# PubChain: A Decentralized Open-Access Publication Platform with Participants Incentivized by Blockchain Technology

Taotao Wang
*College of Electronics and Information Engineering*
*Shenzhen University*
Shenzhen 518060, China
ttwang@szu.edu.cn

Soung Chang Liew
*Department of Information Engineering*
*The Chinese University of Hong Kong*
Hong Kong SAR, China
soung@ie.cuhk.edu.hk

Shengli Zhang
*College of Electronics and Information Engineering*
*Shenzhen University*
Shenzhen 518060, China
zsl@szu.edu.cn

***Abstract*—We design and implement Publication Chain (PubChain), a decentralized open-access publication platform built on decentralized and distributed technologies of blockchain and IPFS peer-to-peer file sharing systems. The existing publication platforms have some severe drawbacks. First, instead of promoting widespread knowledge sharing, access to publications on the platforms owned by publishers is often on a fee basis. This drawback of pay wall prevents researchers from "standing on the shoulders of giants". Moreover, the peer review process on most all existing publication platforms (including both open-access and publisher platforms) is prone to be ineffective, since there is no proper incentive to reviewers for performing high-qualified reviews. PubChain is an alternative platform to the existing publication venues aiming to address their drawbacks. No central third-party owns the contents (i.e., papers and reviews) of PubChain. Exploiting blockchain technology, we devise an elaborate incentive scheme on PubChain to incentivize key stakeholders (i.e., authors, readers and reviewers) to participate publication activities on PubChain in a substantive manner by earning credits and rewards through self-motivated interactions. We have performed simulations to investigate the robustness of our proposed incentive scheme against fraudulent publications and reviews. We also have implemented a prototype of PubChain to demonstrate its key concepts.**

## I. Introduction

Publications of research results are an important activity to disseminate new knowledge. "Standing on the shoulders of giants" is a vivid expression that points out that new discoveries and innovations are often built on prior work by others [1], [2]. Researchers thrive on free exchange of information.

### *A. Drawbacks and Limitations of Existing Publication Platforms of Publishers*

To date, the most successful venues for academic paper publication are journals and magazines owned by large entrenched publishers, such as Nature Publishing Group, Institute of Electrical and Electronic Engineers (IEEE), Association for Computing Machinery (ACM) and Elsevier of RELX Group. These publishers publish a huge number of research papers every year. Their journals and magazines are platforms on which researchers exchange their latest research results and where latest research breakthroughs are announced. Despite their success, these publication platforms have significant drawbacks and limitations from the standpoint of key players—authors, reviewers, and readers—that matter most.

*1) Pay Wall:* The power to publish, store and share academic literature is concentrated in the hands of a few dominant publishers. These publishers are for-profit outfits. They charge authors for publications on their venues and they charge readers for accessing the papers. In other words, they charge both the producers and consumers. Furthermore, conferences organized by some publishers often charge exorbitant registration fees for conference attendance, and outrageous sums of money for page charge for pages that do not incur much additional cost on their electronic platforms. They get away with these exploitations because they can. They have built up their brands over the years.

But who help them build and maintain their brands? Well, they leverage the free service of editors and reviewers to maintain the quality of the publications. In most businesses, workers who do work receive compensation rather than the other way round. Publication business is an exception—publishers charge both the workers (the authors) as well as the customers (the readers) and receive free services from both the workers and the customers (the authors and readers themselves often serve as the reviewers).

Their charges can be quite expensive to the extent that only large organizations, such as corporations, research institutions and universities, can afford the fees. The pay wall put up by the publishers excludes small organizations and individuals from accessing the latest research publications. These publishers stand in the way of knowledge dissemination and the pay wall prevents a level playing field among researchers.

This work was supported in part by the National Natural Science Foundationof China under grants 61701311, in part by the Guangdong NSF Project under Grant 2020A1515010708 and in part by the Shenzhen NSF Project under Grant JCYJ20170818095107583

T978-1-7281-5628-6/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: San Francisco State Univ. Downloaded on June 16,2021 at 00:43:48 UTC from IEEE Xplore. Restrictions apply.

*2) Information Island:* The authors are forced to transfer the copyrights of their papers to the publishers. The publishers typically do not mutually share their literature resources. This gives rise to information islands with unsynchronized contents. There are many intrinsic disadvantages associated with such isolated information islands. Readers and researchers lacking resources will have difficulty getting a complete set of past papers unless they subscribe to all these publishers. These islands are hurdles to knowledge dissemination.

*3) Disintegration of Peer Review Process:* Peer reviews of papers should be performed by experts with the same level of competence as the authors of the papers in their particular field. The peer review process is crucial to maintaining paper quality. As a rule of thumb, journals and conferences with a rigorous peer review process and with a low paper acceptance rate are considered to be more prestigious by readers and authors.

With the growth of research participants, research papers are also growing exponentially. It is getting increasingly difficult to find quality reviewers to review the large number of papers. Competent reviewers are researchers themselves. As researchers, they need to balance their time between reviewing others' papers and doing their own research. Unless these papers are directly related to their current research topics, they have little incentive to do the review, even if they have the technical expertise to do so. Review is a form of "technical auditing" as far as scientific papers are concerned.

When accountants perform financial auditing for corporations and organizations, they often charge a large sum of money for their service, and as such they are obligated to do a professional job that meets a certain minimum quality threshold. Otherwise, the accountants would not receive future jobs. When reviewers perform technical auditing, reviewers receive zero compensation, and the quality of review varies much from reviewer to reviewer. There are no incentives other than the conscience of the reviewers to meet certain minimum quality target. Arguably, serious technical auditing can be a lot more time-consuming than financial auditing. Why should technical auditing be free? Are scientists worth less than accountants?

Without proper incentives, there is little reason for reviewers to spend time on paper review. As a result, because of paper explosion, many reviews are quite shallow in nature, even for prestigious venues such as IEEE. Many senior researchers (e.g., professors) may relegate the responsibility of paper review to junior novice researchers (e.g., beginning graduate students of the professors) who at least have the incentive to review papers as part of their learning process—some of them probably have no choice because their superiors ask them to do the job. Where did the money—page charges, membership fees—go? Did any go to those responsible for quality assurance? Without proper incentives to reviewers, the current peer review process can break down easily, especially in the face of paper explosion [3], [4].[1]

### B. Other Publication Platforms and Services

Literature search and citation index services, such as web of science and google scholar, can partially overcome the information island effect. Papers from multiple publishers can be listed and their citations can be indexed. Since these services do not really publish papers, they still cannot overcome the handicaps of pay wall and peer review disintegration.

To overcome the pay wall of publishers, Free Open Access aims to make academic literature a free public resource on a global scale. For example, the arXiv preprint system allows authors to upload their papers for free access by all researchers. By the year of 2014, more than 1 million articles have been uploaded on arXiv [5]. The founder of arXiv, physicist Paul Kingsbagh, won the 2002 MacArthur award for his contribution to Free Open Access. Although Free Open Access platforms allow everyone to access research outputs freely and easily, they still suffer from peer review disintegration. In fact, arXiv does not even have a peer review process. Low-quality papers abound on Free Open Access platforms. As of today, papers published on Free Open Access platforms do not earn the same prestige that papers on the publication platforms of Publishers.

Some open access platforms such as Peerj also support peer review openly—review comments can be posted along with the published papers. However, even with this open peer review scheme, reviewers still lack incentives for investing efforts to provide high quality reviews. Therefore, the open peer review scheme does not address the drawback of peer review process's disintegration.

All publication platforms today are centralized—they are owned or managed by a single organization. As a consequence, they are prone to single points of failure—there is no guarantee that the organization will never close the access to the database. The power to publish, store and share academic literature is concentrated in the hands of publishers and owners of open-access platforms.

Using blockchain technology, we propose Publication Chain (PubChain) that aims to overcome the limitations of the current publication platform. Papers posted on PubChain by authors are automatically distributed to a distributed file sharing system and are registered in a decentralized blockchain. PubChain is a decentralized publication platform, where authors, readers and reviewers are incentivized to participate in a meaningful and substantive manner. In particular, these key players can earn credits and rewards through self-motivated interactions. The assets of PubChain are owned by these key players, not by a separate profit-focused publisher. PubChain does not own the copyrights of the papers; the authors retain their copyrights. PubChain is not a central authority. The authors do not need permissions to publish papers on PubChain. We next discuss why we choose blockchain as our solution.

[1] Besides the lack of incentives for reviewers, other limitations of the current peer review system are discussed in [3], [4]. For example, the review process can be slow and cumbersome; it also exhibits various forms of bias.

Authorized licensed use limited to: San Francisco State Univ. Downloaded on June 16,2021 at 00:43:48 UTC from IEEE Xplore. Restrictions apply.

TABLE I
COMPARISON OF DIFFERENT PUBLICATION PLATFORMS

| | Pay Wall | Single Point of Failure | Disintegration of Peer Review Process |
|---|---|---|---|
| Publishers | × | × | × |
| Free Open Access | √ | × | × |
| Consortium Web App | √ | √ | × |
| PubChain | √ | √ | √ |

×: having such problem; √: having no such problem.

### C. The Motivation and Justification for the use of Blockchain

Blockchain is a decentralized and distributed digital ledger that stores data in chronological order in a way that can ensure that the data in the chain cannot be falsified [6]–[8]. It is natural to employ blockchain to solve the problems of pay wall and single point of failure. Although other solutions (e.g., a web application run by a consortium) can also address some problems of the existing publication platforms, i.e., pay wall (if the consortium's platform subscribes to free open access), single point of failure, there are still a number of advantages enjoyed by blockchain that cannot achieved by these solutions.

Using blockchain, our PubChain can construct a decentralized publication platform that does not belong to any entities. Besides solving the problems of pay wall and single point of failure, the decentralized publication platform can establish an ecosystem for publication activities, where authors, reviewer and readers are incentivized to make positive contributions.

We believe that there is no free lunch—but good lunch should not cost too much, and people who contribute to the lunch quality should be rewarded. For certain production and sales activities, there must be people who pay money (the ones who enjoy the services/products) and people who receive the money (the ones who provide the services/products). Authors need to pay for their publications, due to that they use others' services (i.e., platforms' publication services, reviewers' review services). However, on all current publishers' platforms (i.e., conferences, journals), authors just pay the publishers and not the reviewers, and fees paid to publishers are rather high. On free open-access platforms, authors do not pay anybody and there is no review service provided. Besides the platforms' publication services, reviewers' services are also important to paper quality assurance. Reviewers provide some sort of a "consumer reports" service. However, reviewers on current publication platform lack incentives to perform high quality reviews. Therefore, we design PubChain to incentivize reviewers to provide high quality reviews by letting them to receive an amount of financial rewards commensurate with the quality of their reviews.

How to credit reviewers is not an easy task, since how to evaluate their review comments and scores is not straightforward. Using blockchain technology, we design an incentive mechanism for authors, reviewers and readers that incentivizes these participants to form an ecology that promotes paper quality and achieves free open access at the same time (see Section III). In particular, we devise a decentralized scoring mechanism that is robust to dishonest scoring behaviors (see Section IV). We compare the current existing publication platforms, a web application run by a consortium and our PubChain by indicating the problems that can be solved by them in TABLE I.

### D. Related Work and Our Contributions

Refs. [3], [4] discussed the limitations of the current publication platforms, especially the limitations imposed by the peer review process. Works [9], [10] proposed to incentive reviewers using digital cryptocurrency of blockchain. Work [11] proposed to permanently record educational records (i.e., exam credentials, record of learning, the authorship of something) and the corresponding reputation rewards on blockchain. Works [12] and [13] proposed to use blockchain to record the copyright of publications and use peer-to-peer file storage system IPFS to store the publications.

These existing works mainly talked about the idea of using blockchain to record the copyright of papers and/or issuing tokens to authors/reviewers. However, a complete framework with rigorous designs is still lacking. Our PubChain is a complete framework with rigorous designs for implementing blockchain based publication platform. Compared to these existing works, PubChain has the following new contributions.

- First, PubChain has an incentive mechanism to authors, reviewers and readers that incentivizes these participants to form a self-sustaining ecology that promotes paper quality on an open access platform (Section IV).
- Moreover, PubChain has a decentralized scoring mechanism that is robust to dishonest scoring behaviors (Section IV).
- Finally, we consider many practical aspects of the system, including the system architecture and the financial model (Section II and Section III of our technical report [14]).

The remainder of the paper is structured as follows. Section II overviews the system design of PubChain. Section III presents the incentive scheme of PubChain. Section IV discuss the decentralized scoring mechanism of PubChain. Section IV gives the current implementation of PubChain. Finally, Section V concludes the paper.

## II. SOLUTION OF PUBCHAIN

### A. Design Concept

A central design concept of PubChain is to use blockchain [6]–[8] and off-chain peer-to-peer distributed file storage (i.e., InterPlanetary File System (IPFS) [15]–[17]) as building blocks to decentralize the publication platform. Such decentralization also means that there is no single central party that controls the running of the platform. If properly designed, the decentralized system can also be more robust than a centralized system given its replication of data across multiple parties.

PubChain uses the IPFS system as the database system for storing papers. IPFS is a distributed and decentralized storage system consisting of a network of peer-to-peer nodes. The techniques and features of IPFS can be found in [15]. With IPFS, papers are content-addressed in the database. Authors

Authorized licensed use limited to: San Francisco State Univ. Downloaded on June 16,2021 at 00:43:48 UTC from IEEE Xplore. Restrictions apply.

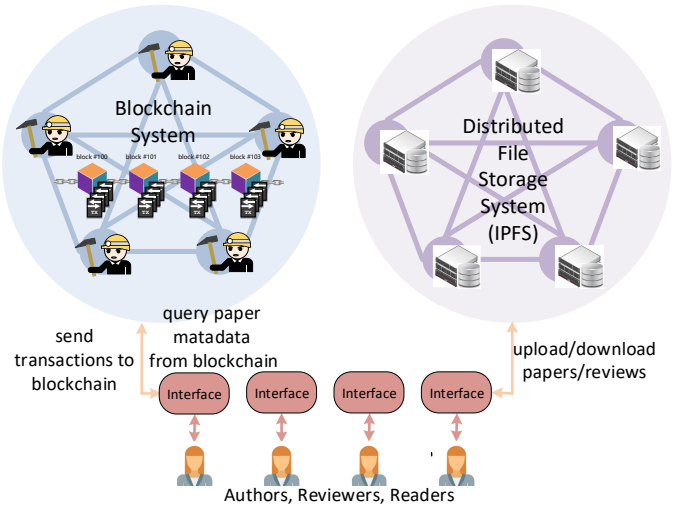


Fig. 1. The overview of PubChain platform, where we have three entities: a group of publication players (authors, reviewers, readers), a blockchain system sustained by miners, an IPFS system with distributed storage nodes.

can back up their papers to the network and freely download papers without the risk of single-point failure. The IPFS repository is physically owned by all users and not by a single entity.

PubChain exploits blockchain technology to confirm the registration of the paper ownership, to track index citations, and to incentivize participants. Blockchain is a distributed and decentralized append-only ledger for digital assets. Data in blockchain is replicated and shared among all the participants. Past records are made tamper-resistant through its append-only paradigm. There are many successful existing blockchain systems, e.g., Bitcoin [18], Ethereum [19]. We can reuse and modify their software code to build the blockchain of PubChain.

The operation of PubChain blockchain is divided into two consecutive phases, with the first phase being a temporary phase before the final second phase takes over. In the first phase, PubChain operates as a consortium blockchain using the Proof-of-authority (PoA) consensus protocol [20]. In the second phase, PubChain operates as a public blockchain using the Proof-of-work (PoW) consensus protocol [18].

Fig. 1 gives an overview of the PubChain platform. There are three entities in the platform: a group of publication players, a blockchain system sustained by miners, an IPFS system with distributed storage nodes. The blockchain system and the IPFS system are the infrastructure of PubChain. A network node can be a miner of blockchain or/and a storage node of IPFS. Blockchain miners run the distributed consensus protocol to maintain the data on blockchain. In the consortium blockchain phase, the miners are the super nodes that are selected to run the PoA protocol. In the public blockchain phase, the miners are the nodes that devote their computing powers to solving hash puzzles of the blockchain. IPFS storage nodes share their memory space for the distributed and persistent storage for PubChain. Through a PubChain interface, the publication players (i.e., authors, reviewers, and readers) interact with the blockchain and IPFS systems in the conduct of their activities on PubChain. We have developed a PubChain system that combines blockchain and IPFS. We will describe the system architecture of PubChain in Section III.

When an author uploads his/her paper to PubChain, the paper is time-stamped and registered on Pubchain as a permanent record. The citation index for every paper is also tracked on PubChain. Tokens are used to financially incentivize players to engage in publication activities on PubChain and to incentivize miners to sustain and maintain PubChain. We will elaborate our proposed incentive mechanism in Section IV.

The tokens issued by PubChain are called PubCoins. PubChain is a non-profit project and we will not sell the issued PubCoins through initial coin offering (ICO) and private placements to any other entity to make money. PubCoins will be distributed to all the participants as the rewards for their contributions to the platform, rather than to the project team or other organizations.

To endow PubCoin with real monetary value, we design PubChain as a side chain of another parent chain whose tokens are in wide circulation and are considered to have real monetary values, e.g., Bitcoin, Etherum, Bitcoin Cash. Using the two-way pegging technique of side chain [21], we can transfer the tokens on the parent chain to PubChain and vice versa. The technical details of two-way pegging and side chain can be found in [21]. At the beginning stage, PubChain operates separately from the parent chain, and PubCoin has no real monetary values. Donation to PubChain can be injected into PubChain from a parent chain using two-way pegging, and PubChain will then operate as a side chain after that. The details about the financial model of PubChain can be found in our technical report [14].

## III. Incentive Mechanisms of PubChain

Papers on PubChain will attract large readership only if it is a reputable publication platform. Authors and reviewers contribute toward making PubChain a quality publication platform by submitting high-quality papers and reviews. In that light, effective incentive mechanisms to encourage substantive and meaningful participation by authors and reviewers are a core part of Pubchain. In this section, we describe the incentive mechanisms of PubChain[2].

### A. Incentive Mechanism to Authors

In Pubchain, an author submits her/his paper via a transaction $T_{post}$. The transaction $T_{post}$ includes meta-information associated with the submitted paper, such as the author's address on the blockchain, the paper's IPFS address, title, keywords, and the transaction hashes of papers cited by the paper.

[2]We focus on the incentive mechanism to publication players here. The incentives to blockchain miners are the mining rewards and transactions fees. Filecoins [22] are the incentive to IPFS storage nodes who share their storage spaces.

Authorized licensed use limited to: San Francisco State Univ. Downloaded on June 16,2021 at 00:43:48 UTC from IEEE Xplore. Restrictions apply.

The author pays $X$ PubCoins in the transaction of each submitted paper.[3] A fraction of $a_1X$ tokens, $0 < a_1 < 1$, are allocated to a *reviewer bonus pool*. The tokens in the reviewer bonus pool are distributed to reviewers according to the mechanism presented in Section III.B. Another fraction of $a_2X$ tokens, $0 < a_2 < 1$, $a_1 + a_2 < 1$ are given to the papers cited by the submitted paper [4]. The remaining $(1 - a_1 - a_2)\,X$ tokens are taken as the transaction fees given to the miner that records the transaction onto blockchain.

In the public-chain phase of PubChain, every mined block contains a coinbase transaction that mints $Y$ new PubCoin tokens. Among these minted tokens, a fraction of $b_1Y$ tokens, $0 < b_1 < 1$, are released to the reviewer bonus pool, a fraction of $b_2Y$ tokens, $0 < b_2 < 1$, $b_1 + b_2 < 1$ are released to authors as rewards according to the following reward distribution mechanism, and the remaining $(1 - b_1 - b_2)\,Y$ tokens are released to the miner of that block.

To incentivize authors to submit quality papers, rewards are distributed to authors according to the review scores of their papers. Specifically, when a new block is mined, the author of paper $i$ receives a reward of $G_i$ PubCoins from the new block, computed as follows:

$$G_i = b_2Y \times \frac{\max\left(S_i - \lambda, 0\right)}{\sum_i \max\left(S_i - \lambda, 0\right)} \tag{1}$$

where $S_i$ is the current review score of paper $i$ (computation of the current review score $S_i$ of paper $i$ will be presented in Section IV.A), $\lambda$ is a quality threshold for papers, the summation of $i$ is over all papers that were been published on PubChain during the previous $M$ block intervals. In other words, a paper will be rewarded within a reward window of $M$ blocks after it is published on PubChain.

The review score $S_i$ of paper $i$ is initialized to zero, $S_i = 0$, and thus $\max\left(S_i - \lambda\ \ , 0\right) = 0$ initially. With the threshold $\lambda$, posting and overwhelming PubChain with low-quality papers whose review score does not pick up over time will not be rewarding. Furthermore, since $X$ PubCoins are charged for each posted paper, there is a disincentive for authors to submit low-quality papers.

Besides receiving rewards from good reviews, a paper can also receive rewards when it is cited by another paper. Specifically if paper $j$ cites $K$ other papers that are also posted on PubChain, $a_2X$ tokens paid by paper $j$ will be given to the authors of the $K$ cited papers. Each cited paper $i$ will receive $a_2X/K$ tokens from paper $j$. In this way, if a paper has a long lasting influence on other papers, it may continue to receive rewards through the citation mechanism (i.e., long after the review reward window has transpired, citation reward may continue).

### B. Incentive Mechanism to Reviewers

In Pubchain, a reviewer submits the review of a paper by sending a transaction $T_{review}$ to the blockchain. The transaction $T_{review}$ includes the hash of the transaction that publishes the paper, the numerical score of the paper given by the reviewer, and the blockchain address of the reviewer. In addition to the numerical score, reviewers can also write comments on papers. Insightful comments are useful to the authors in terms of improving future versions of their papers; they also let readers identify high-quality papers. The comments on papers are stored on IPFS and their IPFS address are included in the transaction $T_{review}$ that is sent to PubChain for record keeping.

We denote the review of paper $i$ by reviewer $j$ by $R_{i,j} \triangleq (Z_{i,j}, C_{i,j})$ where $Z_{i,j}$ is the numerical score and $C_{i,j}$ is the comments. PubChain treats the comments by reviewers as some sort of a "special paper" that are reviewed by readers – paper reviews are also reviewed, but with a numerical score only. The score of review $j$ depends on its review numerical scores given by readers. Readers will not give high scores to a paper review with only a numerical score without insightful comments. Review $j$ of paper $i$ receives a reward of $g_{i,j}$ PubCoins computed as follows:

$$g_{ij} = \alpha F \frac{\max\left(W_{i,j} - \lambda, 0\right)}{\sum_{i,j} \max\left(W_{i,j} - \lambda, 0\right)} \tag{2}$$

where $W_{i,j}$ is the current average score of comment $C_{i,j}$, $F$ is the total reward in the review bonus pool during the current block interval [5], $\alpha$ is a ratio ($0 < \alpha < 1$) that governs how much bonus in the current pool are distributed to reviewers during this block interval, the summation of $i$ and $j$ is over the comments recorded onto PubChain during the previous $M$ blocks. The bonus not used in the current block, $(1 - \alpha)F$, will be kept in the pool for release in subsequent blocks.

To incentive miners to include review transactions into their blocks, a fraction $\beta \ll 1$ of the reward obtained by a review $g_{ij}$ (i.e. $\beta g_{ij}$ tokens) is released to the miner who included this review transaction into its block. Therefore, during each block interval, $\beta\alpha F = \beta \sum_{i,j} g_{ij}$ tokens from the review bonus pool are released to miners who included review transactions associated with all reviews $j$ in the past $M$ blocks.

The incentive mechanism relies on the scores that can objectively reflect the qualities of papers and reviews. The next section will present a decentralized scoring system that can prevent malicious nodes from tampering with scores.

[3] On PubChain, when users generate the blockchain addresses, they must provide their affiliation emails or ORCID IDs, and each affiliation email/ORCID ID can only be linked to one address. At the beginning stage of PubChain, we will send a few tokens to each address for free, so that authors can upload their papers to the platform. After PubChain is widely accepted by authors and it has gained recognition as a reputable publications platform, we will stop issuing free tokens. From then on, to publish papers, authors will have to find ways to earn tokens (e.g., by serving as reviewers for others' papers) or else purchase tokens from others to pay the author fees. We believe the payment scheme will not reduce the number of submitted papers.

[4] Only authors of cited papers having a registered account with PubChain will be rewarded. Each Paper is identified by the hash of the transaction that registered the paper on the blockchain.

[5] In the consortium-chain phase of PubChain, $F = (1 - \alpha)\,F' + a_1XN$, where $F'$ is the total reward in the review bonus pool during the last block interval, $N$ is the number of the published papers and $a_1XN$ is the total amount of the tokens paid by the published papers during this block interval. In the public-chain phase of PubChain, $F = (1 - \alpha)\,F' + a_1XN + b_1Y$, where $b_1Y$ is the amount of tokens released to the review bonus pool by the current block interval.

Authorized licensed use limited to: San Francisco State Univ. Downloaded on June 16,2021 at 00:43:48 UTC from IEEE Xplore. Restrictions apply.

## IV. Decentralized Scoring System of PubChain

The financial rewards of PubChain are issued to authors and reviewers according to the scores of their papers and their reviews. To earn more rewards, malicious nodes may deliberately give scores that deviate from the true quality of papers and reviews. Therefore, a decentralized scoring system that can ensure objective scores in the presence of malicious nodes is very important. In this section, we first propose a decentralized scoring system to compute the scores of papers and reviews. We then perform simulations to investigate the integrity of the proposed decentralized scoring system.

### A. Decentralized Scoring System

The effective score $W_{i,j}$ of review $R_{i,j} = (Z_{i,j}, C_{i,j})$ is computed by averaging readers' scores on $R_{i,j}$. If review $R_{i,j}$ has received scores from less than $N_{rs}$ readers, its effective score is fixed to $W_{i,j} = 0$; otherwise, the effective score $W_{i,j}$ of review $R_{i,j}$ is obtained by excluding the highest and lowest 10% readers' scores and then averaging the remaining scores. To avoid conflict of interest, if a participant has submitted a review of a paper, she/he cannot score the other reviews of the same paper as a reader. In addition, to avoid score flooding, a reader can at most score $N_{rc}$ review comments of the same paper.[6] The effective score $W_{i,j}$ of review $R_{i,j}$ is for two purposes. First, it is used in (2) to incentivize reviewers to perform high-quality reviews. Second, it is used to compute the effective score $S_i$ of paper $i$.

We employ the review results of paper $i$, encoded in the form of $\{(Z_{i,j}, C_{i,j}, W_{i,j})\}_{j=1,2,\cdots}$ to compute the review score $S_i$ of paper $i$. First, we normalize the scores $W_{i,j}$ of review $R_{i,j}$ given by the readers as:

$$\widetilde{W}_{i,j} = \frac{W_{i,j}}{\sum_j W_{i,j}} \tag{3}$$

for all $j$. The normalized score $\widetilde{W}_{i,j}$ takes value between 0 and 1. Then, we compute $S_i$ as a weighted sum of scores $Z_{i,j}$ given by reviewers to paper $i$ using the normalized scores $\widetilde{W}_{i,j}$ as their weights:

$$S_i = \sum\nolimits_j \widetilde{W}_{i,j} Z_{i,j} \tag{4}$$

The computed $S_i$ is an evaluation on the quality of paper $i$ and is used to reward the author by the reward distribution mechanism. In essence, the effective score $W_{i,j}$ made by readers to review $R_{i,j}$ reflects the quality of that review and is an indication of the extent to which readers agree with the score $Z_{i,j}$ by reviewer $j$ on paper $i$.

[6]For implementation, we need a way to identify participants on PubChain and associate each participant ID with a unique address on blockchain. To achieve this, we can use the affiliation emails or ORCID IDs of the participants as their IDs on PubChain. Moreover, identifying reviewers using their affiliation emails or ORCID IDs means that our scoring system is a real-name system. This can avoid the dishonest review behavior in which a paper is intentionally given high scores by many reviewers from the same affiliation.

### B. Simulation Investigations

We next present simulation results to validate that our proposed decentralized scoring system can ensure fair reviews of papers, even in the presence of adversary reviewers with a biased interest.

Consider one poor-quality paper, paper $i$ with a ground-truth score of $S$. The author of this paper is an attacker who wants to gain more rewards by controlling a set of malicious nodes faking as reviewers and readers so that the paper can obtain a much higher score $S' \gg S$ on the platform. We assume the scores $\{Z_{i,j}\}_j$ for a paper $i$ given by honest reviewers are Gaussian distributed with mean $S$ and variance $\sigma_s^2$. The scores given by honest readers to a particular review $j$ of paper $i$ are Gaussian distributed with mean $W_P - |Z_{i,j} - S|$ and variance $\sigma_s^2$, where $W_P$ is the mean score for a "perfect review" that assigns the same score to the ground-truth score (i.e., $Z_{i,j} = S$).

We consider two strategies for the attacker. The first strategy is to have all malicious nodes serve as reviewers of the paper. All malicious nodes will give a high review score $Z_{i,j} = S'$, for all $j \in MS$, where $j \in MS$ is the set of the malicious nodes controlled by the attacker. In our simulations, we assume there are totally 1000 review scores given by reviewers to paper $i$, among which $N_{mn}$ scores are given by the malicious nodes. We assume each review $R_{i,j} = (Z_{i,j}, C_{i,j})$ is scored by $N_{rs}$ honest readers. Then, the effective score $W_{i,j}$ of review $R_{i,j}$ is obtained by first excluding the highest and lowest 10% scores from $N_{rs}$ readers' scores and then averaging the remaining scores. Finally, we compute the final score $S_i$ of paper $i$ according to (3) and (4). The results are shown in Fig. 2, where the final scores $S_i$ are evaluated with respect to different numbers of malicious nodes $N_{mn}$ and $\sigma_s^2 = 10$. We treat the simple average of the review scores, i.e., $S_i = \sum_j Z_{i,j}$, as our benchmark. In the simulations, we set $S = 40$, $S' = 80$ , $W_U = 90$ and $N_{rs} \in \{10,\ 100,\ 300,\ 600\}$. As we can see, our scoring method is robust to the attacker's fake reviews. When more and more malicious nodes are involved (large $N_{mn}$), the attacker becomes more successful in biasing the score toward the fake score. Large readership on the PubChain platform means large $N_{rs}$ , and large $N_{rs}$ makes the system more robust against large $N_{mn}$.

The second strategy is to have a fraction $\delta$ of the malicious nodes be fake reviewers and the rest be fake readers; and half of the fake readers will support the fake reviews by giving high scores, and the other half of the fake readers will attack the honest reviews by giving them low scores. For example, suppose that there are $N_{mn} = 100$ malicious nodes and $\delta = 0.1$. Then, 10 of the malicious nodes are fake reviewers that give review score $S'$ to paper $i$ and 90 malicious nodes are fake readers that can give a total of $90N_{rc}$ fake scores to all reviews of paper $i$. Among the $90N_{rc}$ fake scores to reviews, $45N_{rc}$ scores of $V_U$ are given to the fake reviews put up by the attacker (each of the 10 fake reviews is assigned with $45N_{rc}/10$ scores of $V_U$), where $V_U$ is a very high score used to support the fake reviews ; $45N_{rc}$ scores of $V_L$ are given

Authorized licensed use limited to: San Francisco State Univ. Downloaded on June 16,2021 at 00:43:48 UTC from IEEE Xplore. Restrictions apply.

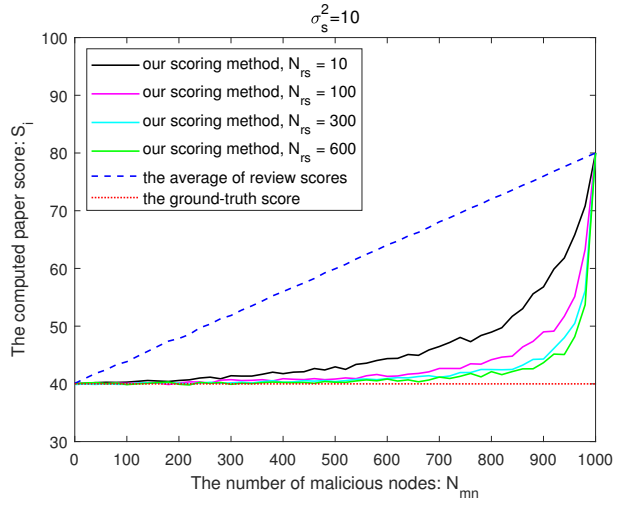


Fig. 2. The curve of $S_i$ vs. $N_{mn}$ with $\sigma_s^2 = 10$ for the first attack strategy.

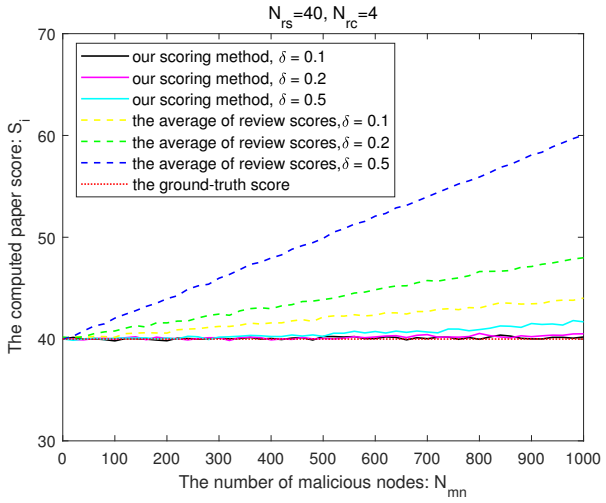


Fig. 3. The curve of $S_i$ vs. $N_{mn}$ with $N_{rs} = 40$ for the second attack strategy.

to the honest reviews (each honest review is assigned with $45N_{rc}/90$ scores of $V_L$), where $V_L$ is a very low score used to attack these honest reviews.

In the simulation, we set $S = 40$, $S' = 80$ , $W_P = 90$, $V_L = 20$, $V_U = 100$, $N_{rc} = 4$, $\delta \in (0.1\ , 0.2,\ 0.5\}$. Fig. 3 shows the results for $N_{rs} = 40$, respectively. From the results, we can observe that with large $N_{rs}$, our scoring method is still robust to this attack strategy.

## V. System Implementation

We have implemented a proof-of-concept prototype for the PubChain platform. The implementation of the blockchain reuses Ethereum, which means we can realize the virtual machine layer of PubChain using the EVM smart contract mechanism. The prototype uses IPFS for the storage layer. We have deployed the PubChain interface to a network node. Users (i.e., publication players) can use the JSON-RPC protocol to remotely deploy and invoke smart contract via this network node to conduct their activities on PubChain.

**Algorithm 1** Smart Contract for Posting Papers

```
1: struct Paper{
2:    string ownership;
3:    string paperName;
4:    string paperHash;
5:    string paperKeywords;
6: }
7: function stringsEqual(string storage _a, string memory
   _b) internal view returns (bool) {
8: bytes storage a = bytes(_a);
9: bytes memory b = bytes(_b);
10: if (a.length != b.length)
11:    return false;
12: for (uint i = 0; i ¡ a.length; i ++) {
13:    if (a[i] != b[i])
14:       return false;
15:    return true;
16:    }
17: }
18: function storeData(string userName_, string paper-
   Name_, string paperHash_, string paperKeywords_) pub-
   lic {
19:    require(!stringsEqual(paperInfo[paperName_].paperName,
   paperName_) && !stringsE-
   qual(paperInfo[paperName_].paperHash, paperHash_),
   "one paper one upload");
20:    paperInfo[paperName_] = Paper({
21:      ownership: userName_,
22:      paperName: paperName_,
23:      paperHash: paperHash_,
24:      paperKeywords: paperKeywords_,
25:    });
26: function getPaperHash(string paperName_) public view
   returns (string) {
27:    return paperInfo[paperName_].paperHash;
28: }
29: function getPaperOwnership(string paperName_) public
   view returns (string) {
30:    return paperInfo[paperName_].ownership;
31:    }
32: }
```

With smart contracts, we have implemented the functions of paper posting, paper reviewing, review scoring. The script codes of smart contracts are stored on blockchain. The smart contracts are triggered by transactions sent sent to their address on blockchain. Algorithm 1 shows the script codes of the smart contract that implements the function of paper posting. To post a paper on PubChain, an author needs to carry out the following procedure: 1) upload the paper (possibly including some program codes, multimedia materials) with her/his signature to the IPFS system and obtain the IPFS address of this paper (i.e., the paper hash); 2) include the publication information about the paper, i.e., its ownership (the address of the author on blockchain), IPFS address, paper

Authorized licensed use limited to: San Francisco State Univ. Downloaded on June 16,2021 at 00:43:48 UTC from IEEE Xplore. Restrictions apply.

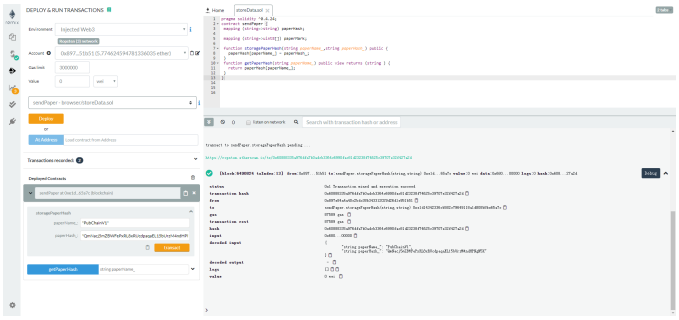

Fig. 4. The window of Remix Ethereum IDE after the paper posting smart contract is triggered by a transaction that posts our paper to the Ethereum testnet.

title, key words, etc.) into a paper metadata record; 3) pack the metadata of the paper to a transaction; 4) issue the transaction to the blockchain system. After the smart contract receives the transaction, it can be executed by some miner to write the metadata of the paper to the blockchain. Fig. 4 shows the window of Remix Ethereum IDE after the paper posting smart contract is triggered by a transaction that posts our paper to the deployed Ethereum testnet. The procedures and smart contracts for other functions are designed and implemented in similar ways.

Currently, we have not implemented the proposed incentive mechanism that requires extensive modifications on the blockchain program codes. This is the most important part of our follow-up work.

## VI. Conclusion

To overcome the drawbacks and limitations of existing publication platforms for research papers, we exploit recent advances in decentralized technologies (i.e., blockchain, IPFS) to design a decentralized open-access publication platform named PubChain. Compared with the existing centralized publication platforms, PubChain has several advantages: (i) PubChain breaks the pay wall imposed by publishers so that everybody can enjoy free access to papers. (ii) PubChain eliminates undesired effects of information islands and has the potential to become a unified database for global sharing and recording of papers. (iii) PubChain, as a decentralized system, provides uninterrupted service without single points of failure. (iv) PubChain incentivizes participants to make positive contributions to the platform with an incentive scheme implemented over blockchain technology.

Importantly, unlike many other publication platforms, PubChain is not meant to be a profit-oriented platform. The donation of cryptocurrency injects initial financial values to Pubchain. We propose to use a two-way pegging technique to lock donated cryptocurrency to a special address of the parent chain that cannot be spent by any individual address. The project development team, as volunteers, will not receive any cryptocurrency

This project will be successful only if it can recruit the participation of authors, reviewers, and readers who believe in the tenet of free dissemination and free open access to timely research results. We invite more volunteers to join the project and work with us to improve the design of Pubchain, and to serve as advocates for the new way of knowledge dissemination for the benefit of humanity.

## References


[1] B. Keith, K. Vitasek, K. Manrodt, and J. Kling, *Strategic sourcing in the new economy: harnessing the potential of sourcing business models for modern procurement*. Springer, 2015.

[2] I. Newton, "Letter from Sir Isaac Newton to Robert Hooke," *Historical Society of Pennsylvania*, Retrieved 7 June 2018.

[3] R. Smith, "Peer review: a flawed process at the heart of science and journals," *Journal of the royal society of medicine*, vol. 99, no. 4, pp. 178–182, 2006.

[4] J. R. Adler Jr, "A new age of peer reviewed scientific journals," *Surgical neurology international*, vol. 3, 2012.

[5] T. Vence, "One million preprints and counting: A conversation with arxiv founder Paul Ginsparg," *The Scientists*, 2014.

[6] D. Puthal, N. Malik, S. P. Mohanty, E. Kougianos, and G. Das, "Everything you wanted to know about the blockchain: Its promise, components, processes, and problems," *IEEE Consumer Electronics Magazine*, vol. 7, no. 4, pp. 6–14, 2018.

[7] K. Christidis and M. Devetsikiotis, "Blockchains and smart contracts for the internet of things," *IEEE Access*, vol. 4, pp. 2292–2303, 2016.

[8] W. Wang, D. T. Hoang, P. Hu, Z. Xiong, D. Niyato, P. Wang, Y. Wen, and D. I. Kim, "A survey on consensus mechanisms and mining strategy management in blockchain networks," *IEEE Access*, vol. 7, pp. 22 328 – 22 370, 2019.

[9] M. Spearpoint, "A proposed currency system for academic peer review payments using the blockchain technology," *Publications*, vol. 5, no. 3, p. 19, 2017.

[10] K. Janowicz, B. Regalia, P. Hitzler, G. Mai, S. Delbecque, M. Fröhlich, P. Martinent, and T. Lazarus, "On the prospects of blockchain and distributed ledger technologies for open science and academic publishing," *Semantic web*, no. Preprint, pp. 1–11, 2018.

[11] M. Sharples and J. Domingue, "The blockchain and kudos: A distributed system for educational record, reputation and reward," in *European Conference on Technology Enhanced Learning*. Springer, 2016, pp. 490–496.

[12] N. Nizamuddin, H. R. Hasan, and K. Salah, "Ipfs-blockchain-based authenticity of online publications," in *Blockchain - ICBC 2018 - First International Conference, Held as Part of the Services Conference Federation, SCF 2018, Seattle, WA, USA, June 25-30, 2018, Proceedings*, 2018, pp. 199–212.

[13] A. Tenorio-Fornés, V. Jacynycz, D. Llop-Vila, A. Sánchez-Ruiz, and S. Hassan, "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.

[14] T. Wang, S. C. Liew, and S. Zhang, "Pubchain: A decentralized open-access publication platform with participants incentivized by blockchain technology," *arXiv preprint arXiv:1910.00580*, 2019.

[15] J. Benet, "IPFS-content addressed, versioned, p2p file system," *arXiv preprint arXiv:1407.3561*, 2014.

[16] S. Wang, Y. Zhang, and Y. Zhang, "A blockchain-based framework for data sharing with fine-grained access control in decentralized storage systems," *IEEE Access*, vol. 6, pp. 38 437–38 450, 2018.

[17] Y. Chen, H. Li, K. Li, and J. Zhang, "An improved p2p file system scheme based on IPFS and blockchain," in *Proc. 2017 IEEE International Conference on Big Data (Big Data)*. IEEE, 2017, pp. 2652–2657.

[18] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.

[19] V. Buterin *et al.*, "Ethereum: A next-generation smart contract and decentralized application platform," 2013.

[20] P. Technologies, "Parity: Fast, light, robust ethereum implementation," retrieved 12 Dec. 2017.

[21] A. Back, M. Corallo, L. Dashjr, M. Friedenbach, G. Maxwell, A. Miller, A. Poelstra, J. Timón, and P. Wuille, "Enabling blockchain innovations with pegged sidechains," 2014.

[22] J. Benet and N. Greco, "Filecoin: A decentralized storage network," *Protoc. Labs*, 2018.


Authorized licensed use limited to: San Francisco State Univ. Downloaded on June 16,2021 at 00:43:48 UTC from IEEE Xplore. Restrictions apply.