Distributed ledgers application in Science: Smart Papers on the Ethereum

YixuanXu (yx5u17@ecs.soton.ac.uk)

University of Southampton

August 25, 2018

Contents

List of Figures

List of Tables

Abstract

This work is all about . . .

1 Introduction

1.1 Project aim

Digitization and Web technologies are now changing the way of publishing and disseminating the knowledge. It becomes more convenient and less expensive for people to access the knowledge. The knowledge creation process is more dynamic right now. Text/graphics/rich media can be changed quickly and easily while at the same time being available to all the audiences. However, most of current methods of academic publication are static, that means, they cannot be revised over time[1]. Web technologies actually have the power to make it more dynamic but is currently underused. On the other hands, journals, publishers and funders fully control the entire process of academic publishing. The view of authors who should also participate in the publishing process tends to be underrepresented. Despite that fact that the current academic publishing system is advance and productive, authors still want a more open and decentralize publishing process [2].

In the past, the scholarly books were keeping improving and updating for the centuries by releasing the new editions. Mistakes would be corrected, new result would be added and feedbacks would be used for improvement. Revising books allowed author to keep track with novel development [1]. Many handbooks and schoolbooks have been revised over and over again, resulting massive amount of quality publications. In the contrast to books, academic paper were a kind of snapshot of certain scientific knowledge. Most of them were just published once. If there is a new finding, usually a new articles need to be published. But this kind of process are currently under debate and development [1]. The number of authors who want a more open process in scientific publishing is increasing rapidly. When it comes to the traditional of academic publishing, publishers play an important role of filtering good research, rejecting papers without sufficient conclusions. They make their decisions based on the peer-review process which is fully controlled by themselves. Since this kind of peer-review process usually will take a significant amount of time and is one of the main reason of delaying publications, researcher came up with a new idea of doing publishing. A initial version will be firstly released, then it could be updated after receiving the feedbacks from pre-peer-review. All the version will be always available and the changes made in the pre-peer-review process will also be stored after the

publishing of final version. The models allows the tracking of the development of academic papers. This improvements make the process of publishing process more dynamic and flexible. But, there are some vital problems that have been discussed under such models which is the mechanism to manage the interactions between authors and contributor in a trust way [3]. How can authors make agreements with each other about which version should be available ? How can authors determine their contributions to the papers in a unprejudiced way ? On the other hand, the contributions of reviewers is ignored in this model.

In last few years, Distributed Ledger Technology have attracted public attentions as the most advanced tool that can provide a decentralized solutions to manage the interaction between people that may not trust each other. It also could guarantees of security and consistence without the need for admin. The special tools which could achieve such functionality is called Smart Contract. For the question about the publishing model, Smart contract could programed to help authors to making decision in a decentralized way. The aim of this project is trying to provide a prototype of decentralize application to help authors to manage their publish and their attribution agreements in a dynamic and trusted way. The application itself will use the blockchain technology [4], so nobody can fully control the whole process. It could be more reliable than the current publishing system. The implementation will be evaluated by the cost of using such system. A detailed cost analysis and data visualization will also presented.

1.2   Outline

In Section 2, firstly it will present some modern publishing systems based on the blockchain. Then a general background of blockchain will be presented. After that, a more specific description about the smart contract on Ethereum would be shown for the following part

In Section 3, it would give detailed design of the application. By comparing the blockchain technology with the current Web technology, It should gives more clear explanation why blockchain is preferred for this project

In Section 4, it will focus on the Implementation. All the important implementation details will be described here

Section 5 demonstrate the cost of using the decentralize application. The analysis consists of the visualization and specific code review.

Section 6 will discuss the advantage and disadvantage of the current implementation. Potential improvement would also be covered in this section.

Section 7 will give a summary about the project and conclusion.

2 Background Research

2.1 Modern publishing models example

Because of the development of web technologies, several models have been provided based on the modern web technologies to improve the process of the academic publish, making authors manage and produce their works more easier. DEIP [5] is a platform that aimed at effective and fair distribution of resource allocated to scientific and research activities. It proposed a community-driven models that encourage the open knowledge without restriction. The users of this platform could be both author and reviewers and have free access to all the publication. All the record of review will be recorded and reviewers will be awarded with the decentralized protocol. Scienceroot [6] is another blockchain-based publishing platform, it tries to creating a scientific publishing model which will reward and sustain researchers instead of maximizing the publishers' profit by using a fully decentralized storage platform called IPFS [7]. The Pluto [8] is also a platform that is trying to help researchers to get funding in a decentralized ways. It also wants to establish a more proper evaluation index for the academic publications. The ideas of these kinds of platforms are really aggressive but inspiring. They are trying to give the power back to authors and reviewers and to make the knowledge more open and accessible to everyone. It could be found that modern publishing system requires more transparent review process and more trustable collaborations between authors. The current Blockchain technologis have huge potential in aspect of creating decentralized application. So it

could be found that all of above platform are using the Ethereum [4] as their core framework.

2.2 Blockchain

Technology On its most basic level, the blockchain is a new kind of information technology which is the combination of encryption methods with distributed computing. Satoshi Nakamoto [9] combine them to make new ways to create a model where a network of computers collaborate to maintain a shared and secure database. This database consists of a string of blocks each one a record of data that has been encrypted and given a unique identifier called a hash. Mining computers on the network validate transactions, add them to the block they are trying to build, and then broadcast the completed block to other nodes so that all have a copy of the database. Because there is no centralized server to validate the transaction to the database, the blockchain depends upon a distributed consensus algorithm. In order to make an entry on to the blockchain database, all the computers have to agree about its state so that no one computer can make an alteration without the consensus of all others. Once completed, a block goes into the blockchain as a permanent record. Each time a block get completed, a new one will be created. Countless number of block connect to each other, like links in a chain. The kind of data structure makes the transactions immutable. Every block contains hash value that is dependent on the hash of the previous block, so they are all linked meaning if one is changed than all the other blocks linked to it going forward will be altered. This works to make the data entered tamper-proof. This model is the workings of the first generation blockchain which is known as Bitcoin. Its functionalities are mainly focused on the small data storage, which is the history of transactions.

Features Blockchain are trying to create a secured, trusted, shared database and they do this through encryption and hashing, proof of work and network consensus. The hashing and linking of blocks makes it difficult to go back and change a previous block once entered. And the proof of work system intentionally makes it computationally more difficult to alter the database, thus making it extremely difficult to change all the blocks. It leads to a distributed consensus mechanism so that even if someone did it, it is almost impossible to convince others to accept as the valid record. The bitcoin blockchain is a very good proof of the stability of such system. It now secure

hundreds of billions of dollars using this method without the network having been hacked. What this technology enables is a database that is secure with automatic trust which is powered by open source community and encryption. It is tamper proof, once information is put into the database it can not be altered. It is a shared database as many people would have a up-to-date copy so that all have a single source of truth. Likewise, it is transparent, all the transactions and alteration made to the blockchain is visible to everyone.

Evolution   With in a few years,the second generation of blockchain emerged design as a network on which developers could build applications. It is going to be more like a distributed virtual computer than just database. This was made technically possible by the development of the Ethereum platform. Ethereum is an open-source, public, blockchain-based distributed computing platform featuring smart contract functionality. It provides a decentralize Turing-complete virtual machine, which can execute computer programs using a global network of nodes. Ethereum was initial described in a white paper by Vitalik Buterin [4]. The system has been very successful attracting a large and dedicated community of developers, supports and enterprises. The important contribution of Ethereum as the second generation of blockchain is that it worked to extend the capacity of the technology from primary being a database to becoming more a general platform for running decentralized application and smart contracts. As of 2018 Ethereum is largest and most popular platform for building distributed application. It has been a major step forward to become a global distributed computer, a massive globally distributed cloud computing platform upon which we can run any application at scale and speed, with the assurance that it has the security, resilience and trustworthiness of today's blockchain.

Smart Paper model   By reviewing the blockchain technology, it is easy to find out that Ethereum is really powerful to help authors to make agreements and set their outcomes in stone so that they cannot be later repudiated. In [3], the Smart Paper model is proposed to provide a collaborative platform that preserves a single version of the truth through the collaborative process. Author will use the smart contract to publish their works and collaborate with others. Once decision is made, smart contract will guarantee the result of decision won't be change by anyone. All the versions and meta data about contributions will be safely stored on the Ethereum.

3 Design

3.1 Requirements

The first step of software developing is to identify the requirements from the specification, then finding the correspond user cases. The aim of this project is to provide a prototype of decentralize application to help authors to manage their publish and their attribution agreements. Based on this, we could have these simple user cases:

- Researchers should have their identities based on addresses of their Ethereum accounts

- Researchers should be able to use smart contract to publish a research artifacts, which will be represented by the hash of files

- Researchers should be able approve the contracts that belongs to themselves after verifying their local files with the online files with respect to a specific version

- The paper contracts should contains the description, metadata and version list of the research artifacts.

- New researchers could be added to an ongoing paper with the agreements of all the previous researchers.

- The application should be open sourced and decentralized. Anyone could deploy their own instances.

The paper smart contract will be design based these requirements.

4 Implementation

5 Evaluation

6 Discussion

7 Conclusion

References

[1] L. Heller, S. Bartling *et al.*, "Dynamic publication formats and collaborative authoring," in *Opening Science.* Springer, 2014, pp. 191–211.

[2] B. D'Souza, S. Kulkarni, and C. Cerejo, "Authors' perspectives on academic publishing: initial observations from a large-scale global survey," *Science Editing*, vol. 5, no. 1, pp. 39–43, 2018.

[3] M. R. Hoffman, I. Luis-Daniel, F. Huw, and S. Elena, "Smart papers-dynamic publications on the blockchain," 2018.

[4] V. Buterin *et al.*, "Ethereum white paper," *GitHub repository*, 2013.

[5] D. Platform, "Decentralized research platform," 2017.

[6] G. Vlad and A. Chirita, ""scienceroot" whitepaper," 2017.

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

[8] P. Network, "Pluto white paper," 2018.

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