Decentralized Storage - Literature Review

Jonathan Healy

Decentralized storage is a concept that has roots in peer to peer file sharing technologies of the past. It has emerged as a response to modern cloud computing and naturally shares the ideas and philosophies that helped drive the early development of Bitcoin. The five decentralized storage solutions covered in this review share many things in common. The first and maybe most obvious is that each has adopted the idea of using a blockchain to secure a digital currency in order in incentivize their ecosystem.

The core logic behind decentralized storage is relatively simple. Users contribute hard drive space to store other Users files and are rewarded for doing so. The challenges involved with implementing this model however, are not so straightforward. One common approach is to implement a solution, like the one presented in the paper, “Sia: Simple Decentralized Storage” (1), where a blockchain is utilized to not just store payment history but also to track and activate smart contracts between data providers and subscribers. Data itself is not stored on a blockchain as doing so would be prohibitively expensive and the system would suffer from serious latency issues.

Another issue that is of concern in decentralized storage models is in auditing the network. If nodes are being rewarded for storing data or are under contract to do so, how can the system ensure that they have been? David Vorick and Luke Champine in their paper, “Sia: Simple Decentralized Storage” (1) take the position that contracts themselves should, “require the storage provider to prove, at regular intervals, that they are still storing their client’s data.” In the paper, “Storj: A Decentralized Cloud Storage Framework” (2) a similar idea is used based on probabilistic per-file audits called ‘proofs of retrievability’. Storj also introduces the idea of using a reputation system to rate storage providers.

Filecoin, presented by Protocol Labs in the paper, “Filecoin: A Decentralized Storage Network” (3), claims to introduce a novel proof-of-storage scheme called ‘proof-of-replication’. They also have developed a consensus model based on sequential ‘proofs-of-replication’, using storage as a measure of power called, ‘proof-of-spacetime’. The team behind Filecoin previously developed IPFS, Interplanetary Filesystem which is in use today, “serving billions of files across a global peer-to peer network.” Filecoin essentially acts as an incentive layer on top of IPFS. They present 2 verifiable markets, a storage market and a retrieval market. These markets run on top of a blockchain which is similar to Bitcoin except that storage providers will essentially be minting new tokens by providing storage resources. The chances that any one entity has to hash and create a new block (and receive the reward) will be proportional to the amount of storage they have being used in the network.

The need for auditing data on a network is important for many reasons. Nodes storing data may be unreliable and unfit to be participants in the network or they may be looking to find ways to game the system and receive profit for data that they are not sufficiently making available. If, after being audited, a node is deemed as being unreliable or incapable of hosting data, the network must redistribute the data it is holding to new storage nodes. This introduces a large amount of complexity. Contracts need to be revoked, other contracts need to be created, and pieces of data must find a new home and be accounted for.

Node churn, where nodes join and leave the network, is a related issue and one that has traditionally plagued peer-to-peer networks. Designing a system that rewards storage provider nodes for remaining online for long periods of time should help combat these concerns. This challenge is recognized in the Storj paper when they say that the system must, “detect when a storage node stops storing data correctly or goes offline and then repair the data it had to new nodes.” Repairing data is done with erasure code reconstruction from remaining pieces. Erasure code reconstruction is something that is talked about in all of the papers here. Other systems help mitigate data loss as well by having multiple copies of any data spread out across the network in different locations. In some systems a Client seeking data storage can specify how many copies of their data they want to have stored in the network and this can not only provide redundancy but also higher availability. Potentially data could be downloaded from multiple sources at the same time similar to BitTorrent which could allow the system to offer much higher download speeds.

All data being stored in a decentralized network should be encrypted. This can be done client side or by the network itself and encrypting data before storing it on the cloud may be something that should be looked at even when using centralized cloud services. Ideally encryption should be performed locally before data ever leaves a User's computer.

1. Vorick, David, and Luke Champine. “Sia: Simple decentralized storage.” *Nebulous Inc* (2014)
2. Storj Labs, Inc. “Storj: A Decentralized Cloud Storage Framework” (2018)
3. Benet, J., and N. Greco. “Filecoin: A decentralized storage network.” *Protoc. Labs* (2018).