Content Sharing Network based on IPFS and Blockchain

Haimei Xu¹, Yan Cheng¹, Zhengdong Li² and Chunyan You³

¹ ChongQing teleCommunication polytechnic college, Chongqing, China
² Chongqing Vocational College of Public Transportation, Chongqing, China
³ Chongqing Municipal Education Commission, Chongqing, China

E-mail:xuhaimei2019@hotmail.com

Abstract. In this paper, we propose one content-sharing system InnerLight based on IPFS and BlockChain, which is a creation and public discussion platform about mental health. InnerLight put copies of articles from creators on IPFS to achieve distributed storage of contents and complete the first step of returning the data to creators. At the same time, it also encourages creators and readers to maintain the sustainable development of the system through blockchain-based cryptocurrency. In addition to IPFS and Blockchain, ranking algorithms contribute to make Innerlight to be a decentralized autonomous ecosystem.

1. Introduction

In the information society, information resource is not only a kind of resource, but also a kind of social wealth. Information resource should be shared by the public in some way under an efficient organization and allocation mechanism, and be used fairly, orderly and fully. Information resource sharing is an important goal of information resource construction and integration. With the in-depth development of social informatization and the rapid advancement of IT technology, new concepts and new business forms, such as the InterPlanetary File System (IPFS) [1], Blockchain, Cloud Computing, Big Data and Artificial Intelligence (AI), have been continuously influencing the mode and means of information resource sharing in recent years.

Blockchain as a new information infrastructure, to which the banking, securities, insurance, logistics, notarization, electronic commerce, Internet of things and other Internet industries have attached great importance, is a new force of the Internet which will innovate information resource sharing model. Along with other major cutting-edge information technologies, blockchain has been included in China National Informatization Plan [2]. It can be predicted that with the continuous improvement of technology and the increasing maturity of applications, blockchain will exert a profound impact on the construction and the sharing of information resources under the driving force of the government’s vigorous promotion and huge social demand.

Blockchain technology comes from Bitcoin. Bitcoin and blockchain were first proposed by Satoshi Nakamoto [3]. Bitcoin is a brand new digital currency in an innovative payment system while blockchain is the underlying technology and infrastructure supporting bitcoin. A blockchain is a list of several blocks linked by hash pointers, as shown in Figure 1. With different versions and types of blockchain, multiple applications can be built in the application layer, and information resource sharing system is one of them [4-7].
Figure 1. The basic structure of blockchain.

IPFS is a Peer-to-peer protocol to make the web faster, safer, and more open [8], and it is also a distributed file system that seeks to connect all computing devices with the same system of files, which provides a high through-put content-addressed block storage model with content-addressed hyperlinks. IPFS combines important properties of successful peer-to-peer systems. Firstly, IPFS uses Distributed Hash Table (DHT) to coordinate and maintain metadata of P2P systems. Secondly, IPFS design includes BitTorrent’s block exchange protocols which succeed in coordinating networks of untrusting peers to cooperate in distributing pieces of files to each other. Thirdly, because Version Control Systems provide facilities to model files changing over time and distribute different versions efficiently, IPFS introduce the popular version control system Git which can provides a powerful Merkle DAG object model to captures changes to a filesystem tree in a distributed-friendly way. Fourthly, IPFS introduced a technique SFS for building Self-Certified Filesystems. SFS [9-10] proposed compelling implementations of both distributed trust chains and egalitarian shared global namespaces, which synthesizes successful techniques from previous peer-to-peer systems, including DHT, BitTorrent, Git, and SFS, to present a new system for distributing and versioning large data.

As a new force of the internet, IPFS has brought great challenges to the information resource sharing and discovery as shown in Figure 2 compared with traditional HTTP mode. This paper presents one content-sharing system InnerLight based on IPFS and BlockChain. InnerLight is a creation and public discussion platform about mental health, which is based on distributed network and driven by cryptocurrency. By InnerLight, people interested with psychology and mental health can form a community where everyone can create, share contents and discuss some specific topics. InnerLight put copies of articles on IPFS to achieve distributed storage of contents and complete the first step of returning the data to creators. At the same time, it also encourages creators and readers to maintain the bright future of community development through blockchain-based cryptocurrency.

Figure 2. HTTP vs IPFS.

2. InnerLight based on IPFS and Blockchain

The premise of information resource sharing is information aggregation, that is, information is stored and organized by some means. In the centralized mode, all information resources are allocated to the information resource centre, while in the non-centralized mode all information resources are distributed to each node, which will not accumulate to form the information resource centre. IPFS runs on Peer-to-Peer networks and follows a decentralized pattern, and therefore there is no centre in information resource sharing system.

2.1. The Content-sharing Chain based on IPFS
In this chain, information is stored directly in IPFS nodes, these nodes are also used for the query or search of information resources in the Peer-to-Peer network. InnerLight select IPFS for the distributed storage of contents. Every piece of content published in InnerLight, including text, pictures, and links, will be uploaded to IPFS nodes. Once the article is uploaded to IPFS nodes, permanent storage is achieved. Each article has several copies, each copy has a public access address, which we call a public node as shown in Figure 3.

For example, the operation process of IPFS below presents how to upload and download the content C1 to IPFS nodes as shown in Figure 4 and Figure 5 respectively.

Assuming that one author wants to post an article C1 to InnerLight. InnerLight will be responsible for adding the C1 article to the IPFS system. The IPFS system will hash C1 article and give a hash value starting with Qm. A feature of IPFS is that it replaces traditional Internet domain-based addresses with content-based addresses (hash values beginning with Qm). The IPFS system then copies the C1 article into multiple copies and splits each one, and each part of the split is stored on the decentralized IPFS network nodes. Because multiple copies can perform redundant backups, even if some nodes are offline, C1 article can be founded from other nodes, and thus security can be ensured. However, it should be noted that IPFS cannot ensure that it is 100% to avoid partial file loss.

| Hash Value for one pieces of content C1 (Unique ID) |
|--------------------------------------------------|
| QmVvuTYncwWQX6ZC2AZ6hr5QKznYxByQGcWroNfwHqaYA8 |

Public Address 1:
https://13.230.162.124/ipfs/QmVvuTYncwWQX6ZC2AZ6hr5QKznYxByQGcWroNfwHqaYA8/

Public Address 2:
https://13.114.30.87/ipfs/QmVvuTYncwWQX6ZC2AZ6hr5QKznYxByQGcWroNfwHqaYA8/

Public Address 3:
https://13.114.247.2/ipfs/QmVvuTYncwWQX6ZC2AZ6hr5QKznYxByQGcWroNfwHqaYA8/

Public Address 4:
https://ipfs.io/ipfs//QmVvuTYncwWQX6ZC2AZ6hr5QKznYxByQGcWroNfwHqaYA8/

**Figure 3.** Four public addresses for one piece of content C1 with unique ID.

If anyone wants to get C1 article, according to the content address (the hash value at the beginning of Qm), he can download C1 article from the IPFS system.

IPFS system downloads different parts of C1 article ranged from public addess1 to public address 4 and finally joins all the parts to restore the original file by BT’s transport technology. Compared to the traditional HTTP protocol, users have to upload and download data from the centralized server, the IPFS transmission speed is faster and the more users at the same time, the faster the transmission speed.

| C1 article | QmVvuTYncwWQX6ZC2AZ6hr5QKznYxByQGcWroNfwHqaYA8 |
|------------|-----------------------------------------------|
| IPFS nodes |

**Figure 4.** Upload c1 article to four public addresses (yellow nodes).
2.2. The Incentive Chain for Content Sharing Network

It is well known for people who used Steemit [11] that on Steemit, the top articles are often those with the most gimmicks and the most attractive clickbait, but the highest quality contents may not necessarily float out of the water. Some social platforms like today’s headline [12] use algorithms that capture user preferences to determine content recommendations while for some traditional Medias, contents are sorted out by internal teams. Lack of effective incentives in the content sharing system will hinder the sustainable development of a community and lead to tragic of the commons [13]. InnerLight use three mechanisms to incentive all the participants of the system to build a benign development of quality community.

Firstly, Micropayment based on cryptocurrency is introduced to incentive authors and readers to contribute to community. Creators and readers may use cryptocurrency as a convenient tool to interact with the decentralized ecosystem—InnerLight. Authors can earn virtual coins by creating quality contents, and readers can also gain rewards by reasonable comments on quality contents while they can express their admiration by giving virtual coins to their favourite articles. For example, LikeCoin[14-16] is a public blockchain built on the Cosmos SDK for content monetization, attribution and distribution. Empowered by LikeCoin, every content is a registered asset on LikeCoin chain and every Like is a reward to creators in LikeCoin token. LikeCoin can be used on WordPress, Medium, Vocus, blogspot and any other content platforms. Creators and readers may use Likercoin as a convenient reader and gateway to interact with the decentralized ecosystem. LikeCoin token, we propose to build a utility token lightcoin based on Ethereum and ERC-20 compliant for realigning creativity and reward. Lightcoin can reward content creators by proof of creativity mechanism. While blockchain miners usually regard mining as solving a mathematical problem, it does not have to be the case. With a focus on rewarding content creators, Lightcoin can measures the importance of creative contents, i.e. creativity, and allows content creators to mine Lightcoin with their creative contents. The more upvotes the content receives, the more creative it is. A creator’s pool of LightCoin will be minted for distribution to content creators.

Secondly, in order to prevent colluding attack from over 50% participants of the system, Innerlight introduces initial Trusted User as Seeds Users with high initial trust value, who can guide the community to develop better and better.

Thirdly, a sorting algorithm represents a kind of arrangement and organization of content. The presentation quality of contents is ultimately selected by algorithms. The right to choose an algorithm is returned to all users. Users bind their own interests with the quality of information. People will vote for the best content. The better the community develops, the more they will benefit. People who contribute more to the community will be given higher weights, while articles awarded by people with higher weights will be ranked higher. Innerlight has a set of its own algorithms, which can convert users’ upvotes to the content to the quality level of this content, so as to make the more valuable content in the community emerge, and guide the entire community to develop in a better direction.

3. Ranking Algorithms

In order to ensure that the high quality content in the platform is selected for readers, we adopt a sorting mechanism.
So far, there are three main modes of organizing content.

The first is the distribution mode based on social relations represented by WeChat circle of friends and Facebook. In the WeChat friend circle, everyone tends to post more content. As a result, the friend circle becomes more and more bloated, filled with all kinds of advertisements and other messy content, which leads to public tragedy.

The second is Toutiao algorithm-based distribution model. In essence, machine algorithms through large data analysis, get the common feature extraction of a group of people, for example, the common features may be "20 years old girl like romance", and these characteristics are abstract into some labels, which can be used to search for more similar contents that are finally delivered to the corresponding group. In pursuit of profits, Businessmen use machine algorithms to exploit human weaknesses for the efficient delivery of content. As a result, such algorithms have created a content cage for each of us.

The third way is represented by Douban group. This model classifies the content based on people with the same interests. This classification method relies on more granular interest tags, which allow people to aggregate through subdivided topics, share relevant content within the topic, and get everyone involved in managing the content. Compared with the previous two models, this model effectively reduces the redundancy of information, but it does not fundamentally solve the problem of content quality.

We propose to return the organization of content to users. If all the content is controlled by a group of moderators, who can decide what is quality and what is bad, or, all content controlled by the company behind a platform, such as Toutiao that tends to use a fixed set of algorithms to determine what information you want to see, and what information you don't want to see, the quality problem of the content will never be improved. Coinbase co-founder Fred Ehrsam came up with the idea that the quality of the presentation of the content is ultimately selected by the algorithm. However, the right to select the algorithm must be returned to the user. Users binding their own interests to the quality of the content will choose the best algorithm for their own interests. So it's a good idea to create a self-healing, less cold algorithm that is truly beneficial to all users.

We propose to use a sorting mechanism to present quality content to users. Sorting represents the filtering, organization and maintenance of all the content by all users. We propose two algorithms to present quality contents for Innerlight as follows.

3.1. The hottest articles ranking algorithm

Compared with Literature [20], we take into account the number of comments in addition to the number of upvotes (likes).

Given the time the content of C1 posted A and the time of 8:08:08 a.m. December 8, 2020 B as the running time of InnerLight, we have t as their difference in seconds

\[ t = A - B \]  \hspace{1cm} (1)

And x as the number of upvotes, c as the number of comments

We have the rating as a function

\[ f(t, x, c) = \log_2(x + c) + \frac{t}{172800} \]  \hspace{1cm} (2)

172800 seconds equal 24hours,

The above algorithm shows two main factors influence the ranking. Firstly, submission time is a very important parameter, generally newer contents will rank higher than older. Secondly, the logarithm function is applied to weight the first votes higher than the rest.

3.2. The Latest articles ranking Algorithm

\[ f(x,t) = \frac{(x + c - 2)}{(t + 2)^g} \]  \hspace{1cm} (3)
Where

\( x \) = the number of upvotes
\( t \) = time since submission (in hours)
\( g \) = Gravity, defaults to 1.5
\( c \) = the number of comments

Effects of gravity \((g)\) and time \((t)\)

Gravity \((g)\) and time \((t)\) have a significant impact on the rating of a piece of content. Generally these things hold true:

- the rating decreases as time \((t)\) increases, meaning that older contents will get lower and lower scores and the rating decreases much faster for older items if gravity is increased, after 24 hours (i.e. \( t = 24 \) hours) no matter how many upvotes one article gets, it still gets a very low score, which means this algorithm can make old articles disappear faster, one article that are more than 24 hours old, will not remain on the list of the latest articles with a high probability.

Compared with Literature [21], we rank the newest popular contents with upvotes and comments. And we have the rating as a function.

3.3. Comparison

The latest articles ranking algorithm can make old articles disappear faster, like articles from 24 hours ago, it's hard to stay on the list. In contrast, the hottest articles ranking algorithm allows the best posts to stick around for a long time and then get pushed down by new, better posts over time. Another advantage of the hottest articles ranking algorithm is that when the number of votes or comments changes, the rank of the article needs to be recalculated, regardless of the time change.

4. Conclusion

InnerLight is a decentralized community in cyberspace among content creators and readers, who collectively own and govern the community. It is also a decentralized autonomous ecosystem. Cryptocurrency can provide creators with tangible rewards, which make appreciation and rewarding for creativity a core value in an ecosystem. InnerLight uses two chains, one is IPFS content chain and the other is a blockchain, which is a cryptocurrency chain. Each article is posted to IPFS nodes and cannot be deleted or modified at will. The ownership of contents belongs to their authors. The fingerprints of contents and their addresses of IPFS nodes are publicly visible and can be used to share the corresponding contents. If an author needs to modify the content, he can hide it from the website, and republish the revised content. The hidden content is only visible to its author. Cryptocurrency based on blockchain is introduced to reward creativity of contents' creators.

InnerLight is literally a decentralized autonomous organization (DAO) among content creators and consumers, who collectively own and govern the community.

5. Acknowledgments

The authors would like to acknowledge the supports provided by the Program of Science and technology from ChongQing Municipal Education Commission (KJZD-K201805501), China Scholarship Council, the Key Project of “13th five” Chongqing Education Science in 2019 (2019-GX-185), the Research Project on Teaching Reform of Chongqing Education Commission in 2019 (193561) and the Research Project on Science and Technology of Chongqing Education Commission in 2019 (KJQN201905803).

6. References

[1] IPFS, The Permanent Web. Retrieved May 18, 2017 from https://github.com/ipfs/ipfs
[2] Notification of the national informatization plan. http://www.gov.cn/zhengce/content/201612/27/content_5153411.htm
[3] Satoshi Nakamoto, Bitcoin: A Peer-to-Peer Electronic Cash System. https://bitcoin.org/bitcoin.pdf
[4] Melanie Swan 2015 *Blockchain: Blueprint for a New Economy* (Sebastopol, CA: O’Reilly Media Inc)

[5] Wang Yong. Blockchain: unprecedented technological breakthrough and legal challenge *Shanghai Securities News* 2016-06-22(08)

[6] Bitcoin Developer Guide[EB/OL]. https://bitcoin.org/en/developer-guide.[2018-01-01].

[7] Andreas M Antonopoulos 2017 *Mastering Bitcoin* ISBN-10:1491954388

[8] IPFS powers the Distributed Web , https://github.com/ipfs/ipfs#quick-summary [2020-04-30] 

[9] D. Mazieres and F. Kaashoek, Self-certifying file system, 2000, Massachusetts Institute of Technology201 Vassar Street, W59-200 Cambridge, MA United States

[10] D. Mazieres and M. F. Kaashoek 1998 Escaping the evils of centralized control with self certifying pathnames. In Proceedings of the 8th ACM SIGOPS European workshop on Support for composing distributed applications 118-25 ACM

[11] Chohan, Usman W., The Concept and Criticisms of Steemit (March 29, 2020). CBRI Working Papers: Notes on the 21st Century, Available at SSRN: https://ssrn.com/abstract=3129410 or http://dx.doi.org/10.2139/ssrn.3129410

[12] Ximing Li and Yi Zhang 2019 Empirical Research on Usage Intention and Behavior of "Toutiao" Users, Proceedings of the 4th International Conference on Economy, Judicature, Administration and Humanitarian Projects (JAHP 2019) 94 885-90

[13] G. Hardin 1968 “The tragedy of the commons,”Science 162(3859) 1243-48

[14] LikeCoin Foundation Limited, LikeCoin is a public blockchain for content monetization, attribution and distribution. [online] Available:https://github.com/likecoin, 2020-09-25

[15] LikeCoin Foundation Limited LikeCoin: Reinventing the Like, Rewarding Contents by Proof of Creativity, White Paper v0.9.3, [online] Available: https://like.co/in/likecoin-whitepaper-en.pdf, 2020-09-25

[16] Trishita Tiwari, David Starobinski and Ari Trachtenberg 2018 Distributed Web Mining of Ethereum *International Symposium on Cyber Security Cryptography and Machine Learning* 38-54

[17] Chunmei Gan 2017 Understanding WeChat users' liking behavior: An empirical study in China *Computers in Human Behavior* 68(C) 30-9

[18] Yang, J. and Yecies B. Mining Chinese social media UGC: a big-data framework for analyzing Douban movie reviews. Journal of Big Data 3, 3 (2016). https://doi.org/10.1186/s40537-015-0037-9

[19] Fred Ehrsam,Token curation markets use economic incentives to create crowd curated information feeds, https://twitter.com/FEhrsam/status/958388803655184386, 2020-09-25

[20] Amir Salihefendic, How Reddit ranking algorithms work, https://medium.com/hacking-and-gonzo/how-reddit-ranking-algorithms-work-ef111e33d0d9, 2020-09-25

[21] Amir Salihefendic, How Hacker News ranking algorithm works, https://medium.com/hacking-and-gonzo/how-hacker-news-ranking-algorithm-works-1d9b0cf2e08d, 2020-09-25