D-Drive: Decentralized Storage Space using Blockchain and IPFS Protocol

Dr. Piyush Samant
Apex Institute of Technology(CSE))
Chandigarh University
Mohali, India
piyushsamantpth@gmail.com

Anuj Srivastava
Apex Institute of Technology(CSE)
Chandigarh University
Mohali, India
innovativeanujsrivastava@gmail.com

Smita Shinde
Apex Institute of Technology(CSE)
Chandigarh University
Mohali, India
smitash3011@gmail.com

Ananya Singh
Apex Institute of Technology(CSE)
Chandigarh University
Mohali, India
ananyasgh941@gmail.com

Siddharth Singh
Apex Institute of Technology(CSE)
Chandigarh University
Mohali, India
siddharthsj06@gmail.com

Ronak Chaudhary
Apex Institute of Technology(CSE)
Chandigarh University
Mohali, India
ronakchoudhary2528@gmail.com

***Abstract*—Centralized cloud-based storage has received great attention and has been extensively used by many companies in recent years. However, these cloud based storage are not secure because of the involvement of a centralized entity or a third party. On the other hand, there is a need for blockchain based decentralized storage to maximize data privacy and security. This paper proposed D-Drive, an IPFS-based decentralized storage space to solve the problem. D-Drive is a software solution trying to prove that centralized cloud-based storage applications can be decentralized, more secure, and efficient. This paper 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.**

Keywords*— *Blockchain, Data Security, IPFS, Encryption, Cloud Storage, Decentralized storage

I. Introduction

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 [15], [16] which motivates us to shift from centralized storage to a decentralized storage 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 [5]. 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) [10], has been introduced. The Inter-Planetary File System (IPFS) is a

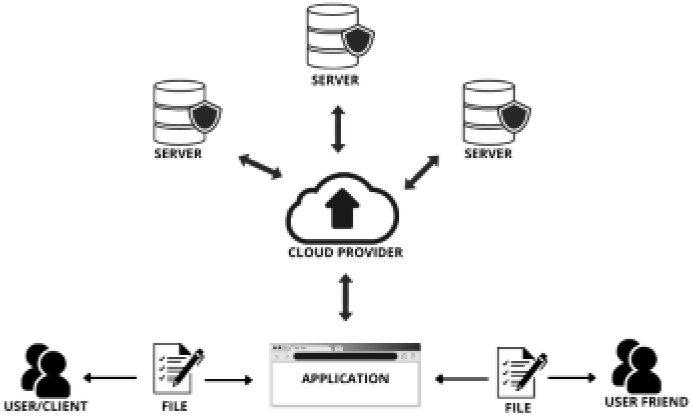
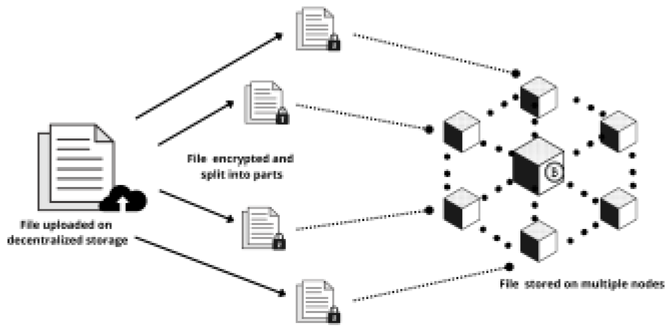


Fig. 1. Schema of centralized cloud-based storage system



File encrypted and split into parts
File uploaded on decentralized storage
File stored on multiple nodes

Fig. 2. Schema of decentralized storage system

version controlled decentralized file system that uses Distributed Hash Table (DHT) technology for storing data in a peer-to-peer network [7].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 [4]. To sort out above problem cryptographic techniques [3], 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.

II. Problem Identification

Cloud-based storage pointed out the problem of data privacy and security, as there is an involvement of a centralized entity or a third party. The proposed system introduced the need for blockchain based decentralized storage to maximize the data privacy and security.

III. Background

A. Blockchain Technology

There are three types of blockchain named as public, private, and consortium blockchain in decentralized systems [2]. The concept of cryptocurrency is more related to solving issues of a public blockchain. Blockchain [1], [4] 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 [8], and is beneficial for record-keeping, digital notary, and smart contracts [9]. This technology has been initially used for digital currency [6], and secure distributed transaction storage systems. Bitcoin is a great example of cryptocurrency. Ethereum [11], 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) [14],which can execute scripts using an Ethereum computer network.

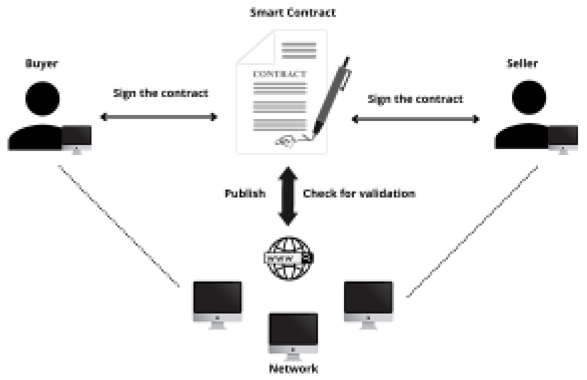


Fig. 3. Schema of blockchain transaction

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 (Message Digest 5) or SHA (secure hash algorithm) [13]. 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 [13].

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 [12], is used to learn about data in nodes. Kademlia is a hash table for decentralized computer networks designed by Petar Maymounkov and David Mazi‘eres 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.

IV. Methodology and Architecture

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.

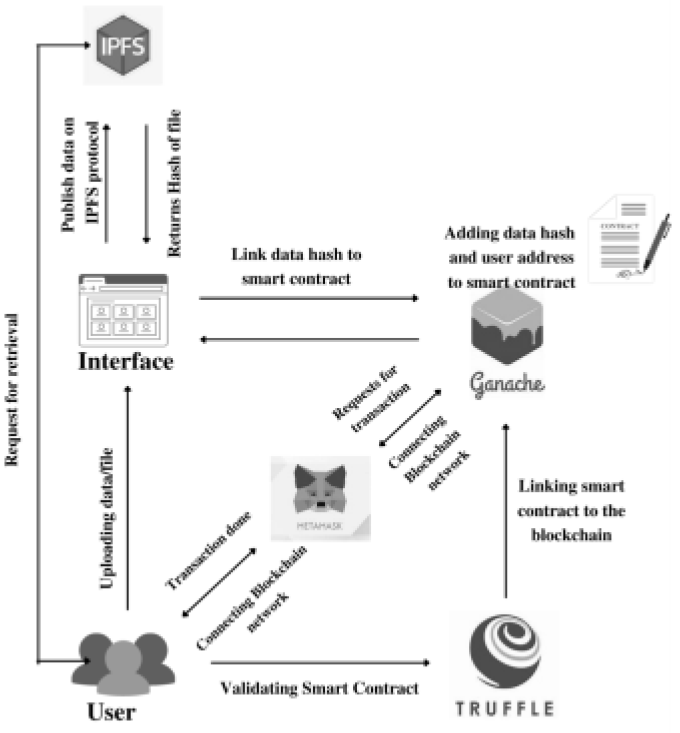

IPFS
Publish data on IPFS protocol
Returns Hash of file
Link data hash to smart contract
Adding data hash and user address to smart contract
CONTRACT
Interface
Ganache
Request for retrieval
Requests for transaction
Connecting Blockchain network
METAMASK
Linking smart contract to the blockchain
Uploading data/file
Transaction done
Connecting Blockchain network
User
Validating Smart Contract
TRUFFLE

Fig. 4. Architecture of D-Drive

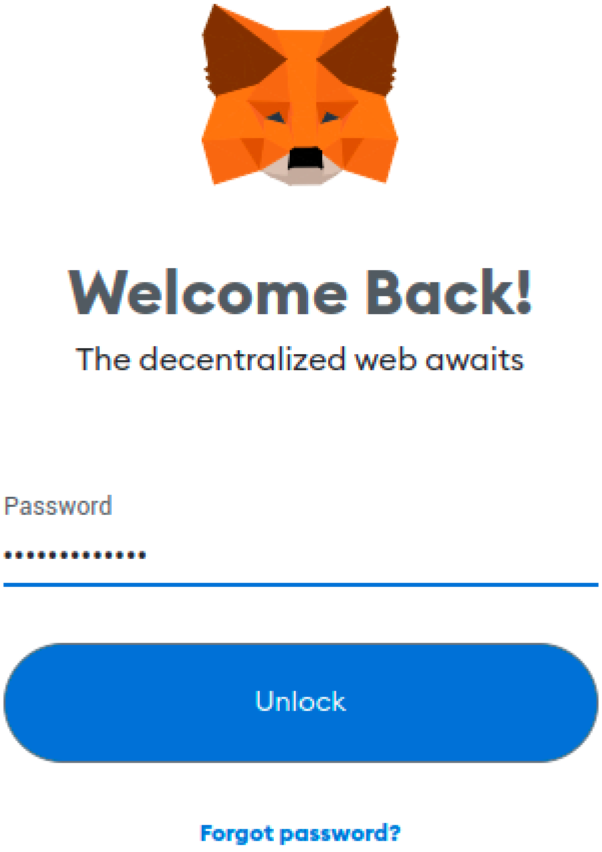
From the fig 4. the user needs a blockchain network, which is provided by Ganache. The accounts provided by Ganache are added to Metamask for transaction purpose. A Specific amount of Gas is needed in the form of Ether. Ether is a type of crypto token that fuels the blockchain network. Then, the user 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.

V. Results 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.



Welcome Back!

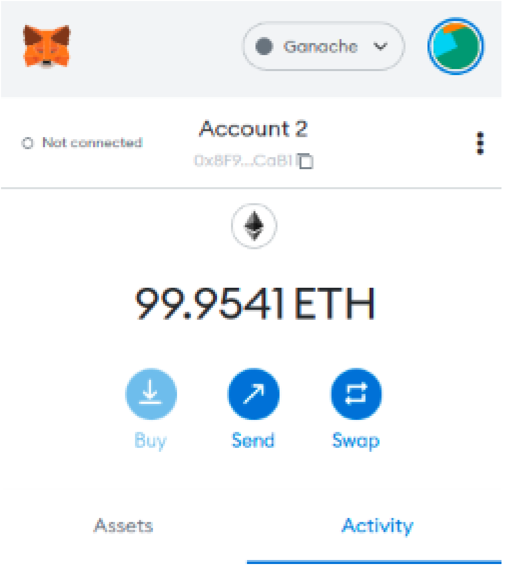
The decentralized web awaits

Password

•••••••••••••

Fig. 5. Metamask login interface

- 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



Not connected

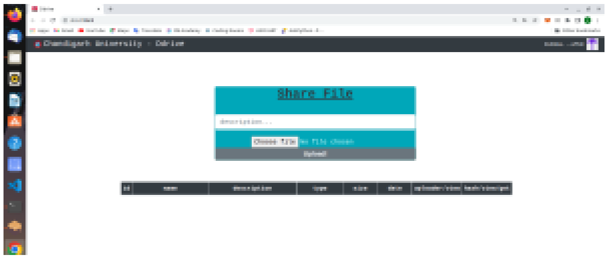
Account 2

0x8F9...CaB1

99.9541 ETH

Fig. 6. Account balance fetched

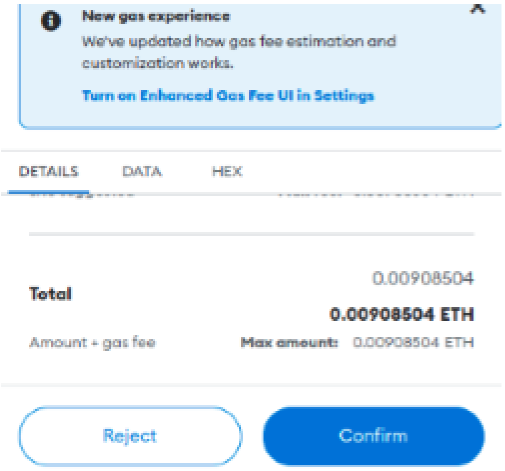
- After that, the user needs to open Dapp(D-Drive) and select the files to upload.



id	name	description	type	size	date	uploader/view	hash/view/get

Fig. 7. User Interface

- 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.



New gas experience

We've updated how gas fee estimation and customization works.

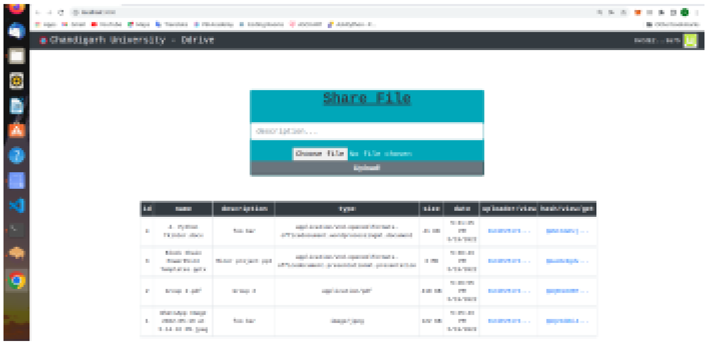
Total 0.00908504

0.00908504 ETH

Amount + gas fee Max amount: 0.00908504 ETH

Fig. 8. Payment confirmation dialogue box

- 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.



Chandigarh University - Ddrive

Share File

description...

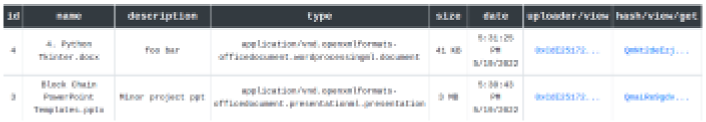
Choose file No file chosen

Upload

id	name	description	type	size	date	uploader/view	hash/view/get
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

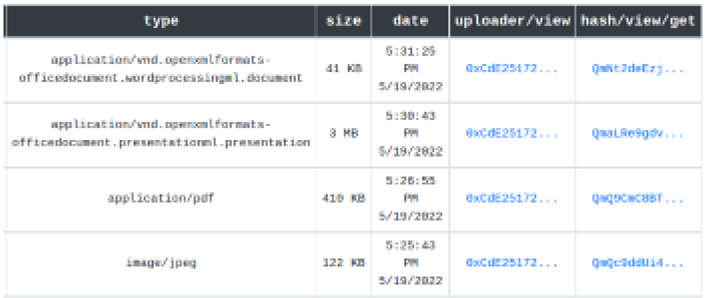
Fig. 9. File Uploaded with hash value

- 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...	QmNt2deErj...
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...

Fig. 10. Hash value of a uploaded file



type	size	date	uploader/view	hash/view/get
application/vnd.openxmlformats-officedocument.wordprocessingml.document	41 KB	5:31:25 PM 5/19/2022	0xCdE25172...	QmNt2deErj...
application/vnd.openxmlformats-officedocument.presentationml.presentation	3 MB	5:30:43 PM 5/19/2022	0xCdE25172...	QmaLRe9gdv...
application/pdf	410 KB	5:26:55 PM 5/19/2022	0xCdE25172...	QmQ9CmC8BT...
image/jpeg	122 KB	5:25:43 PM 5/19/2022	0xCdE25172...	QmQc9ddUi4...

Fig. 11. Hash value of a uploaded file

VI. Conclusion and Future Scope

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.

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.