

Incentive Mechanisms for a Decentralized Blockchain-based Search Engine for the Decentralized Web

Michael Chau
Faculty of Business and Economics
The University of Hong Kong
Hong Kong
mchau@business.hku.hk

Eric Lo
Department of Computer Science & Engineering
The Chinese University of Hong Kong
Hong Kong
ericlo@cse.cuhk.edu.hk

Ben Kao
Department of Computer Science
The University of Hong Kong
Hong Kong
kao@cs.hku.hk

Siu-Ming Yiu
Department of Computer Science
The University of Hong Kong
Hong Kong
smyiu@cs.hku.hk

Abstract—Decentralized Web, or DWeb, has been proposed as a promising model for the future of the Web. An important infrastructure to support the functioning of DWeb is a decentralized search engine in order to allow users to look for the information they need. There are various parties involved in the operation of a decentralized search engine, such as website owners, crawlers, indexers, and advertisers. This research focuses on the incentive mechanism behind a decentralized blockchain-based search engine. Three research questions on this area are proposed and our research plan is discussed in this paper.

Keywords—Decentralized Web, decentralized search engine, incentive mechanism, business model

I. INTRODUCTION

Decentralized Web, or DWeb, has been proposed as a promising model for the future of the Web [1]. DWeb is a model in which there is no centralized web server. Web contents are stored on decentralized web servers. The decentralized nature of DWeb provides several advantages over the traditional web, such as data redundancy, protection against denial-of-service attacks, and higher throughput. A prototype of the DWeb has been implemented and can be found on the Internet¹.

An important infrastructure to support the functioning of DWeb is a decentralized search engine in order to allow users to look for the information they need. A decentralized search engine should support indexing, ranking, and retrieval of DWeb contents. Such a decentralized search engine has several advantages over traditional search engines as it does not suffer from search bias and censorship [2, 3].

Many decentralized infrastructures have relied on the use of cryptocurrency, often called tokens or coins, as the incentive mechanism to maintain its operation. The most notable example is Bitcoin. The Bitcoin network rewards workers (usually called miners), who are vital to its

operations by verifying transactions, with a certain number of Bitcoins based on a proof-of-work system [7]. These tokens often have actual value and can be exchanged for other currencies. One issue of this mechanism is that when the monetary value of the cryptocurrency is low, especially when the reward falls below the computing cost (e.g., hardware and electricity cost), workers may lose the motivation to work and may leave the blockchain network.

The focus of our research is to study the incentive mechanism behind a decentralized blockchain-based search engine. There are various parties involved in the operation of a decentralized search engine. First, contents need to be submitted by website owners or crawled by workers, which are machines on the DWeb that constitute the search engine. Second, workers need to perform indexing of the contents collected. On the other hand, advertisers want their advertisements to be seen when users perform search, and therefore they are the ones who may provide incentives to support the operation of the search engine. In this paper, we study the incentive mechanism involved and raise some research questions.

II. RESEARCH BACKGROUND

A. Decentralized Search Engines for the Decentralized Web

A search engine is needed in order to allow users to search for contents in the DWeb. Following the spirit of decentralized and distributed computing, it would be natural to have a decentralized search engine to serve this purpose. A decentralized search engine called QueenBee has been proposed in our previous work [4]. QueenBee aims to revolutionize the search engine business model by offering incentives to both content providers and peers that participate in QueenBee's page indexing and ranking operations. QueenBee's core architecture, shown in Figure 1, is designed based on smart contracts on a cryptocurrency blockchain like Ethereum [5]. QueenBee has the vision of having advertisers directly deliver advertisements through smart contracts and

¹ <https://ipfs.io/ipfs/QmXoyvizjW3WknFiJnKLwHCnL72vedxjQkDDP1mXWo6uco/wiki/index.html>

the ad revenue is shared among the content creators and peers that maintain the search engine.

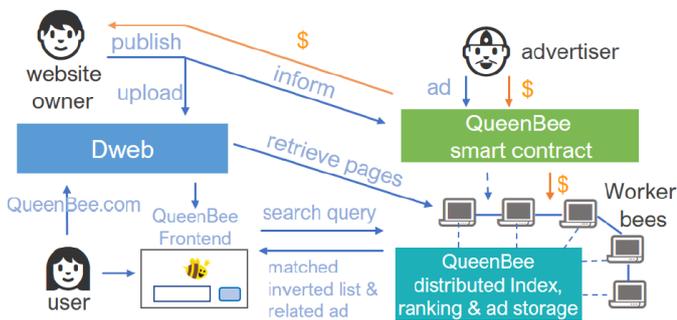


Fig. 1. Architecture of QueenBee [4]

There are various technical challenges in the design and implementation of a decentralized search engine for the DWeb. First, performance would be a critical issue. Internet users nowadays often expect websites and search engines to respond quickly and are used to getting high quality search results from sophisticated search engines like Google.

Second, it is important to know how to divide tasks into smaller pieces to be allocated to workers. It will not be trivial to perform crawling, indexing, and retrieval in a decentralized manner. For example, issues such as duplicated contents, duplicated URLs, and mirror sites need to be considered [6].

Third, security would be another important issue. A blockchain-based decentralized search engine would need to deal with typical blockchain attacks such as fake search results. Besides, the new decentralized search engine model may attract new form of attacks. For instances, some workers may aim at manipulating the search engine's indexes or ranking or the privacy issues. Advanced algorithms would be needed to protect the blockchain from such types of attacks.

B. Incentive Mechanism

Many blockchain applications have relied on the use of cryptocurrency (e.g., Bitcoin's BTC, Ethereum's ETH) as the incentive mechanism to maintain its operation. For Bitcoin and Ethereum, nodes that help to maintain the blockchain get BTC/ETH as rewards. Users have to pay BTC/ETH in order to carry out transactions and operations.

III. RESEARCH QUESTIONS

In a traditional web search engine such as Google, the search engine crawls contents from as many websites as possible, perform indexing and ranking, and process queries and return search results [8]. All these are done at the search engine's cost. Website owners and search engine users in general do not need to pay any fee. The search engine company expects that a good quality search service can attract a large number of users, which will in turn attract advertisers to pay for putting advertisements on its search engine result pages.

A decentralized search engine would need to have a new business model to sustain. The traditional model cannot be applied directly because of the decentralized nature. This paper focuses on the business and economic model of a decentralized search engine and the following are some interesting research questions that we are planning to investigate in this study:

1) To crawl or not to crawl?

Should a decentralized search engine crawl pages from the DWeb? Crawling can actively control the coverage of the search space, but involve a large amount of computing resources. Or, should a decentralized search engine incentivize website owners, in the form of a cryptocurrency, to publish their contents via smart contract? If we incentivize website owners to actively notify the search engine with the new contents as created, the search engine can always have the freshest index, which is a very strong advantage over existing centralized services. This is particularly so given the increasingly dynamic nature of the web. If we incentivize crawlers, there are technical issues like how to assign websites to different crawlers. Either way, when there are rewards, it is necessary to deal with issues like duplicated contents (e.g., to prevent duplicated rewarding to mirror sites and to prevent pirating sites). It may also be interesting to explore the effectiveness of a combination of the two, i.e., incentivizing websites owners who publish and incentivizing crawlers who crawl from other websites. The incentive scheme has to be so designed such that we can cover a similar, if not more comprehensive, search space as existing search engines. Unlike in the traditional centralized model, the cost of crawling not-so-common webpages can be covered by the incomes of advertisements of the popular webpages. In our blockchain model, we need to incentivize crawlers or content providers to handle both popular and not so popular web contents. Otherwise, another forms of bias may occur.

2) How much to pay workers?

A decentralized search engine relies on workers to perform indexing and other processing (e.g., PageRank calculation) of contents. These workers need to be incentivized to keep processing their work. An obvious way is to reward them with cryptocurrency tokens. As seen in the case of Bitcoin and other blockchain applications, workers are often willing to contribute as long as the value of the cryptocurrency payment is larger than the cost of the computing power. However, one problem is that the value of cryptocurrencies often suffers from high volatility. Cryptocurrencies such as Bitcoin and Ether have a much larger fluctuation than fiat currencies. We may need a more dynamic pricing scheme to balance the value fluctuation of cryptocurrencies. It would be very interesting to study how to devise a reward mechanism that would ensure that enough workers will stay in the network in order to keep the search engine running. There are also many issues such as the problem of dealing with not-so-popular web contents as mentioned in the last paragraph and how to share the incentives among different workers who worked on the

processing of the same contents. With the security issue of not allowing any worker to manipulate the indexing process, this becomes a very challenging research problem.

3) *How much do advertisers pay?*

When there is good traffic to the search engine, advertisers would be willing to pay for posting advertisements to be displayed in the search result pages. In traditional search engines, auctions are often run among advertisers where advertisers try to bid for their advertisement being displayed for a particular search keyword, based on the highest price that they are willing to pay per click (or per a thousand views). Game-theoretic and other mathematical models have been used to analyze the decisions of advertisers [9].

In a decentralized search engine, one possible way is to follow the process of a traditional search engine like Google, i.e., keyword auctions can be run among advertisers who bid by cryptocurrency. We may, however, also incorporate other components into the pay. For example, shall we consider the popularity of the website or the cost of indexing the page? Shall part of the advertisement fee be paid directly to the worker that processed the related search index? It would be interesting to study whether this would still be the best strategy to maximize income while maintaining advertisement relevance in a decentralized search environment and ultimately decreasing the advertisement fee.

IV. RESEARCH PLAN

Our current research plan is to mathematically model the various incentive mechanisms of a decentralized search engine as discussed in the previous section. The mechanisms will then be tested in three ways. First, they will be analytically studied and see whether a theoretical optimal solution can be derived from the models. Second, simulation testing will be conducted and the performance of the

different mechanisms will be evaluated. Different values of parameters, such as number of publishers, number of workers, number of advertisers, and budget of advertisers, will be used in our evaluation in order to establish the sensitivity and robustness of our findings. Third, and perhaps most importantly, the proposed mechanisms will be empirically tested in the QueenBee prototype. This will allow us to demonstrate the real-world applicability of the incentive mechanisms.

REFERENCES

- [1] B. Kahle. Locking the web open: A call for a decentralized web. <http://brewster.kahle.org/2015/08/11/locking-the-web-open-a-call-for-a-distributed-web-2>, 2015.
- [2] L. M. Hinman. Searching ethics: The role of search engines in the construction and distribution of knowledge. In *Web search*. Springer, 2008.
- [3] J. Kulshrestha, M. Eslami, J. Messias, M. B. Zafar, S. Ghosh, K. P. Gummadi, and K. Karahalios. Quantifying search bias: Investigating sources of bias for political searches in social media. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 2017.
- [4] Z. Lai, C. Liu, E. Lo, B. Kao, and S.-M. Yiu. Decentralized Search on Decentralized Web. arXiv:1809.00939v1 [cs.IR] 18 Aug 2018.
- [5] G. Wood. Ethereum: A secure decentralised generalised transaction ledger. Ethereum project yellow paper, 2014.
- [6] M. Chau, J. Qin, Y. Zhou, C. Tseng, and H. Chen. SpidersRUs: Creating specialized search engines in multiple languages. *Decision Support Systems*. 2008 Jun 1;45(3):621-40.
- [7] S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system. 2008.
- [8] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*. 1998 Apr 1;30(1-7):107-17.
- [9] P. Maillé, E. Markakis, M. Naldi, G. D. Stamoulis, and B. Tuffin. *Electronic Commerce Research*, 2012, 12:265.