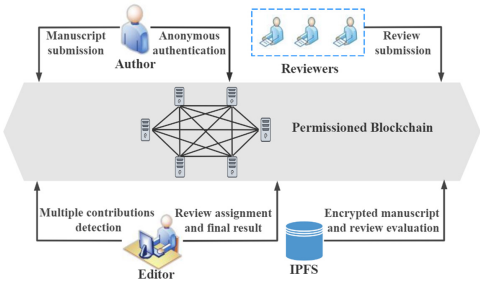

Fig. 1. Overview of MaSRChain system.

Manuscript Submission. Authors upload manuscripts encrypted by DPUPH-CP-ABE to IPFS and blockchain. Then, the system provides a copyright certificate to authors.

Multiple Contributions Detection. When journals or conferences receive manuscript, the editor executes multiple contributions detection to avoid wasting review resources.

Review Assignment. The editor submits review assignment to the blockchain. Then, the author chooses capable reviewers in this scope to review this manuscript.

Review Submission. The reviewers submit evaluations signed by Ring Signature to IPFS and Blockchain. The tamper-proof storage of evaluations can restrict reviewers to review more fairly.

Final Notification and Anonymous Authentication. The editor sends final notification to authors. The author can verify the authenticity and creditability of evaluations.

3.2 Manuscript Copyright Protection Protocol

To realize copyright protection and access control to manuscripts for authors, we design a manuscript copyright protection protocol based blockchain and DPUPH-CP-ABE. Next, we will explain this protocol in detail by submitting a manuscript in Fig. 2.

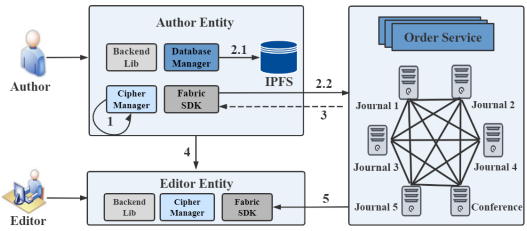

Fig. 2. The process of submitting a manuscript.

- ***Step 1.*** Before submitting a manuscript, the author entity runs $\boldsymbol{Setup}\left(1^{\lambda}, \boldsymbol{U}\right)$ algorithm to generate pk and msk and encrypts the manuscript by $\boldsymbol{Encrypt}(pk, M, \mathbb{A})$ algorithm to generate the ciphertext of the manuscript $ct = \left(C, C^{(1)}, \left\{C^{(2)}\right\}_{i\in[1,l_s]}, C^{(3)}\right)$.
- ***Step 2.*** Then, the author entity submits ct to IPFS and submits manuscript information (including title, unique identifier returned by IPFS and the fingerprint of manuscript) to the blockchain by invoking manuscript submission smart contract (SC).
- ***Step 3.*** If the fingerprint of manuscript is not similar with the published manuscripts, the blockchain will provide a copyright certificate to the author.
- ***Step 4–5.*** Finally, the author entity sends pk and msk to the editor, and the system will send submission notice to the editor.

3.3 Multiple Contributions Detection and Review Assignment Protocol

To realize multiple contributions detection and choose suitable reviewers to review manuscripts, we propose a multiple contributions detection and review assignment protocol. Next, we will explain this protocol in detail in Fig. 3.

- ***Step 1.*** Editor invokes viewing manuscript SC to get the manuscript information. Then, the editor gets the ciphertext of manuscript through unique identifier from IPFS.

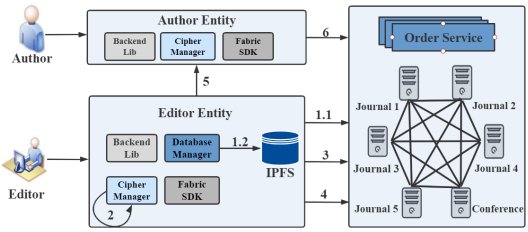

Fig. 3. The process of multiple contributions detection and review assignment.

- ***Step 2.*** The editor runs ***keyGen***$(\boldsymbol{pk}, \boldsymbol{msk}, \boldsymbol{S})$ algorithm generate the private key sk, and runs ***Decrypt***$(\boldsymbol{pk}, \boldsymbol{sk}, \boldsymbol{ct})$ algorithm to get the plaintext of manuscript.
- ***Step 3.*** Subsequently, the editor invokes multiple contributions detection SC to detect multiple contributions. The system calculates the fingerprint $f = (h_s, h_p)$ of manuscript containing Simhash and PHash value. Then, MaSRChain compares the Hamming distance $dis = (f_x, f_y)$ of the fingerprint with that of submitted manuscripts in the blockchain to detect whether it is multiple contributions.
- ***Step 4.*** After multiple contributions detection, the editor invokes review assignment SC to submit review assignment to the blockchain, which contains 5 reviewers' pseudonymous identity attributes and brief introduction to reviewers but no personally identifiable information.
- ***Step 5.*** The editor informs author to choose 3 reviewers as he thinks suitable to review his manuscript.
- ***Step 6.*** Then, the author gets pseudonymous identity attributes of reviewers he chose from blockchain and runs $\boldsymbol{Update}\left(\boldsymbol{E_{info}}, \boldsymbol{C_i^{(2)\prime}}\right)$ to update access policy and generate new ciphertext $ct^{'}$, Finally, the author submits new ciphertext to IPFS and blockchain.

3.4 Accountable and Anonymous Review Protocol

To restrict reviewers to review manuscripts more fairly, and verify the authenticity and creditability of evaluations on the premise of anonymity, we design an accountable and anonymous review protocol based blockchain and Ring Signature as shown in Fig. 4.

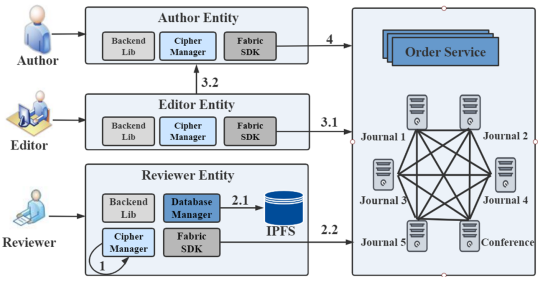

Fig. 4. The process of review evaluations submission and anonymous authentication.

- ***Step 1.*** When reviewers finish review evaluations, they run ***RKeyGen***(P) algorithm to generate temporary secret key σ_i and temporary public key y_i. Then, the signer runs ***RSign***(m, L, x_i) to generate the Ring Signature result σ.
- ***Step 2.*** The reviewer submits the review evaluation to the IPFS and uploads unique identifier, and Ring Signature result σ to the blockchain.
- ***Step 3.*** When all reviewers have uploaded their evaluations, the editor invokes submitting final notification SC to upload the final review result of this manuscript, and informs the author final decision through the system.
- ***Step 4.*** When the author receives the final review result of manuscript, he runs ***RVerify***(m, σ) to verify the authenticity and creditability of evaluations that submitted by the reviewers of his choice under the condition of anonymity.

4 Security and Privacy Analysis

It is important to protect the copyright and content of manuscript. The timestamp and tamper-proof infrastructure of blockchain prove that the author begins to own this manuscript at a specific time. Moreover, the manuscript is encrypted by DPUPH-CP-ABE to protect the content of manuscript, and DPUPH-CP-ABE is indistinguishable under chosen plaintext attack, which meets the security requirement in scholarly publishing.

Now, double blind reviews are becoming more and more common. The anonymity of author can be achieved simply by hiding authors' identity information. In MaSRChain, the author needs to authorize reviewers to access the manuscript. To prevent the author from getting reviewers' identity information, we utilize pseudonymous identity attributes of reviewers to construct DPUPH-CP-ABE cryptosystem, which realizes access authorization without knowing reviewers' identity.

5 Performance Evaluation

MaSRChain is built on Fabric 1.2, which is composed of ten computers running Ubuntu 16.04 and equipped with I7-6700 processor with 3.4 GHz and 8 GB memory. There are four *orderer* and six *organizations*. To fully reflect the performance of the system, we have tested the performance of algorithms utilized in this paper and blockchain system.

5.1 Evaluation of Main Algorithm

we test the performance of DPUPH-CP-ABE with 2 attributes, the performance of Simhash and PHash algorithm with 64 bits length, and the performance of Ring Signature with 5 users. The size of manuscript ranges from 0.5 MB to 5 MB, while the size of figure is 64×64, 128×128, 256×256, 512×512, and 1024×1024 pixels respectively.

Figure 5 shows that the time consumed by ***Encrypt***, ***Decrypt*** and ***Update*** algorithms in DPUPH-CP-ABE increases gradually with the change of data size. The average consumption time is about 72 ms, 31 ms and 6 ms respectively.

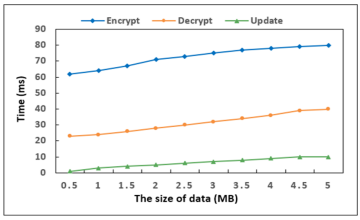

Fig. 5. The performance of DPUPH-CP-ABE.

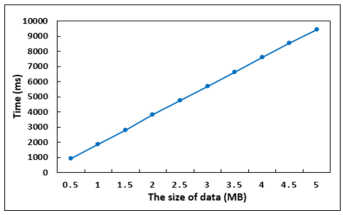

Fig. 6. The performance of Simhash.

Figure 6 and Fig. 7 show that the time consumed by Simhash increases linearly from 900 ms to 9500 ms, and the time consumed by PHash increases exponentially from 10 ms to 210 ms with the change of data size. Although Simhash takes some time, it is tolerated in manuscript submission scenario.

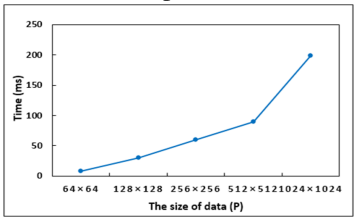

Fig. 7. The performance of PHash.

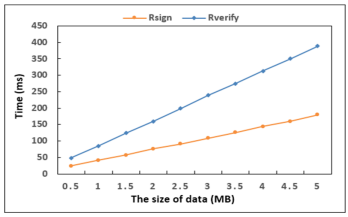

Fig. 8. The performance of Ring Signature.

Figure 8 shows that the time consumed by ***Rsign*** and ***Rverify*** algorithms in Ring Signature increase linearly with the change of data size. The average consumption time is about 218 ms and 101 ms respectively, which meets system requirements.

5.2 Evaluation of Query and Submit Operation

In MaSRChain, there are two types of operations: *query* operation from blockchain and *submit* operation to blockchain. Next, we test the transaction response time and confirmation time with the different number of concurrent transactions per second. Moreover, the consensus mechanism is *Kafka*, batch timeout is 2 s, and block size is 32 KB.

Figure 9 depicts the relationship between response time of *query* operation and number of concurrent transactions. The response time increases slowly at beginning until throughput reaches 300 tps. After that, the response time increases rapidly, higher the number of concurrent transactions brings higher response time duo to processing bottleneck.

Figure 10 shows the relationship between transaction conformation time *submit* operation and the number of concurrent transactions. When the number of concurrent transactions is small, the system needs to wait for batch time to pack a new block. Then, the transaction confirmation time decreases gradually with the change of throughput from 100 tps to 300 tps. This is because the threshold of block size is met, and a new block will be generated before the predefined batch time. As the number of concurrent transactions continues to increase, it exceeds processing capability and transactions cannot be confirmed in time which causes transaction confirmation time increases gradually.

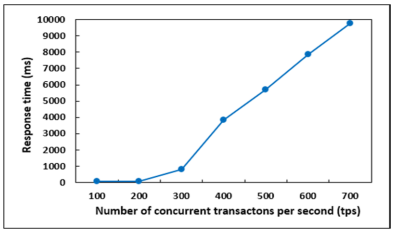

Fig. 9. The performance of *query* operation.

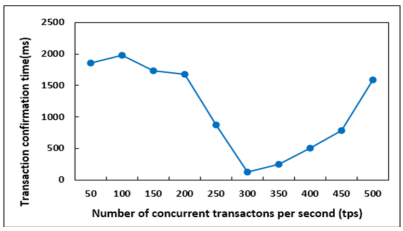

Fig. 10. The performance of *submit* operation.

6 Related Work

Some researchers propose decentralized MaSR systems based on blockchain. Bela G et al. propose CryptSubmit [10]. It uses the trusted timestamp of Bitcoin to provide authors with a reliable certificate. However, to reduce cost, CryptSubmit collects submitted manuscripts from one day, and submits the hashes of manuscripts together to Bitcoin, which prevents it from realizing copyright confirmation in real time. To realize multiple contributions detection, Nitesh E et al. propose a blockchain-based solution that all journals or conferences implement a shared ledger to share the title of submitted manuscripts [11]. However, these solutions are useless because it is easy to change the title of the manuscript and submits to other institutions. Then, For the accountable review mechanism, there are some decentralized publication systems for open science can record review evaluations into blockchain [12]. Although evaluations that are stored in the blockchain cannot be tampered with, authors cannot ensure that the evaluations received are from responsible peer reviewers.

7 Conclusion

In this paper, we proposed MaSRChain to solve common academic misconduct. It can realize manuscript copyright protection, multiple contributions detection, and accountable and anonymous peer review at the same time on the premise of protecting the confidentiality of the manuscript and not affecting the fairness of review. Besides, experimental results demonstrate the performance of system meets the actual requirement.

References

1. Dansinger, M.: Dear plagiarist: a letter to a peer reviewer who stole and published our manuscript as his own. Ann. Intern. Med. **166**(2), 143 (2017)
2. Tie-cheng, J.I.N.: A review on the research of phenomenon of multiple contributions with one manuscript of journal. J. Henan Univ. Technol. (2005)
3. Lee, C.J., et al.: Bias in peer review. J. Am. Soc. Inf. Sci. Technol. **64**(1), 2–17 (2013)
4. Androulaki, E., et al.: Hyperledger fabric: a distributed operating system for permissioned blockchains. In: Proceedings of the Thirteenth EuroSys Conference, pp. 1–15 (2018)
5. Benet, J.: Ipfs-content addressed, versioned, p 2p file system. arXiv preprint arXiv 1407, 3561 (2014)

6. Manku, G.S., et al.: Detecting near-duplicates for web crawling. In: Proceedings of the 16th International Conference on World Wide Web, pp. 141–150 (2007)
7. Venkatesan, R., et al.: Robust image hashing. In: Proceedings 2000 International Conference on Image Processing, pp. 664–666. IEEE (2000)
8. Ying, Z.B., et al.: Partially policy hidden CP-ABE supporting dynamic policy updating. J. Commun. **36**(12), 178–189 (2015)
9. Rivest, R.L., Shamir, A., Tauman, Y.: How to Leak a Secret. In: Boyd, C. (ed.) ASIACRYPT 2001. LNCS, vol. 2248, pp. 552–565. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-45682-1_32
10. Gipp, B., et al.: Cryptsubmit: introducing securely timestamped manuscript submission and peer review feedback using the blockchain. In: JCDL2017, pp. 1–4. IEEE (2017)
11. Emmadi, N., Maddali, L.P., Sarkar, S.: MaRSChain: framework for a fair manuscript review system based on permissioned blockchain. In: Mencagli, G., et al. (eds.) Euro-Par 2018. LNCS, vol. 11339, pp. 355–366. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-10549-5_28
12. Tenorio-Fornés, A., et al.: Towards a decentralized process for scientific publication and peer review using blockchain and IPFS. In: Proceedings of the 52nd Hawaii International Conference on System Sciences (2019)