

D-DRIVE: Decentralized Storage System using Blockchain and IPFS protocol

A Project Work

Submitted in the partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

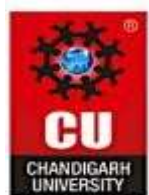
COMPUTER SCIENCE (IoT)

**Submitted by:
ANUJ SRIVASTAVA**

(20BCS4585)

Under the Supervision of

DR. PIYUSH SAMANT



**CHANDIGARH
UNIVERSITY**
Discover. Learn. Empower.

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
APEX INSTITUTE OF TECHNOLOGY**

CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413,

PUNJAB

MAY 2022



BONAFEDE CERTIFICATE

Certified that this project report **“D-Drive: Decentralized Storage System using Blockchain and IPFS protocol”** is the Bonafede work of **“ANUJ SRIVASTAVA”** who carried out the project work under my/our supervision.

SIGNATURE

Dr. Piyush Samant
SUPERVISOR/PL

SIGNATURE

Prof. Aman Kaushik
HEAD OF THE DEPARTMENT

APEX INSTITUTE OF TECHNOLOGY

Submitted for the project viva-voce examination held on 21-05-2022

INTERNAL EXAMINER

EXTERNAL EXAMINER

DECLARATION

I, **ANUJ SRIVASTAVA**, student of **‘Bachelor of Engineering in Internet of Things’**, session: **2020-2024**, Department of Computer Science and Engineering, Apex Institute of Technology, Chandigarh University, Punjab, hereby declare that the work presented in this Project Work entitled **‘Decentralized Storage System’** is the outcome of our own bona fide work and is correct to the best of our knowledge and this work has been undertaken taking care of Engineering Ethics. It contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

(ANUJ SRIVASTAVA)

Candidate UID: 20BCS6092

Date: 21th May, 2022

Place: Mohali, Punjab

ABSTRACT

Centralized cloud-based storage has received great attention and has been extensively used by many companies in the recent years. However, these cloud-based storage are not secure because of the involvement of a centralized entity or a third party.

Hence, in order to maximize data privacy and security, there is a need for blockchain-based decentralized system.

This proposed D-drive, an IPFS-based decentralized storage space to solve the problem. Decentralized storage using blockchain and IPFS will allow us to upload file in a decentralized way, allowing us to share in a trust less censorship resistant fashion.

D-Drive is a software solution trying to prove that centralized cloud-based storage applications Can be decentralized, more secure, and efficient. We proposed developing a web-based application that provides a user interface, from which the user can directly share their data or files. Then, the user file is encrypted and stored across a peer-to-peer network using IPFS protocol instead of HTTP protocol and a cryptocurrency will be used as a payment mechanism. D-Drive's primary objective is to provide secure decentralized storage space

ACKNOWLEDGEMENT

Firstly, I would like to thank my Vice Chancellor Prof. (Dr.) R.S. Bawa, HOD Mr. Vikas Wasson for their enormous support and encouragement.

I would also like to thank my Faculty In charge Mr. PIYUSH SAMANT who helped clear any doubt what so ever I had to encounter while making this project.

It is a matter of great pleasure for me to express my deep sense of gratitude and respect to all who were there on every step to guide me in and helped me make the Project better.

Lastly, I would like to thank my Family for the tremendous amount of support they offered me while making this Project. I also had the invaluable support of my friends and colleagues which helped me in completing this Project.

Yours sincerely

Anuj Srivastava

(20BCS6092)

TIMELINE CHART






TASK NAME	WEEK 01	WEEK 02	WEEK 03	WEEK 04
PLANNING				
RESEARCH				
DESIGN				
IMPLEMENTATION				
FOLLOW UP				

TABLE OF CONTENTS

Title Page	1
Declaration	2
Abstract	3
Acknowledgement	4
List of Figures	5
TIMELINE/Gann Chart	6
1.INTRODUCTION	8
1.1 Problem Definition	8
1.2 Project Overview	
1.3 System Specification	9
1.4 Software Specification	11
1.5 Hardware Specification	13
2. LITERATURE SURVEY	14
2.1 Design	14
2.2 Existing System	14
2.3 Proposed System	14
2.4 Feasibility Study	15
3. PROBLEM FORMULATION	17
4. OBJECTIVES	17
5. METHODOLOGY	18
6. RESULTS AND OUTPUT	20
7. CONCLUSION AND RECOMMENDATIONS	33
8. IMPLICATIONS OF FUTURE RESEARCH	34
9. REFERENCES	35

1.INTRODUCTION

1.1 PROBLEM DEFINATION

Centralized cloud-based storage has received great attention and has been extensively used by many companies in the recent years. However, these cloud-based storage are not secure because of the involvement of a centralized entity or a third party.

Hence, in order to maximize data privacy and security, there is a need for blockchain-based decentralized system.

This proposed D-drive, an IPFS-based decentralized storage space to solve the problem. Decentralized storage using blockchain and IPFS will allow us to upload file in a decentralized way, allowing us to share in a trust less censorship resistant fashion.

1.2 PROJECT OVERVIEW

Our methodology is to create a decentralized storage system using Blockchain and Interplanetary file system (IPFS) IPFS is a peer-to-peer hypermedia protocol designed to power decentralization Nodes participating in the network store data affiliated with globally consistent content addresses (CIDs) and advertise that they have those CIDs available for other nodes to use through publicly viewable distributed hash tables (DHTs).

Nowadays, huge amounts of data are produced every day. To meet the increasing demand for data storage space, cloud-based centralized storage systems have been widely used in terms of data storage and sharing. Cloud drive lets anyone upload and transfer data or files to the cloud and share them with anyone.

However, centralized cloud storage has a lot of disadvantages including data leaking or breaching by malware during the process and a proprietorship of data by a single entity that increases the chances of personal data being used by third parties for their analysis or personal use. We are all aware of information leakage cases of Facebook-Cambridge Analytica which motivates us to shift from centralized storage to a decentralized storage system

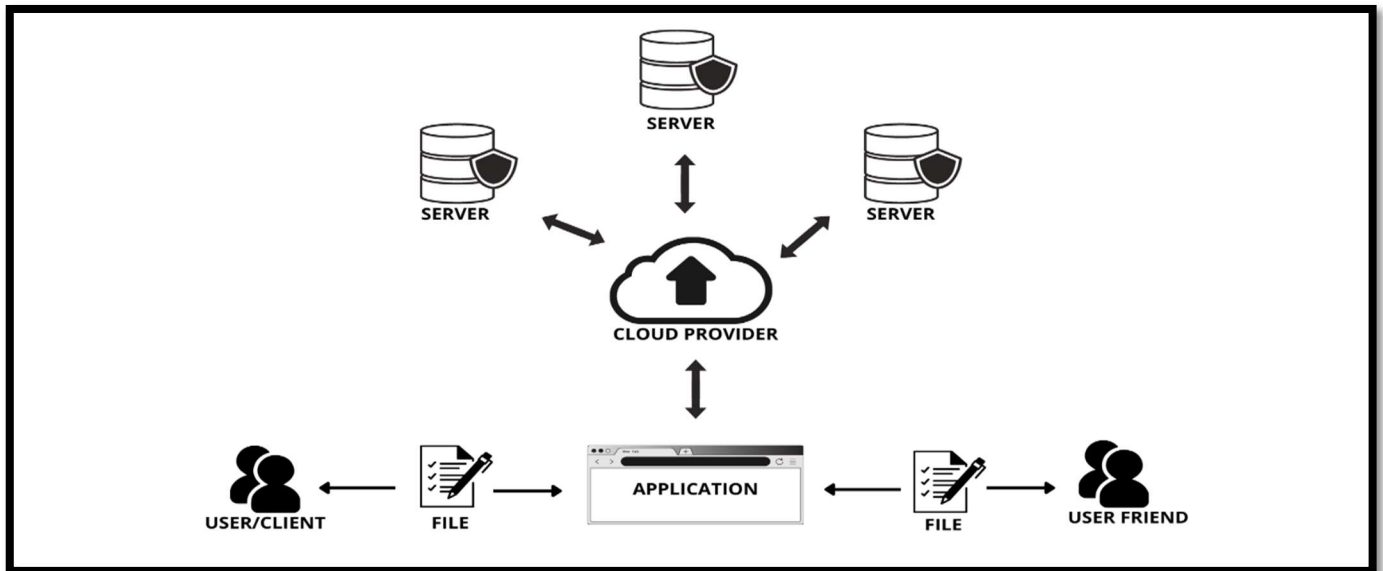


Fig 1. Schema of Cloud Storage

In permissioned or private blockchain only known and identifiable set of participants are explicitly admitted to the blockchain network. This reduces the presence of malicious actors within the network. As a result, only authenticated and authorized actors can participate in the network which increases the security of the system as required by the enterprise applications.

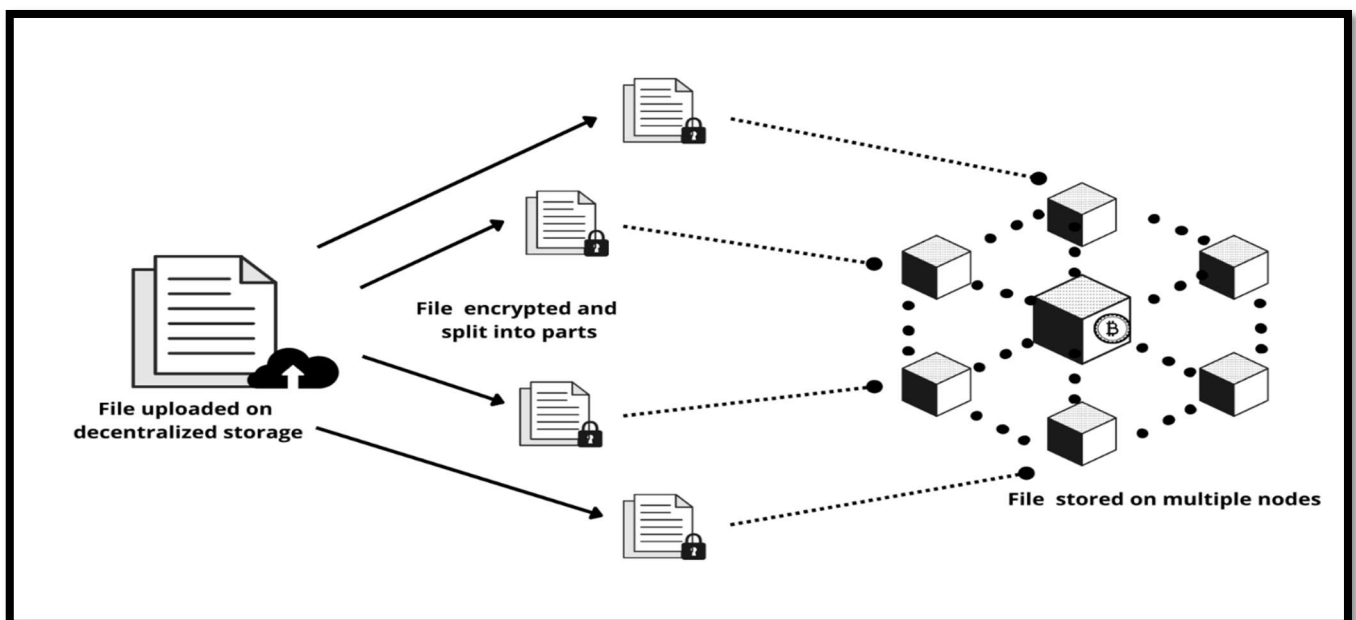


Fig 2. Schema of Decentralized Storage

1.3 SYSTEM SPECIFICATIONS

1. i3 Processor or above
2. 100 MB of free hard-drive space
3. 500 MB of RAM

1.4 SOFTWARE SPECIFICATIONS

1. Operating System: Windows (7 or above)
2. **Truffle:** Framework for creating Ethereum smart contracts (solidity programming language)
3. **Ganache:** It is the blockchain that will run on our computer to run transactions and deploy smart contracts without having to pay real money.
4. **Metamask:** Browser extension to connect the browser to the blockchain. It is Ethereum wallet to turn our browser to Blockchain browser.
5. **Web3.js:** Collection of libraries that allows you to interact with a local or remote Ethereum node using HTTP, IPC, WebSocket (connects our application to the blockchain)
6. **ReactJS:** JavaScript library used for frontend.
7. **Solidity:** Solidity is an object-oriented programming language for implementing smart contracts
8. **IPFS:** The InterPlanetary File System is a protocol and peer-to-peer network for storing and sharing data in a distributed file system.

2.LITERATURE SURVEY

2.1 DESIGN

Centralized cloud-based storage has received great attention and has been extensively used by many companies in the recent years. However, these cloud-based storage are not secure because of the involvement of a centralized entity or a third party. Hence, in order to maximize data privacy and security, there is a need for blockchain-based decentralized system. This proposed D-drive, an IPFS-based decentralized storage space to solve the problem. Decentralized storage using blockchain and IPFS will allow us to upload file in a decentralized way, allowing us to share in a trust less censorship resistant fashion.

Our methodology is to create a decentralized storage system using Blockchain and Interplanetary file system (IPFS)

IPFS is a peer-to-peer hypermedia protocol designed to power decentralization

Nodes participating in the network store data affiliated with globally consistent content addresses (CIDs) and advertise that they have those CIDs available for other nodes to use through publicly viewable distributed hash tables (DHTs)

In permissioned or private blockchain only known and identifiable set of participants are explicitly admitted to the blockchain network. This reduces the presence of malicious actors within the network.

As a result, only authenticated and authorized actors can participate in the network which increases the security of the system as required by the enterprise applications

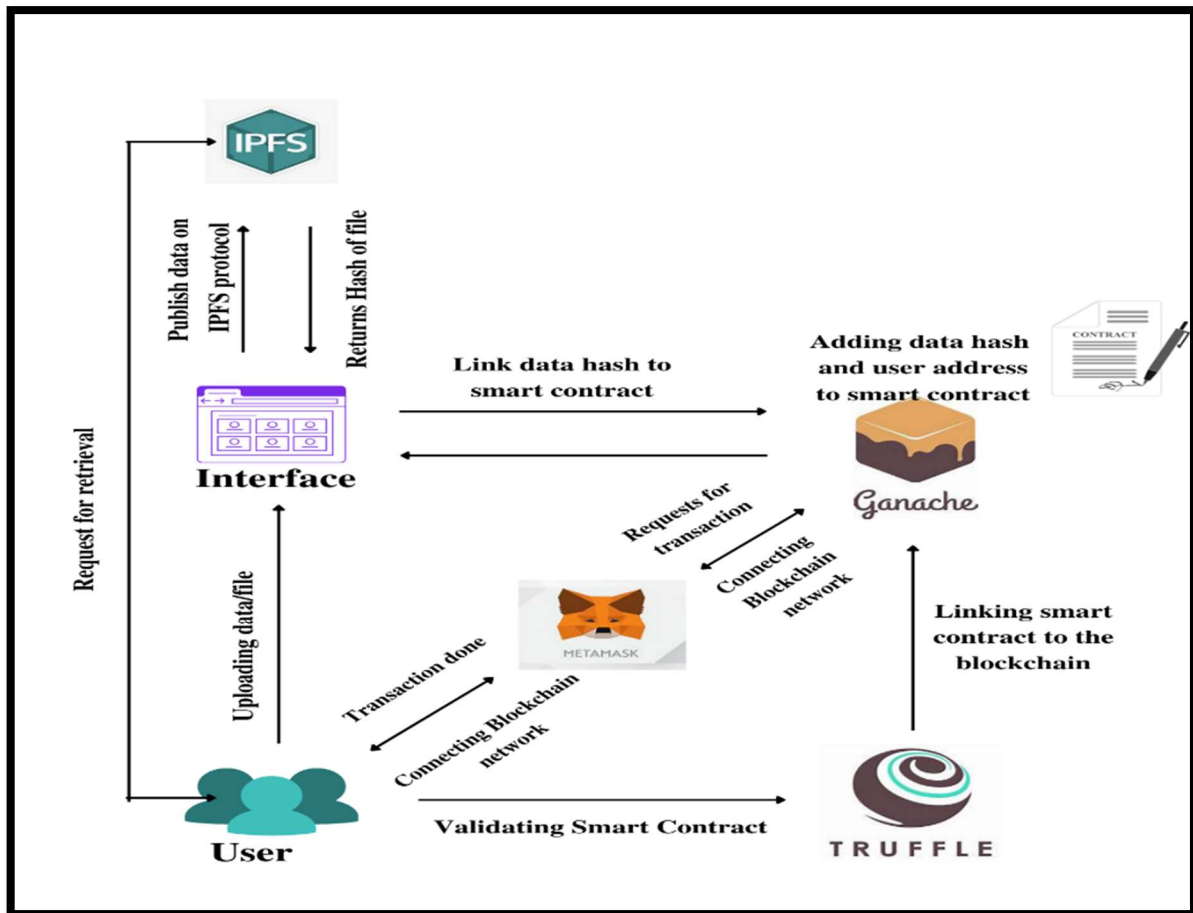
2.2 EXISTING SYSTEM

Nowadays, huge amounts of data are produced every day. To meet the increasing demand for data storage space cloud-based centralized storage systems have been widely used in terms of data storage and sharing. Cloud drive lets anyone upload and transfer data or files to the cloud and share them with anyone. However, centralized cloud storage has a lot of disadvantages including data leaking or breaching by malware during the process and a proprietorship of data by a single entity that increases the chances of personal data being used by third parties for their analysis or personal use. We are all aware of information leakage cases of Facebook -Cambridge Analytica, which motivates us to shift from centralized storage to a decentralized storage system.

2.3 PROPOSED SYSTEM

Decentralization distributes data, applications, power, people, or things into a peer-to-peer rather than on a central authority. If a System is decentralized, it means that it is not controlled, or managed by a single entity or authority. In addition, decentralization facilitates more benefits such as privacy, security, low price, and completely removing trust in a third party. Nowadays, the concept of Inter-Planetary File system (IPFS), has been introduced. The Inter-Planetary File System (IPFS) is a version controlled decentralized file system that uses Distributed Hash Table (DHT) technology for storing data in a peer-to-peer network. It enables us to store and share any type and size of data over a decentralized network without any limitations. Still IFPS needs lot of research to meet specific demands of real world problem, such as how to allow users to share data for

multiple users in different organizations without any trust issues. To sort out above problem cryptographic techniques have been widely used in traditional storage system. This paper proposed a data storage system named D-Drive that provides decentralized, secure, and transparent means of storing and sharing data. For that, we make the use of decentralized technologies to build this system. First, we rely on advantages of IPFS network to store data of users in a decentralized manner.



3.BACKGROUND

A. Blockchain Technology

There are three types of blockchain named as public, private, and consortium blockchain in decentralized systems. The concept of cryptocurrency is more related to solving issues of a public blockchain. Blockchain is a collection of blocks, in which each block is composed by transactions and includes a hash of the previous block. Distributed technology provides immutability of the data because changing data in one block will affect all next blocks and is beneficial for record-keeping, digital notary, and smart contracts. This technology has been initially used for digital currency and secure distributed transaction storage systems. Bitcoin is a great example of cryptocurrency.

Ethereum is another decentralized, open source, public platform based on blockchain technology. The structure of the Ethereum is almost similar to the other blockchain networks. It has a feature called a smart contract, which facilitates online contract agreements.

B. Smart Contract

A smart contract is a small piece of code that executes on the blockchain platform without the involvement of any third party. The platform includes a virtual machine - Ethereum Virtual Machine (EVM), which can execute scripts using an Ethereum computer network. Ethereum has a cryptocurrency named "Ether" which can be transferred between accounts and used to pay miners as a gas fee to help with the calculations.

C. Distributed Systems and Hash Tables

Hash is the output of a hashing algorithm such as MD5 (MessageDigest 5) or SHA (secure hash algorithm). It is used for several different areas such as cryptography and data indexing. A hash function generates a hash value that can only be decoded by looking up the value from a hash table. The table may be any data structure. Hash function is one-way and noninvertible. A DHT is a distributed form of hash table. The main advantages of the DHT is that it makes blockchain faster as all nodes can be added or removed at a minimum time just by redistributing the keys.

D. IPFS

The IPFS (InterPlanetary File System) is a protocol for sharing and storing data on a peer-to-peer distributed network that uses DHT to track the information about data. Hash tables is used to store a data package. Kademlia, is used to learn about data in nodes. Kademlia is a hash table for decentralized computer networks designed by Petar Maymounkov and David Mazières in 2002. When we upload the data, a hash has been generated. IPFS stores the hash and then user can use the hash to get their data back. When data is uploaded on the IPFS network, the data will split into multiple pieces. These pieces of data are identified with its own hash.

4.PROBLEM FORMULATION

Centralized cloud-based storage has received great attention and has been extensively used by many companies in the recent years. However, these cloud-based storage are not secure because of the involvement of a centralized entity or a third party.

Hence, in order to maximize data privacy and security, there is a need for blockchain-based decentralized system. This proposed D-drive, an IPFS-based decentralized storage space to solve the problem. Decentralized storage using blockchain and IPFS will allow us to upload file in a decentralized way, allowing us to share in a trust less censorship resistant fashion.

OBJECTIVES

1. Maximize Data Privacy
2. Ensure Security
3. High data redundancy.
4. Flexible load balancing.
5. Verifiability

ADVANTAGES

1. Network transparency
2. Increased Reliability and availability
3. No involvement of third part
4. No risk of data lost

4.METHODOLOGY

The proposed prototype enables user to upload the data or file to the peer-to-peer network. For that, the user need to configure blockchain network(Ganache is used for local blockchain network) and integrating it into the web browser using the Metamask extension.

Then, the user needs a blockchain network, which is provided by Ganache. The accounts provided by Ganache are added to Metamask for transaction purpose. A Specific amount of Gas is needed in the form of Ether. Ether is a type of crypto token that fuels the blockchain network. Then, the user need to create an account on metamask and connect it with their wallet. Now, as our Web browser supports blockchain network, we can now upload the files through our own designed user interface. When the user select the file to upload that file goes to the IPFS and IPFS returns a hash value that will mapped with smart contract. After that the user need to pay the gas amount from their metamask account. After the successful payment, the smart contract allows the file to get uploaded on a peer-to-peer network. The process of retrieving the file from IPFS requires the previously obtained IPFS hash value, which is generated after uploading the file. We have to put the IPFS hash in the web browser, IPFS will search for the file, and preview is shown. Thus, the file is retrieved back from the IPFS System.

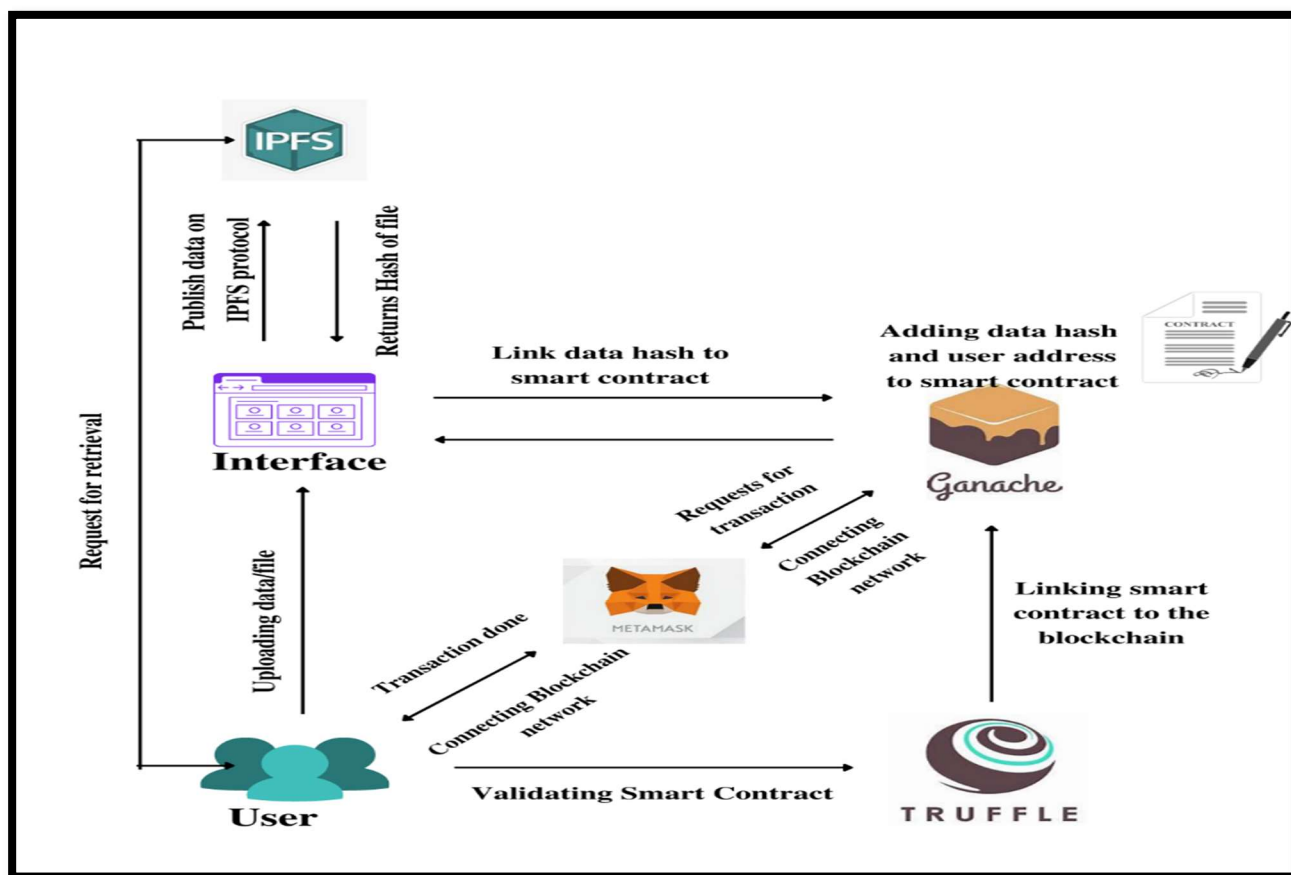
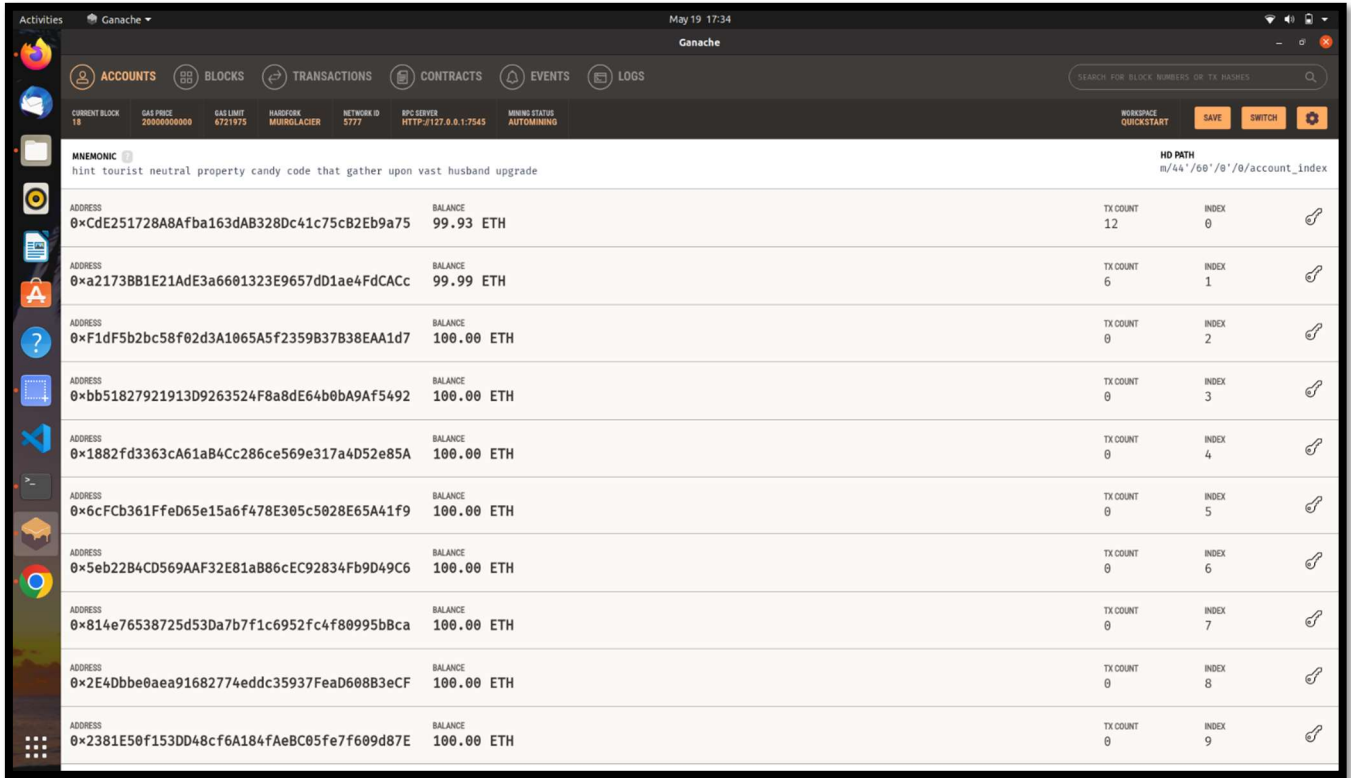


Fig 6. Architecture of D-Drive

5.IMPLEMENTATION

1. ETH COUNT AFTER ALL TRANSACTIONS



ACCOUNTS			
Mnemonic: hint tourist neutral property candy code that gather upon vast husband upgrade			
Address	BALANCE	TX COUNT	INDEX
0xCdE251728A8Afb163dAB328Dc41c75cB2Eb9a75	99.93 ETH	12	0
0xa2173BB1E21AdE3a6601323E9657dD1ae4FdCACc	99.99 ETH	6	1
0xF1dF5b2bc58f02d3A1065A5f2359B37B38EAA1d7	100.00 ETH	0	2
0xbb51827921913D9263524F8a8dE64b0bA9Af5492	100.00 ETH	0	3
0x1882fd3363cA61aB4Cc286ce569e317a4D52e85A	100.00 ETH	0	4
0x6cFCb361FfeD65e15a6f478E305c5028E65A41f9	100.00 ETH	0	5
0x5eb22B4CD569AAF32E81aB86cEC92834Fb9D49C6	100.00 ETH	0	6
0x814e76538725d53Da7b7f1c6952fc4f80995bBca	100.00 ETH	0	7
0x2E4Dbbe0aea91682774eddc35937FeaD608B3eCF	100.00 ETH	0	8
0x2381E50f153DD48cf6A184fAeBC05fe7f609d87E	100.00 ETH	0	9

Fig (1)

2. METAMASK NOTIFICATION TO CONFRIM THE TRASACTION

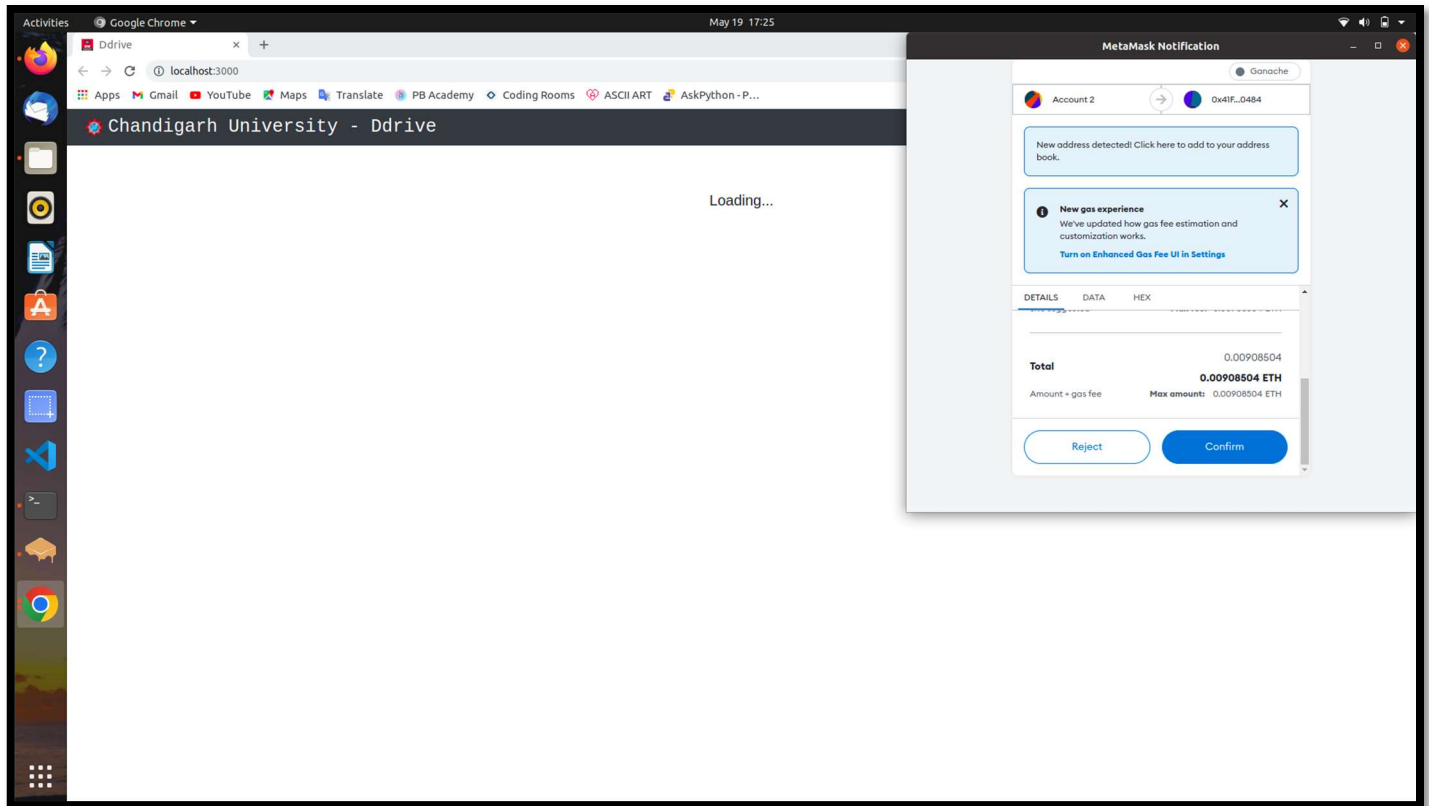
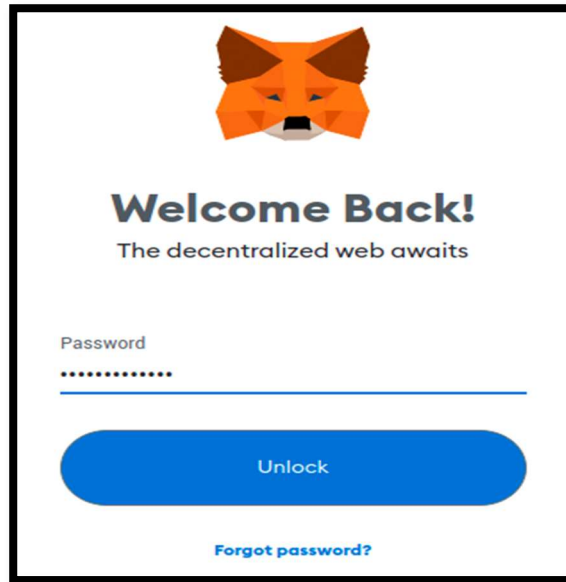


Fig (2)

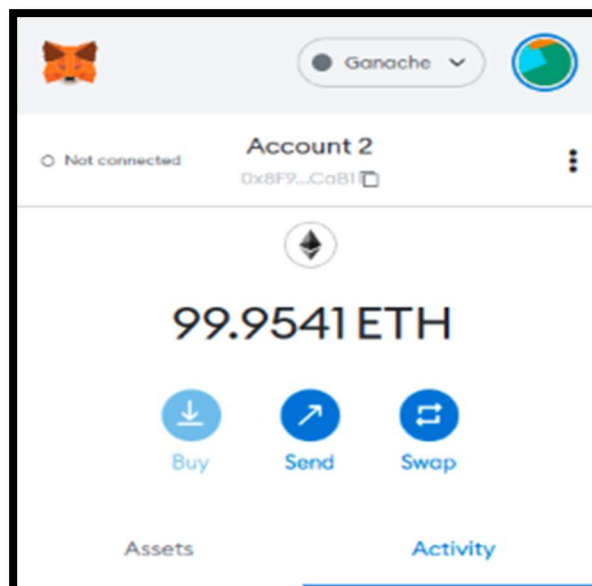
6.RESULT AND ANALYSIS

This paper proposed developing a web-based application that provides a user interface, from which the user can directly upload or share their data and files over a decentralized network. The proposed solution works in multiple segments.

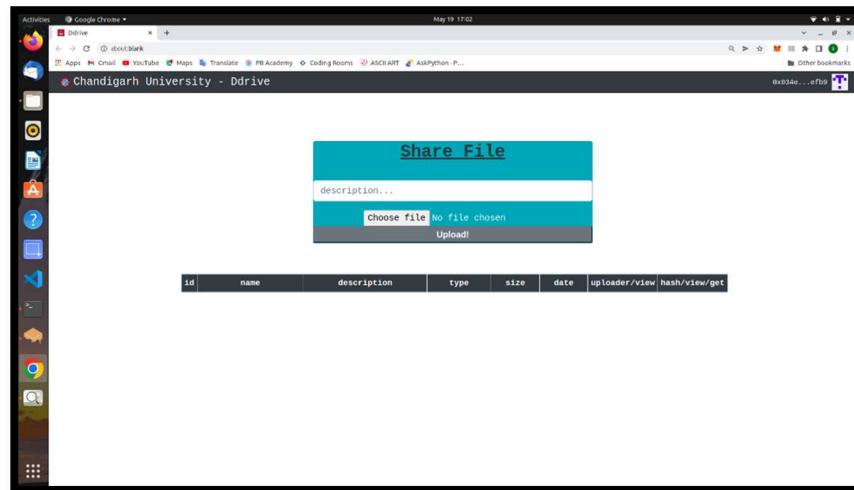
- Firstly, the user needs to create an account on metamask and login with their credentials.



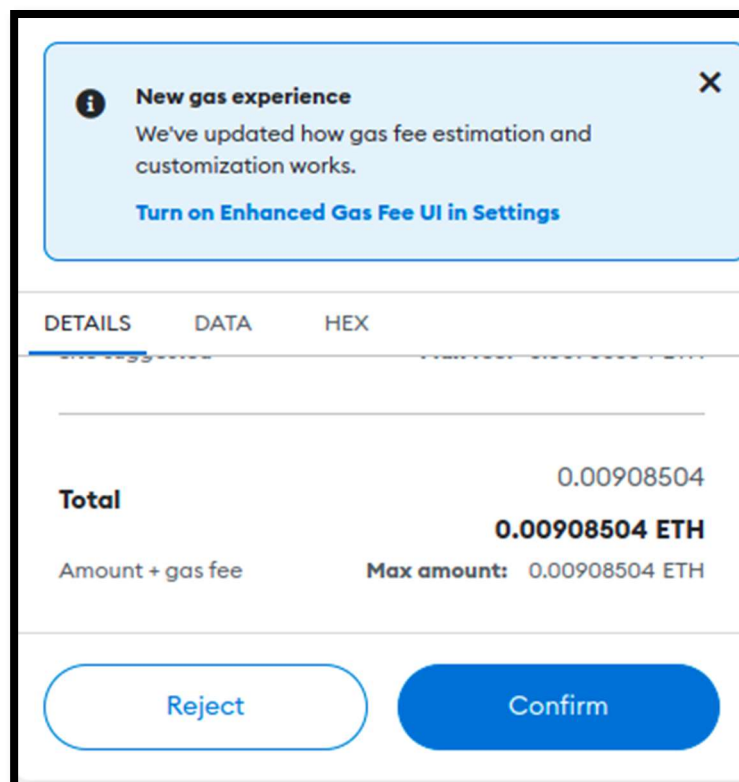
- Then, the user needs to connect their metamask account with their wallet. The user's account address and wallet balance are fetched in the metamask account through web3.js



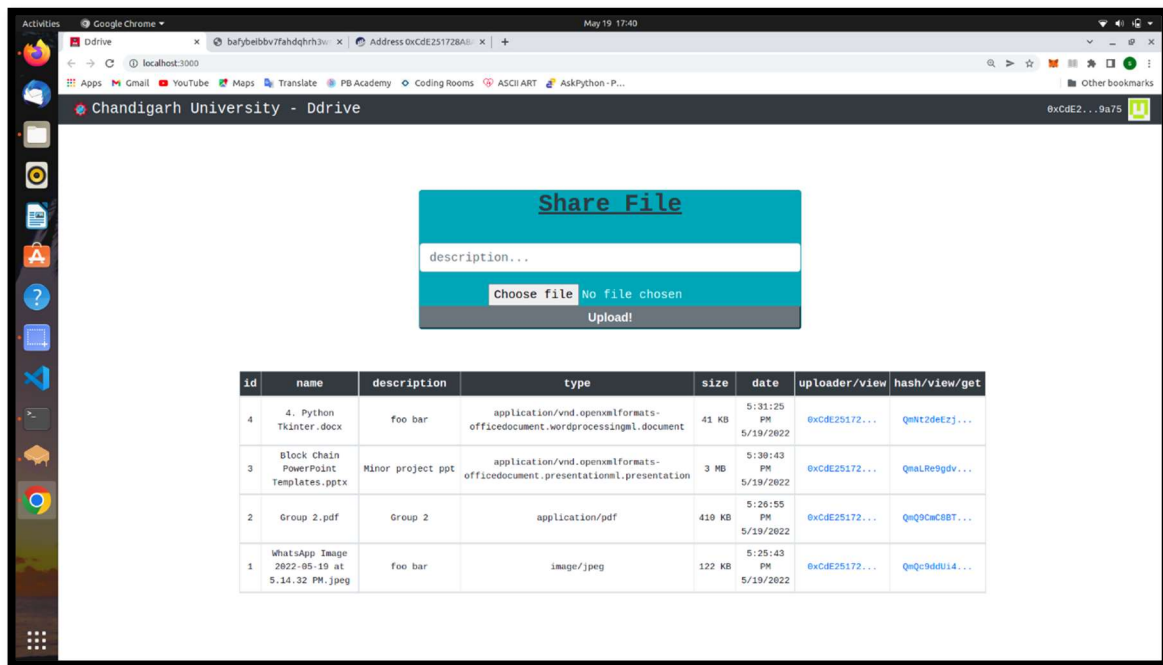
- After that, the user needs to open Dapp(D-Drive) and select the files to upload.



- Further, the AES algorithm link the user wallet address as a key and encrypt the uploaded file. Payment dialogue box pop-ups for the payment confirmation.



- After the successful payment, the user's file stored on the peer-to-peer network using IPFS protocol. IPFS returns a hash value of the uploaded file, that will mapped with address using a smart contract and get stored over a blockchain network.



- Further, if the user want to share their uploaded file, they just need to share the hash view with the other user, so that they can view or download by clicking it.

id	name	description	type	size	date	uploader/view	hash/view/get
4	4. Python Tkinter.docx	foo bar	application/vnd.openxmlformats-officedocument.wordprocessingml.document	41 KB	5:31:25 PM 5/19/2022	0xCdE25172...	QmNt2deEzj...
3	Block Chain PowerPoint Templates.pptx	Minor project ppt	application/vnd.openxmlformats-officedocument.presentationml.presentation	3 MB	5:30:43 PM 5/19/2022	0xCdE25172...	QmaLRe9gdv...

7. CONCLUSION AND RECOMMENDATIONS

This paper proposed an innovative IPFS based decentralized storage system named as D-Drive. The proposed system maximize the data security by distributing our data across peer-to-peer network in a decentralized manner. This system uses the IPFS protocol for ensuring the confidentiality of the user's data.

Apart from these advantages, it needs certain level of improvement for accuracy and speed. In the proposed system, IPFS protocol is used, but if there is a better system in future, that can also be implemented.

Nowadays, huge amounts of data are produced every day. To meet the increasing demand for data storage space cloud-based centralized storage systems have been widely used in terms of data storage and sharing.

Cloud drive lets anyone upload and transfer data or files to the cloud and share them with anyone. However, centralized cloud storage has a lot of disadvantages including data leaking or breaching by malware during the process and a proprietorship of data by a single entity that increases the chances of personal data being used by third parties for their analysis or personal use.

8. IMPLICATIONS FOR FUTURE RESEARCH

Nothing is 100% perfect, there always a scope of improvement. Similarly, this project can be improved further. There are many areas in this project which can be expanded more on a larger scale that could not be achieved due limited time and resources. Some more research needs to be done to meet the requirements of a learner. Following are some implications and changes that can be done in this project.

- ❑ **Network transparency:**

This basically refers to the freedom for the user from the operational details of the network. These are of two types Location and naming transparency.

- ❑ **Replication transparencies:**

It basically made user unaware of the existence of copies as we know that copies of data may be stored at multiple sites for better availability performance and reliability.

- ❑ **Fragmentation transparency:**

It basically made user unaware about the existence of fragments it may be the vertical fragment or horizontal fragmentation.

- ❑ **Increased Reliability and availability**

Reliability is basically defined as the probability that a system is running at a certain time whereas Availability is defined as the probability that the system is continuously available during a time interval.

REFERENCES

- [1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," [Online; accessed 01 Nov. 2019]. Available: <https://bitcoin.org/bitcoin.pdf>.
- [2] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An overview of blockchain technology: Architecture, consensus, and future trends," in IEEE International Congress on Big Data, 2017, pp. 557–564.
- [3] Y. Peng, W. Zhao, F. Xie, Z.-h. Dai, Y. Gao, and D.-q. Chen, "Secure cloud storage based on cryptographic techniques," Journal of China Universities of Posts and Telecommunications, vol. 19, pp. 182–189, 2012.
- [4] I.-C. Lin and T.-C. Liao, "A survey of blockchain security issues and challenges.," International Journal of Network Security, vol. 19, no. 5, pp. 653–659, 2017.
- [5] 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.
- [6] M. Conti, E. S. Kumar, C. Lal, and S. Ruj, "A survey on security and privacy issues of bitcoin," IEEE Communications Surveys & Tutorials, vol. 20, no. 4, pp. 3416–3452, 2018.
- [7] E. Politou, E. Alepis, C. Patsakis, F. Casino, and M. Alazab, "Delegated content erasure in IPFS," Future Generation Computer Systems, vol. 112, pp. 956–964, 2020.
- [8] M. H. ur Rehman, K. Salah, E. Damiani, and D. Svetinovic, "Trust in blockchain cryptocurrency ecosystem," IEEE Transactions on Engineering Management, 2019.
- [9] V. L. Lemieux, "Blockchain and distributed ledgers as trusted recordkeeping systems: An archival theoretic evaluation framework," in Future Technologies Conference (FTC), vol. 2017, pp. 1–11, 2017.
- [10] J. Benet, "IPFS - Content Addressed, Versioned, P2p File System," [Online; accessed 01 Nov. 2019]. Available: <https://www.hirego.io/>.
- [11] X. Xu, I. Weber, M. Staples, L. Zhu, J. Bosch, L. Bass, C. Pautasso, and P. Rimba, "A taxonomy of blockchain-based systems for architecture design," in International Conference on Software Architecture (ICSA), 2017, pp. 243–252.
- [12] P. Maymounkov and D. Mazieres, "Kademlia: A peer-to-peer information system based on the XOR metric," in International Workshop on Peer-to-Peer Systems, pp. 53–65, Springer, 2002.
- [13] I. I. Yiakoumis, M. E. Papadonikolakis, H. E. Michail, A. P. Kakarountas, and C. E. Goutis, "Maximizing the hash function of authentication codes," IEEE Potentials, vol. 25, no. 2, pp. 9–12, 2006. [14] Y. Hirai, "Defining the ethereum virtual machine for interactive theorem provers," in International Conference on Financial Cryptography and Data Security, 2017, pp. 520–535.
- [15] J. Isaak and M. J. Hanna, "User data privacy: Facebook, cambridge analytica, and privacy protection," Computer, vol. 51, no. 8, pp. 56–59, 2018.
- [16] C. Cadwalladr and E. Graham-Harrison, "Revealed: 50 million facebook profiles harvested for cambridge analytica in major data breach," The Guardian, vol. 17, p. 22, 2018.

***** THANK YOU *****