# Research Paper on Decentralized Storage

**Daniel Maliro**

**Abstract**: In a forever-changing society, we are losing more and more privacy in exchange for security. Decentralized Storage could help increase privacy along with security. We differentiate between what centralized storage provides and what decentralized storage can offer. The current uses of decentralized storage also promote the benefits of why this system is very beneficial to today's society. There can even be a more local use of decentralized storage could be implemented to higher elevate the security and privacy of your data as well.

## I. Introduction

In the technological age, people use computers as a profession and as a necessity for day-to-day processes. Due to our dependence on our devices, malicious individuals seek to take advantage of them. In our current age, data that isn't stored on our devices are stored in big servers. These servers are mostly managed by big companies, e.g. Google, Dropbox, AWS, etc. There are several cases where these companies have been breached and had personal information leaked to people who would want to buy your information. In July 2019, Capital One, an American Bank, faced a security failure that resulted in many people who use their services to have their data public [1]. This data breach affected 106 million customers [1], which raised questions about cybersecurity from the tech giants. Since our data is in the hands and care of big organizations, consumers are subject to their will. I.e. they can decide to censor what they don't agree with off of their platforms. This creates limited mobility in an environment where freedom is one of its core attributes.

Decentralized Storage has become more and more popular as a remedy to our current questions about keeping our data safe from other people. It promotes the use of peer-to-peer networking. This type of connection was the initial purpose of the web before centralized storage became popular. The reason why most consumers rely on these companies is that they don't have to worry much about security, and it is very easy and fast to retrieve your data. In the 1960s, Professor Börje Langefors was against total management systems and argued that they would ultimately fail [4]. These systems at the time were called MIS (management information systems). Langefors gave speeches on why these types of systems will not work, however, many centralized storage methods were becoming more and more popular at the time.

Overall, Decentralized Storage is a system of storing data that does not include any third party interfering or altering the data we want to represent on the World Wide Web.

## II. Comparison with Centralized Storage

Centralized Storage was the favored system of management, that implemented MIS and sprouted in the 1970s [4]. The word "centralized" implies that all clients have their information "centralized" at a single point. This can result in a "single point of failure" situation. If one thing goes wrong with that server, your data is at risk of being lost.

Decentralized Storage is a system where your data is spread across multiple devices through the P2P network. Your data is split into multiple smaller pieces, which may not even be all on the same device. These smaller pieces are normally called blocks, and there are many copies of these blocks on different computers to prevent data from being lost. Due to the exclusion of a third party, your information is sent directly from and to its intended positions. This coupled with low latency [3] when downloading makes it appealing to consumers.

### *A. Performance*

Companies like Google and AWS have a centralized storage architecture for storing user data. This method is proven to be extremely fast compared to other architecture. This is because of how the data is saved to their servers.

In a blockchain storage system, accessing data takes constant time O(1) [2]. This is because every node has a hash ID. This means whenever someone needs to read from or write to a file, they need to present the computer with the hash ID, where the computer will find the file with a hash algorithm. Therefore, its drawbacks are its space complexity. Clients who choose not to centralize their data due to concerns have to keep in mind that there have to be available nodes for their data to be saved. Centralized Storage almost always has storage available due to consumers buying for their employees and servers to be maintained and managed very well. Decentralized Storage has to depend on people willingly giving up storage, so it can be used by other people.

### *B. Privacy*

Privacy is one of the most important human rights someone can have. Yet, within big corporations, your data is technically handled by their employees regularly. All it takes is a little negligence and your data could be subject to intrusion. Personal Data is the key to even accessing what you want to retrieve on their services. This can be seen as intrusion due to companies needing to know more about their clients to even do business with them. For example, to reset your password, you may need to answer security questions only you and the company knows you know. All of this information is ultimately saved in their servers, which can be subject to theft.

In a blockchain, this issue is resolved by having encrypted your data within each block created when in the system. The only way to access your block is with a key given to you once you have uploaded your content Through the utilization of the P2P network, nodes of

data can be saved on other devices on your network or even devices of other people. Even if one person's device is down, there is statistically always someone with a copy of that node available for use elsewhere.

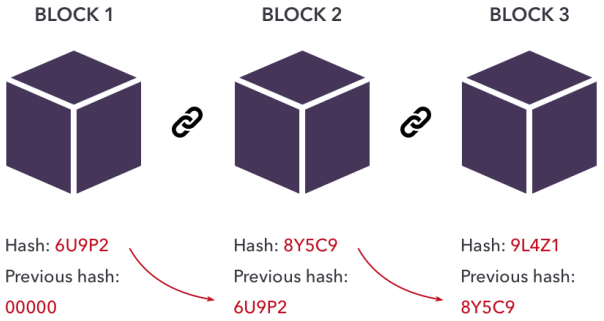
### C. *Security*

In a centralized storage system, security is controlled solely by the people who manage the system. All your data is contiguously stored in one place. Decentralized Storage, however, in decentralized storage data is stored in many steps. Whenever you want to save something within this system, your file will be broken into "shards" and each shard will become a node [2]. The node will have the contents of a hash ID, pointers to other nodes, and encrypted data of that specific shard.

Why is this method more secure than just storing it all contiguous in one place? In a P2P network, there will be several copies of the same node, hence there will be less of a chance you will not be able to get your data if One device in your network is down. In other words, there are multiple fail-safes to prevent you from losing your data. For example, if you went and saved the project you worked so hard on in Google Drive, and Google goes down for 2 hours, you will not be able to access that project until Google gets back online.

**Figure 1**

*Blockchain of a file*




*Note. The image shows three blocks chained together. Blockchain 101: The Simplest Guide You Will Ever Read*. (n.d.). https://www.velotio.com/engineering-blog/introduction-to-blockchain-and-how-bitcoin-works

### D. *Addressing*

The architecture of a system highly defines the behavior of saving and receiving data. On centralized servers, we locate what we want by giving a computer an address and having the computer retrieve whatever we are looking for. For example, websites are saved to servers, and whenever clients want to visit said website they type in a URL to do so. It just happens that if that server the website was on was currently down or has moved, the computer has no way of knowing how to find that website after it reached "NOT FOUND." This is a major flaw of centralized servers since they are dependent on data being exactly where they are concerning the consumer. This type of addressing is called location-based addressing. When it works, it takes constant time, otherwise, it will never reach where the client intends.

The addressing the decentralized system uses is called content-based addressing [2]. If you have a key to your data, and you know exactly what you want, you can tell your computer what you are looking for instead of where to find it. This has the computer convert whatever information given in a hashing algorithm, which outputs a hash ID. With this hash ID, it'll locate the parent block and return the whole chain by iterating through the blocks and representing the data to the screen.

## III. Current Uses

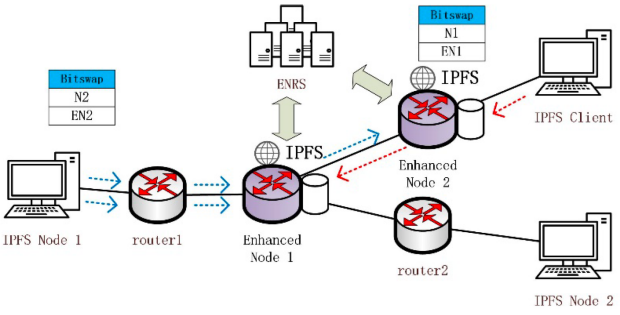
In recent times, decentralized storage is becoming more and more popular. This is due to the fact of censorship becoming a major issue for a lot of content developers. Services like the InterPlanetary File System (IPFS), Bitcoin, etc. have surfaced to solve the issues of not only censorship but also security.

### A. *IPFS*

IPFS contain nodes like how blockchain does, but the nodes contain IPFS Objects [5]. These objects help check the version of the program currently installed on it. Whenever it is changed, the ID for the object changes as well. This implementation helps notice changes to one's data and can quickly recognize it if it was not intended. This type of version checking is very similar to Git, in which we can retrieve a specific version of a program. Whenever a client updates an application in IPFS, it'll create a new block in the chain and assign it a new hash ID. If we alter a version of the application that was before the previous update, it'll create a branch. This system of storing helps us track updates and changes to an application. This in turn adds a layer of security to the system.

**Figure 2**

IPFS Branching




*Note. The image shows an IPFS object being updated twice from the same node.* Zeng, R., You, J., Li, Y., & Han, R. (2022). An ICN-Based IPFS High-Availability Architecture. Future Internet, 14(5), 122. https://doi.org/10.3390/fi14050122

### B. *Bitcoin*

Bitcoin is a recent system that plans to give more freedom to the economy. The strongest economies are those that are of gold-based currencies, these are the types of systems Bitcoin is trying to get overcome [6]. This service does not plan to be controlled by

governments are any system that could turn corrupt. It is based on the decentralized method. Bitcoin is similar to IPFS in terms of tracking versions. It is different from IPFS in terms that it helps validate transactions [6]. In regular central banks, there are usually lower transaction fees for payments, but with Bitcoin, this is not the case. The Bitcoin currency can be tracked by checking the most recent "block" within the blockchain. Whenever a transaction is made, it is transmitted to the Bitcoin network, and then transmitted back to all parties of the transaction. By having this system, both parties have the entire history of the transactions [6].

## IV. My Method

My method is to implement a decentralized storage system but with only devices from within the network. This method will still have the benefits of the security of not having everything saved to one point where it is vulnerable if found. If someone has multiple devices on their network that can save information, it is best to decentralize files downloaded over the internet.

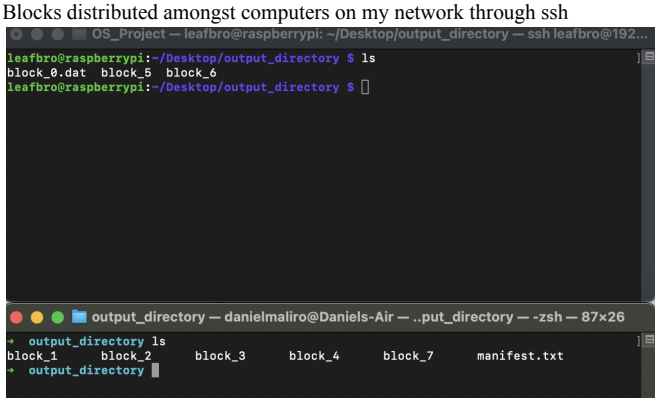
This method still includes having blocks created whenever the blockchain program is inputted with a file. That file will then be broken into block objects/structs where it is ultimately saved to devices within your current network. Each block object will contain the attributes: index, hash, *filename, *data, *next.

```
typedef struct block {
   int index;
   byte hash[HASH_SIZE];
   char *filename;
   byte *data;
   struct block *next;
} Block;
```

The filename is simply the name of the block when it is saved to the hard drive, the hash is the hash assigned to it based on its data. The data is up to 25KB per block, if an input file is more than 25KB, it'll create a new block and start writing to that block where it left off from the previous block. The *next attribute is what links everything together so it can be read.

```
while ((size = fread(data, 1, BLOCK_SIZE, fp)) > 0) {
      Block *block = new_block(data, size, num_blocks,
argv[2]);
      blocks[num_blocks++] = block;
   }
```

**Figure 3**
Blocks distributed amongst computers on my network through ssh




## V. Conclusions

Up to this day, big tech companies emphasize that they prioritize the security of the data of their clients. Though centralized storage is still the most widely used method of storage to date, decentralized storage stays true to keeping the responsibility of your data in your hands. We have to educate ourselves more on the management of our data outside our hands to understand the importance of managing it by ourselves. Bitcoin already is trying to turn away from the banks due to incidents like the breach of Capital One. Peer-to-peer networking is the next step in managing connections.

### References


[1] Novaes Neto, N., Madnick, S., de Paula, M. G., & Malara Borges, N. (2020). A case study of the capital one data breach. Stuart E. and Moraes G. de Paula, Anchises and Malara Borges, Natasha, A Case Study of the Capital One Data Breach (January 1, 2020).

[2] S. Wang, Y. Zhang and Y. Zhang, "A Blockchain-Based Framework for Data Sharing With Fine-Grained Access Control in Decentralized Storage Systems," in IEEE Access, vol. 6, pp. 38437-38450, 2018, doi: 10.1109/ACCESS.2018.2851611.

[3] Vorick, D., & Champine, L. (2014). Sia: Simple decentralized storage. Retrieved May, 8, 2018.

[4] Hugoson, M. Å. (2009). Centralized versus decentralized information systems: A historical flashback. In History of Nordic Computing 2: Second IFIP WG 9.7 Conference, HiNC2, Turku, Finland, August 21-23, 2007, Revised Selected Papers 2 (pp. 106-115). Springer Berlin Heidelberg.

[5] Benet, J. (2014). Ipfs-content addressed, versioned, p2p file system. arXiv preprint arXiv:1407.3561.

[6] Vranken, H. (2017). Sustainability of bitcoin and blockchains. Current opinion in environmental sustainability, 28, 1-9.