A Framework of Data Sharing System with Decentralized Network

Pengfei Wang[1,2], Wenjuan Cui[1], and Jianhui Li[1](✉)

[1] Computer Network Information Center, Chinese Academy of Sciences,
Beijing 100190, China
lijh@cnic.cn
[2] University of Chinese Academy of Sciences, Beijing 100049, China

Abstract. With the development of technology, more and more data have been accumulated. Utilization of data can benefit many applications, such as facilitate human daily life, promote scientific research, and expedite event detection etc. How to share data within a certain group or a specific person securely and effectively has become a hot research topic. Meanwhile, many distributed technologies have appeared with the triggering of many centralized cloud web services. Thus, we propose a block-chain based data sharing framework with IPFS (https://ipfs.io/), Ethereum (https://www.ethereum.org/) and uPort (https://www.uport.me/). As a result of this design, the system could store data in IPFS system and control sensitive data by block-chain, and the authentication can be managed by the uPort ID system. Finally, we evaluate the effectiveness by a case study.

Keywords: Decentralized system · Blockchain · Data sharing

1 Introduction

Enormous and various data are accumulated by different individuals and organizations [17] from different devices in different application scenarios every day. Meanwhile, there is no one application could exist without any other applications, that is, one application is often related to one or more applications. Similarly, application optimization with utilizing its own dataset often can not get the best effect. Thus, data sharing is critical in data utilizations [3], such as data statistics, data mining and machine learning.

Many centralized systems have been built to provide data sharing service. Specifically, the users could upload their datasets into the centralized server and download the others' datasets from the centralized server. All options are set to rely on the centralized server, including registering, searching, uploading, downloading.

On the decentralized scheme, some people have done some interesting and useful work [2,7,20,21]. Especially, the technologies about blockchain have been developed continuously and well known since the appearance of Bit Coin. There

J. Li et al. (Eds.): BigSDM 2018, LNCS 11473, pp. 255–262, 2019.
https://doi.org/10.1007/978-3-030-28061-1_25

are more and more concerns on these technologies, focusing on governance, security, and privacy, etc.

Traditional data sharing systems, as previously mentioned, could satisfy the requirements when the volume of datasets and number of users are limited. With the explosive growth of data, it is necessary to find a secure, efficient and effective way to build a data sharing system.

Decentralized storage system have been proposed to protect personal data [25]. And some P2P distributed systems have been constructed to share data [18].

However, most of the previous work, including the centralized data sharing systems and the traditional P2P distributed systems, have some disadvantages. For example, the centralized data sharing systems often get the less efficiency compared with the decentralized ones because of all the options are depended on the central server. The traditional P2P data sharing systems often cause disorganized managements and insecure environments. Thus, to address the weaknesses of **Risk Resistance Capacity**, **User Privacy Policy** and **Data Transmission Efficiency** in traditional data sharing systems, we propose a block-chain based data sharing framework with the new decentralized technologies, IPFS and uPort, to provide securely, efficiently and effectively distributed data sharing services.

2 Related Works

This paper is related to the following categories of prior work: data sharing and the application of blockchain in various fields.

Data sharing. Data sharing has been put forward at least from 1980s [10]. There was a discussion about whether the researchers should share their preliminary results in [16]. Data sharing is a complex issue with multiple technical, social, financial and legal facets [11]. And in [11], the authors analyzed the NIH (National Institutes of Health) policy statement and data sharing models to ensure the data sharing is effective and rewarding. Fecher et al. focused on the academic data sharing and wanted to provide evidence for science policies and research data infrastructure [9].

There are two models to share data, central database resources sharing system and peer-to-peer exchange. On central database resources sharing system, every user should access the central server to get the data. More specifically, a person who wants to share data through a centralized system should register on the system to have a user ID, then he/she could upload, download dataset through the central database. [14] focused on discussion on the Archaeological Markup Language (ArchaeoML) and explored approaches to data sharing and data integration which could be used to organize archaeological information. On peer-to-peer exchange situation, the individuals could exchange the local data offline. What's more, we can construct a peer-to-peer network and then share the datasets through the system [15]. For example, in the work [8], the authors proposed future directions for research on P2P systems, and they illustrated some of the trade-offs at the heart of search and security problems in P2P data sharing systems. In [18], the authors designed a peer-to-peer data sharing system. They also implemented and evaluated the PeerDB on a computer cluster. The results showed that PeerDB is efficient. In the work [13], the authors proposed a novel P2P-based MCS(Mobile crowdsensing) architecture, where the sensing data was saved and processed in user devices locally and shared among users in a P2P manner. Then they analyzed the user behavior dynamically and proposed iterative algorithms that were guaranteed to converge to the game equilibrium.

Application with Blockchain. There are many applications in various domains based on blockchain, such as Bitcoin in economic field, voting system [22] in political field and energy supply in resource field, etc.

Decentralized personal data management system was designed in [25]. By using this system, the users could own and control their data. Furthermore, the authors implemented a protocol which made a blockchain work as an automated access-control manager that did not need trust in a third party. In work [19], the authors constructed a framework for cross-domain image sharing based on blockchain. In their framework, the blockchain worked as a distributed data store to establish a ledger of radiological studies and patient-defined access permissions. In the work [1], the authors designed a data sharing framework for electronic health record using ethereum-based blockchain technology.

As mentioned above, the traditional data sharing systems are centralized. The efficiency is limited. The peer-to-peer data sharing system is lack of a kind of effective authorization to manage the datasets. Thus, we propose a blockchain based framework with IPFS and uPort to provide safe and efficient data sharing services.

3 Decentralized Technologies

Distributed technologies have influenced human life in various aspects. For example, from the users' perspective, Twitter[1] is a peer-to-peer alternative social network operating on a decentralized framework, on which users could deliver information and get others' messages; Massive Online Open Courses (MOOCS) is a decentralizing education service system, from which everyone could become accessible to college; And peer-to-peer payment networks have been constructed with online payment framework, such as VenMo and Alipay. Furthermore, the decentralized technologies are more likely to be the Blockchain, Ethereum IPFS and uPort, etc.

Blockchain. A blockchain is a set of records which are called blocks. Given a specific blockchain, each block contains three parts, a cryptographic hash of previous block, a timestamp the block generates and the transaction data. In general, a public blockchain does not have any access restriction. Thus, many algorithms have been used to secure the blockchains, such as Proof of Work (PoW), Proof of Stake (PoS) and the Delegated Proof of Stake (DPoS). For

[1] https://twitter.com/.

example, bitcoin uses a proof-of-work system to create the next block, and EOS[2] uses a proof-of-state to create the next block.

Ethereum. Ethereum [6] is a decentralized platform which can run smart contracts, and smart contracts are the kernel technologies of blockchain. Smart contracts could be authored so that algorithmically specify and autonomously enforce rules of interaction through Ethereum [24]. Many tools could be used to develop smart contracts, such as Truffle[3] and Remix[4]. Many decentralized applications could be programmed without third-party interference.

IPFS. IPFS [4] is a peer-to-peer distributed file system that seeks to connect all computing devices with the same system of files. That is, a terminal could be a client and a server at the same time. IPFS takes advantage of the same peer-to-peer file-sharing capabilities of BitTorrent, and has more expanded functionality [23]. In the framework, we can create an instance for each user, encode the data into JSON format, and put the JSON code into IPFS and generate IPFS hash code.

uPort. uPort is an interoperable identity network for a secure, private, decentralized web. Furthermore, uPort is a decentralized infrastructure for claiming identities and receiving verification from other parties in the network [12]. uPort could help us solve the unique identity for each user [5]. Figure 1 shows the interface of uPort application. First, each user should install the uPort app on mobile devices and register through a Dapp. Then, each user can interact with the ethereum network (such as rinkeby net) by the unique ID generated by uPort.

4 Framework Overview

Figure 2 shows the framework overview of the system. Given a specific user, the general operation steps are as follows.

Register uPort. Firstly, each user log in the uPort system to get an unique user ID in the decentralized network. Every user can manage his/her Dapp and ethers, and the user can broadcast his/her public key into the network.

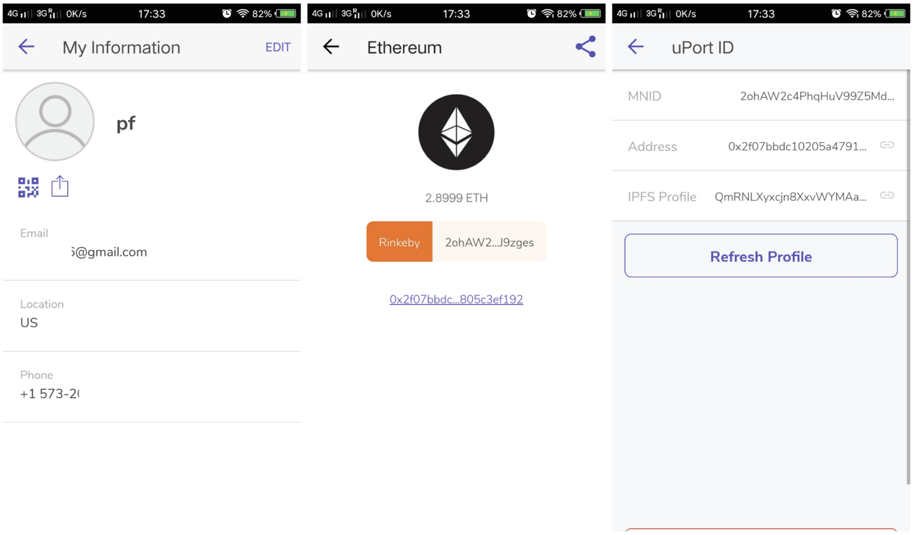
Scan QR Code. Then, a centralized web sever should be constructed in order to manage the users' requests. When the sever receives a request to upload/download a specific dataset, the sever can show a QR code on the screen of the user's device who sends the request.

Authorize Transaction. Once the user scan the QR code through uPort system, the user needs to conduct a confirm operation which is an authorization of a transaction to make the request execute successfully.

[2] https://eos.io/.

[3] https://truffleframework.com/.

[4] http://remix.ethereum.org/.



4G 3G 0K/s 17:33 82%

My Information EDIT

pf

Email
5@gmail.com

Location
US

Phone
+1 573-2

4G 3G 0K/s 17:33 82%

Ethereum

2.8999 ETH

Rinkeby 2ohAW2...J9zges

0x2f07bbdc...805c3ef192

4G 3G 0K/s 17:33 82%

uPort ID

MNID	2ohAW2c4PhqHuV99Z5Md...
Address	0x2f07bbdc10205a4791...
IPFS Profile	QmRNLXyxcjn8XxvWYMAa...

Refresh Profile

Fig. 1. Example of uPort interface.

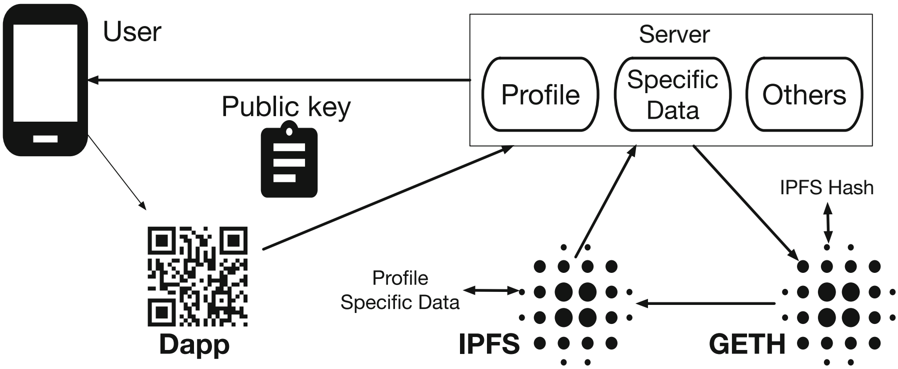
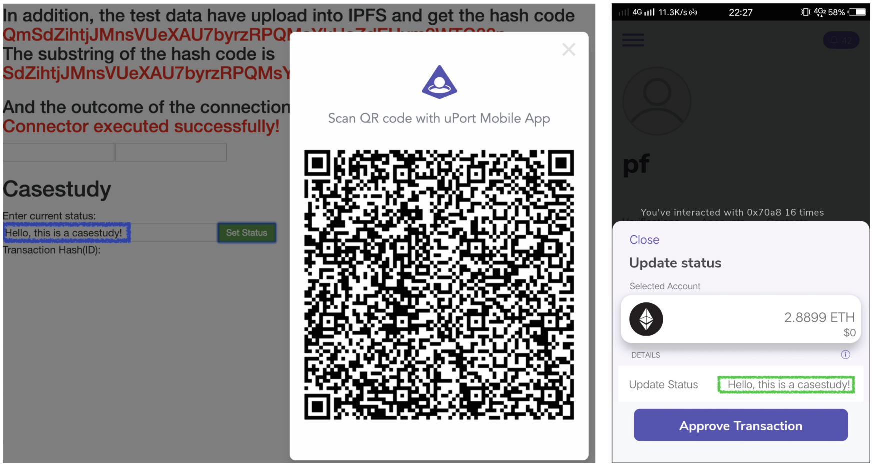


Fig. 2. Framework overview of secure data sharing system.

Execute Operation. If the operation is uploading, the system will charge some ether from the user's account, store the dataset into IPFS and link the generated hash code with the filename, and finally the IPFS's hash code will be stored in Ethereum. If the operation is downloading, the system will charge some ether from the user's account and return the hash-code of store in the IPFS system with a secure way. Then the user can download the data from the decentralized network.



In addition, the test data have upload into IPFS and get the hash code
QmSdZihtjJMnsVUeXAU7byrzRPQ
The substring of the hash code is
SdZihtjJMnsVUeXAU7byrzRPQMsY

And the outcome of the connection
Connector executed successfully!

Casestudy

Enter current status:
Hello, this is a casestudy!
Transaction Hash(ID):

Scan QR code with uPort Mobile App

4G 11.3K/s 22:27 4G 58%

pf

You've interacted with 0x70a8 16 times

Update status

Selected Account

2.8899 ETH
$0

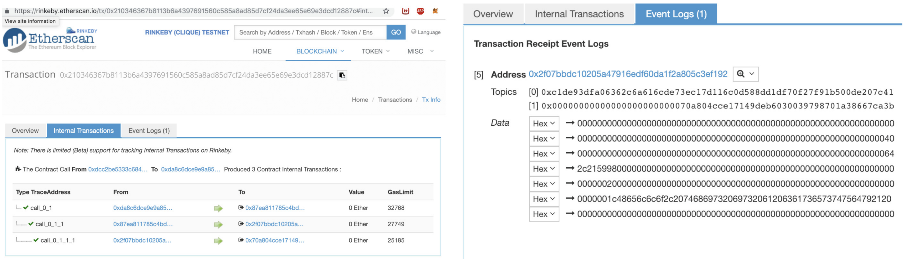
DETAILS

Update Status: Hello, this is a casestudy!

(a) Web interface and the QR code for Uport.

(b) A mobile phone interface for Uport.

Fig. 3. Interface of the simple case.



https://rinkeby.etherscan.io/tx/0x210346367b8113b6a4397691560c585a8ad85d7cf24da3ee65e69e3dcd12887c#int...

View site information
RINKEBY Etherscan The Ethereum Block Explorer
RINKEBY (CLIQUE) TESTNET
Search by Address / Txhash / Block / Token / Ens
GO
Language
HOME
BLOCKCHAIN
TOKEN
MISC
Home / Transactions / Tx Info
Overview
Internal Transactions
Event Logs (1)

Transaction 0x210346367b8113b6a4397691560c585a8ad85d7cf24da3ee65e69e3dcd12887c

Note: There is limited (Beta) support for tracking Internal Transactions on Rinkeby.

The Contract Call **From** 0xdcc2be5333c684... **To** 0xda8c6dce9e9a85... Produced 3 Contract Internal Transactions :

Type TraceAddress	From		To	Value	GasLimit
✓ call_0_1	0xda8c6dce9e9a85...	➡	0x87ea811785c4bd...	0 Ether	32768
✓ call_0_1_1	0x87ea811785c4bd...	➡	0x2f07bbdc10205a...	0 Ether	27749
✓ call_0_1_1_1	0x2f07bbdc10205a...	➡	0x70a804cce17149...	0 Ether	25185

Transaction Receipt Event Logs

```
[5] Address 0x2f07bbdc10205a47916edf60da1f2a805c3ef192
    Topics  [0] 0xc1de93dfa06362c6a616cde73ec17d116c0d588dd1df70f27f91b500de207c41
            [1] 0x00000000000000000000000070a804cce17149deb6030039798701a38667ca3b
    Data    Hex → 0000000000000000000000000000000000000000000000000000000000000000
            Hex → 0000000000000000000000000000000000000000000000000000000000000040
            Hex → 0000000000000000000000000000000000000000000000000000000000000064
            Hex → 2c21599800000000000000000000000000000000000000000000000000000000
            Hex → 0000020000000000000000000000000000000000000000000000000000000000
            Hex → 0000001c48656c6c6f2c20746869732069732061206361736573747564792120
            Hex → 0000000000000000000000000000000000000000000000000000000000000000
```

(a) Example of internal transactions of a transaction

(b) Example of event logs of a transaction

Fig. 4. Schematic diagram of a transaction on Rinkeby net.

5 Case Study

We further use a case to illustrate the effectiveness of this framework. In the framework, users can use the IPFS to store data and use the uPort system to authorize a transaction to be performed. Figure 3 shows the interface of a simple case. When we start the web service, we can upload the test data into the IPFS system, and get a hash code which is the first red string in Fig. 3. When we want to execute a transaction, we can input some information in the "input field" which is circled by a blue rectangle, then we should click the green button and we can see a QR code. Once we scan the QR code in the web page by the uPort system, we can get the information on the mobile as shown in Fig. 3(b).

If the user approve the transaction from the interface of uPort system, the transaction will be processed on Ethernet. As Fig. 4 shows, the transaction information can be seen on the website, such as internal transactions and event logs. And with the control of smart contract, the server can find the hash code of a specific dataset in the IPFS system, then the users could get the dataset in a secure way.

6 Conclusion

In this paper, we try to find a solution to share data securely and effectively in a peer-to-peer network. Specifically, we use the IPFS system to store data, manage the sensitive data by Ethereum, and verify authority through uPort system. With our design, the system can ensure the data transfer safely and each user can upload/download data through a decentralized system. Finally, we evaluate the effectiveness with a case study.

However, there are much work which could be improved. For example, in this framework, we still construct a centralized web service to manage the users' requests, that is, the framework is not an absolutely decentralized structure. Moreover, we only use a simple case study to test the effectiveness of our proposed framework, and more real data sharing experiments cloud be added in the future. Thus, we would follow these available improve-points to continue our work in the future.

Acknowledgment. The work described in this paper was supported by the National Natural Science Foundation of China (No. 91546125), the Chinese Academy of Sciences (No. XXH13505-08-04) and National Science and Technology Infrastructure Center (No. 2018DDJ1ZZ14).

References

1. Adhikari, C.L., et al.: Decentralized secure framework for sharing and managing electronic health record using ethereum-based blockchain technology (2017)
2. Bähnemann, R., Schindler, D., Kamel, M., Siegwart, R., Nieto, J.: A decentralized multi-agent unmanned aerial system to search, pick up, and relocate objects. In: Proceedings of 2017 IEEE International Symposium on Safety, Security and Rescue Robotics (SSRR), p. 8088150. IEEE (2017)
3. Bechky, B.A.: Sharing meaning across occupational communities: the transformation of understanding on a production floor. Organ. Sci. **14**(3), 312–330 (2003)
4. Benet, J.: Ipfs-content addressed, versioned, p2p file system. arXiv preprint arXiv:1407.3561 (2014)
5. Brundo, R., De Nicola, R.: Blockchain-based decentralized cloud/fog solutions: challenges, opportunities, and standards. IEEE Commun. Stand. Mag. **2**(3), 22–28 (2018)
6. Buterin, V., et al.: A next-generation smart contract and decentralized application platform. white paper (2014)
7. Chrysanthakopoulos, G., Nielsen, H.F., Moore, G.M.: Decentralized system services, uS Patent 8,122,427, 21 February 2012

8. Daswani, N., Garcia-Molina, H., Yang, B.: Open problems in data-sharing peer-to-peer systems. In: Calvanese, D., Lenzerini, M., Motwani, R. (eds.) ICDT 2003. LNCS, vol. 2572, pp. 1–15. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-36285-1_1
9. Fecher, B., Friesike, S., Hebing, M., Linek, S., Sauermann, A.: A reputation economy: results from an empirical survey on academic data sharing. arXiv preprint arXiv:1503.00481 (2015)
10. Fienberg, S.E., Martin, M.E., Straf, M.L.: Sharing Research Data. National Academy Press, Washington, D.C. (1985)
11. Gardner, D., et al.: Towards effective and rewarding data sharing. Neuroinformatics **1**(3), 289–295 (2003)
12. Hajialikhani, M., Jahanara, M.: Uniqueid: decentralized proof-of-unique-human. arXiv preprint arXiv:1806.07583 (2018)
13. Jiang, C., Gao, L., Duan, L., Huang, J.: Scalable mobile crowdsensing via peer-to-peer data sharing. IEEE Trans. Mobile Comput. **17**(4), 898–912 (2018)
14. Kansa, E.C., Kansa, S.W., Burton, M.M., Stankowski, C.: Googling the grey: open data, web services, and semantics. Archaeologies **6**(2), 301–326 (2010)
15. Lee, W., Kim, T.Y., Kang, S., Kim, H.C.: Revised P2P data sharing scheme over distributed cloud networks. In: Kim, K.J. (ed.) Information Science and Applications. LNEE, vol. 339, pp. 165–171. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-46578-3_20
16. Marshall, E.: DNA sequencer protests being scooped with his own data. Science **295**(5558), 1206–1207 (2002)
17. McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D., Barton, D.: Big data: the management revolution. Harvard Bus. Rev. **90**(10), 60–68 (2012)
18. Ng, W.S., Ooi, B.C., Tan, K.L., Zhou, A.: Peerdb: A p2p-based system for distributed data sharing. In: 19th International Conference on Data Engineering, Proceedings, pp. 633–644. IEEE (2003)
19. Patel, V.: A framework for secure and decentralized sharing of medical imaging data via blockchain consensus. Health Inf. J. (2018). https://doi.org/10.1177/1460458218769699
20. Pouwelse, J., Garbacki, P., Epema, D., Sips, H.: The bittorrent P2P file-sharing system: measurements and analysis. In: Castro, M., van Renesse, R. (eds.) IPTPS 2005. LNCS, vol. 3640, pp. 205–216. Springer, Heidelberg (2005). https://doi.org/10.1007/11558989_19
21. Shehabi, A., Stokes, J.R., Horvath, A.: Energy and air emission implications of a decentralized wastewater system. Environ. Res. Lett. **7**(2), 024007 (2012)
22. Shukla, S., Thasmiya, A., Shashank, D., Mamatha, H.: Online voting application using ethereum blockchain. In: 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 873–880. IEEE (2018)
23. Swan, M.: Blockchain thinking: the brain as a DAC (Decentralized Autonomous Organization). In: Texas Bitcoin Conference, Chicago, pp. 27–29 (2015)
24. Wood, G.: Ethereum: a secure decentralised generalised transaction ledger. Ethereum Proj. Yellow Paper **151**, 1–32 (2014)
25. Zyskind, G., Nathan, O., et al.: Decentralizing privacy: using blockchain to protect personal data. In: 2015 IEEE Security and Privacy Workshops (SPW), pp. 180–184. IEEE (2015)