Evaluating the Decentralisation of Filecoin

Barbara Guidi
Department of Computer Science,
University of Pisa
Pisa, Italy
guidi@di.unipi.it

Andrea Michienzi
Department of Computer Science,
University of Pisa
Pisa, Italy
andrea.michienzi@unipi.it

Laura Ricci
Department of Computer Science,
University of Pisa
Pisa, Italy
laura.ricci@unipi.it

ABSTRACT

Web3 is progressively decentralising Internet services with the introduction of the blockchain, from social networks to online trading. Filecoin is a Web3 storage marketplace implemented on top of the Interplanetary File System that aims at decentralising storage by rewarding users who pledge storage to the system. Clients form deals with storage miners, who store deals on the Filecoin blockchain for transparency and auditability. Storage miners can also obtain extra rewards when they mine a block. In this paper, we show that although Filecoin aims at complete decentralisation, the rewarding system acts as the opposite. We download a dataset composed of 47.5M messages and expose that most of the blocks are created by the same storage miners, who are also the ones that pledge more storage. An analysis of their identity shows that they belong to cloud companies, that have a lot of extra storage resources. The effect observed is due to the fact that storage miners are able to commit all extra storage to the system, even when there are no deals in them. Additionally, this effect is further intensified by recent updates to Filecoin, which let storage miners update committed storage to deal-containing storage effortlessly.

CCS CONCEPTS

• **Information systems** → **Storage architectures**; **Peer-to-peer retrieval**; • **Computer systems organization** → **Peer-to-peer architectures**; • **Networks** → **Peer-to-peer protocols**.

KEYWORDS

Decentralised storage, Filecoin, IPFS, Rewarding systems

1 INTRODUCTION

The advent of Web2 created an Internet where users are *prosumers*, i.e. users produce data that can be consumed by other users, breaking the well-defined distinction between the two roles. This type of Web created fertile ground for massive companies to monopolise the scenario and gain access to a massive amount of user-generated data that can be used for multiple goals, including large-scale society analysis, advertisement, influencing public opinion and others [4]. This is due to the fact that these companies adopt a centralised approach for data storage, and keep all the data in proprietary cloud storage systems, where knowledge can be extracted from data, not always in a transparent way [15]. However, with the advent of Web3, Internet services are being redesigned in a decentralised fashion to return data ownership to the respective creators, mainly through the usage of the blockchain. The blockchain is one of the possible implementations of a distributed ledger, where the information is organised in a linked list of blocks. It is managed by a Peer-to-Peer (P2P) network of nodes that follow the same consensus protocol, to ensure that the distributed ledger is shared, consistent and available [1].

Filecoin [13] is a project that embraces the revolution of Web3, providing a blockchain-based, decentralised storage marketplace implemented on top of the InterPlanetary File System (IPFS) [7]. The system identifies clients, who request storage through deals, and storage miners, who accept deals to offer storage space. When a deal is formed, a sector (Filecoin's storage unit) is *sealed* and a proof of replication [6] is generated as a result. Once a sector is sealed, the storage miner must prove through a proof of spacetime [11] that it is storing the sealed data every 30 minutes. To make sure the process is transparent, provable, and auditable, all the proofs are stored in the Filecoin blockchain. The blockchain is unique because it requires extreme scalability to let storage miners post their proofs. For this reason, the Filecoin blockchain is made of a chain of tipsets, where a tipset contains the blocks created at a certain round, and each block contains the set of messages (transactions). Every 30 seconds, a set of storage miners are elected to mine a new block following a Proof of Stake approach, were the stake is represented by the storage.

Although the system is designed to be decentralised to let anyone participate as in a P2P system, Filecoin lets storage miners take full advantage of all their storage, even if there are no active deals. Due to this, people who offer their storage through their personal devices are not able to compete with large companies that are able to pledge a lot of storage to the system. This creates a disparity not only regarding the possibility to form deals but also regarding the possibility to mine blocks for the blockchain. Indeed, by tying the amount of storage pledged to the system with the probability to mine a new block, the Filecoin network favours storage miners proportionally to the storage offered. Therefore, the rewarding system acts as a centralisation medium for both Filecoin and IPFS, as it prevents true decentralisation by discouraging small nodes to be competitive in the storage market.

In this paper, we provide an evaluation of the disparity created by this system. We download a dataset made of more than 100k tipsets from the Filecoin blockchain, which includes 500k blocks for a total of 47.5M messages. We analyse the messages downloaded from the blockchain and discover some important insights. To begin with, the block creation process is monopolised by a few miners, a result that is reflected by the fact that the same miners provide the vast majority of space to the system. On top of that, Filecoin lets storage miners commit their storage and make it count towards the chance to be elected as a block producer, even when the committed storage does not contain data from a deal. Lastly, Filecoin recently introduced a way for storage miners to upgrade their committed storage to a sector containing a deal with minimal effort, further sharpening the inequality of block creation possibilities.

The paper is structured as follows. Section 2 provides the an overview of decentralised storage networks. Section 3 provides the relevant concepts of Filecoin. Section 4 provides the result of our study, and finally, Section 5 concludes the paper pointing possible future works.

2 BACKGROUND

Decentralisation of file storage is a broad research topic which has great importance in many application fields. In its most general form, a decentralised storage application stores data in a decentralised network of computers that cooperate to store and make information available. With the advent of Web3, decentralised P2P file storage systems evolved as so to include a blockchain-based rewarding system, that incentivises storing data for other users. Notable examples in this direction are the Arweave[1], Sia[2], and Storj [3].

Arweave is a Decentralised Storage Network that connects people with extra storage space available with those who need more. The protocol allows to store data permanently, sustainably, on-chain, with a single upfront fee. Arweave's consensus mechanism is based on proof of access and proof of work and new blocks are created considering data from a randomly chosen previous block, thus creating a 'weave' of blocks, making it more scalable.

Sia calls file contract an agreement between a storage provider and their client. A file contract includes the hash of the data to be stored, the duration of the deal, a challenge frequency, and the fees payed to the provider. The challenge frequency is a Proof of Storage the provider must post on the Sia blockchain at regular time intervals to prove that it is storing the data. File contracts are managed by transactions stored on the blockchain for transparency reasons. The system also includes a basic reputation system so that client can thoughtfully choose the best hosts.

Storj is a decentralised cloud storage network where customers form deals with satellites, that are lightweight nodes of the protocol that only store metadata. Satellites leverage a network of storage nodes to store the client's data, acting as a medium in the protocol. Network participants earn STORJ tokens, that are Ethereum's ERC-20 tokens, in return for providing unused hard drive space and bandwidth to the network.

BitTorrent uses files, called torrents, that contain metadata of the content to be retrieved, and a list of trackers. A tracker is a special node that keeps track of other nodes connected to the network and shares their IP addresses with other BitTorrent clients, allowing them to connect to each other. Trackers can also be substituted by DHTs. BitTorrent recently launched BitTorrentToken (BTT), a cryptocurrency to reward participants to store files and supply them to the network. BTT is a TRC-10 (similar to an Ethereum's ERC-20) token deployed on top of the Tron blockchain.

The most important project in this direction is the **InterPlanetary File System** (IPFS), which implements a distributed file system over a P2P network [7]. An important feature of IPFS is that files are addressed by their content (a hash of the file) rather than their location (as in traditional file systems or HTTP). This helps with de-duplication and adds a degree of security to the system. References to which nodes are storing which files are stored in a Kademlia DHT [10], while the file distribution is implemented by BitSwap, a protocol similar to BitTorrent. The same creators of IPFS also launched a decentralised market storage called Filecoin. A qualitative comparison among the above listed decentralised P2P file storage systems is presented in [2].

3 THE FILECOIN BLOCKCHAIN

The Filecoin Blockchain is a distributed virtual machine that processes messages issued by storage miners, accounts for storage, achieves consensus, and maintains security in the Filecoin Protocol. Filecoin is a blockchain that regulates a decentralised storage market, where users ask *storage miners*, to store some data behind payment. Filecoin stores data in *sectors*, that have a size (amount of data to be stored) of 32GiB or 64GiB[3], and a lifetime (amount of time the data should be stored) from 6 to 18 months. Data storage and retrieval are provided through IPFS [7]. Deals are formed between users

[1] https://awebanalysis.com/img/whitepaper/pdf/arweave-whitepaper-document-awebanalysis.com.pdf

[2] https://blockchainlab.com/pdf/whitepaper3.pdf

[3] 1 GiB = 2^{30} B

and storage miners on the storage market and are recorded on the blockchain for auditability. When a deal reaches the end of its lifetime, the storage miner is paid for its service, otherwise, the miner pays a penalty.

Storage mining, the focal activity of Filecoin, consists of formalising a deal between a client and a storage miner on the blockchain. It is composed of **Proof of replication** (PoRep) [6] and **Proof of Spacetime** (PoSt) [11]. A PoRep is a proof that a storage miner has correctly generated a unique replica of some data provided by a user in a sector, when a deal is made. A replica is unique to the miner and to the time at which it was created so that a miner can replicate the same data multiple times. To provide a PoRep, sectors must be *sealed*. The raw data provided by a user is included in an (unsealed) sector, which is sealed by computing its Merkle tree [14] and encoding the whole sector .The identifier of the sealed sector is posted on the blockchain through a `PreCommitSector` message. At this time, they also lock a collateral, a small sum of cryptocurrency to prove their intention in storing the sector for all its lifetime. Storage miners prove through a SNARK [12] that sealing was completed and submit the result of the compression to the blockchain as a certification of the storage commitment through a `ProveCommitSector` blockchain message. Once the proof of a sealed sector is posted to the blockchain, storage miners must prove that they are still storing the sector through a `SubmitWindowedPost` blockchain message. A PoSt consists of a collection of proofs, called WindowPoSt, that a storage miner must submit over time to prove it is storing the sectors. Each sector must be proven every 30 minutes, and WindowPoSts can be submitted only for the current time slot. Each WindowPoSts consists of a SNARK, and a single proof can cover a partition, that is made of 2,349 sectors. When a sector naturally expires, the miner can claim the rewards and recover the collaterals. However, if the storage miner fails to provide all the WindowPoSts for a given sector, the sector is declared faulty, the storage miner will not be able to claim the rewards established during a deal, and loses the collateral.

To ensure storage miners work at their top potential, Filecoin introduces the concept of *committed capacity*. An available sector that contains no deal can be pledged to the network so that the storage miner can prove it has more storage available for deals. Adding committed capacity follows a process similar to sector sealing, and the sectors committed can be upgraded at any time to sealed sectors containing deals. Sectors containing data from deals or committed capacity grant the storage miner some storage power, that is used to decide which storage miners will create the next blocks. To encourage storage miners to store actual data in their sectors, Filecoin introduces the concept of sector quality. Indeed, if the sector contains data from deals with *verified clients*, it can be worth up to 10 times the storage power of a regular sector. To become verified, a client must be verified by a decentralised network and can post only a predetermined amount of deals. The quality of a sector also depends on the size of the sector, the length of the deal, and the time passed since the deal was made. A sector with higher quality gives more storage power and therefore grants more chances to mine new blocks for the blockchain.

The blockchain data structure grows through successive rounds of Expected Consensus [9], a leader election process in which a variable number of miners are elected to generate a block. As in all blockchains, including a new block in the chain will earn miners block rewards. The block is the main unit of the Filecoin blockchain and is composed of the block header and the list of signed messages inside the block.

Every 30 seconds (1 epoch), Expected Consensus elects zero or more leaders that will create just as many blocks for the epoch. Blocks from the same epoch are assembled into a tipset. However, due to connection latency, multiple tipset can be created for the same epoch. To mitigate the effects of possible forks, blocks are linked to tipsets, and each block should link to the tipset created in the previous epoch with the highest committed storage. The Filecoin virtual machine interpreter modifies its state by executing all messages in the new tipset, after discarding duplicate messages that may be included in more than one block.

4 DECENTRALISATION EVALUATION OF FILECOIN

The goal of this study is to evaluate the Filecoin storage market and understand if it lives up to its decentralised nature. To begin with, we obtained a dataset made of 103,680 tipsets, that include 504,294 blocks for a total of 47,540,219 messages. The dataset was obtained through Filfox[4], a public Filecoin blockchain explorer, and covers a 36 days long time span, from the 27th of January to the 3rd of March 2022. In the rest of this Section, we provide: an analysis of the miners and a comparison with Bitcoin and Ethereum; a study concerning the message senders and receivers; an analysis of the methods and gas used; an overview of the impact of sealed storage.

4.1 Miners

In Table 1, we report the addresses of the top 10 miners by the number of blocks created. The Table shows that a single miner is not able to monopolise the block creation process, indeed the top miners achieved to block a few thousand of blocks, which are less than 1% of all the blocks created in our time span. However, in some cases, we were able to associate some additional information to each miner id, i.e. a

[4] https://filfox.info/en

Miner id	Blocks	Nation	Tag
f0688165	4625	China	Metaverse Infrastructure
f0127595	4239	China	Metaverse Infrastructure
f0142720	4120	China	RRM-Athena
f049911	3670	China	MakerST
f0123261	3659	China	LDPOOL
f0135467	3222	China	RRM asia 20
f0134867	3155	China	DCPool-Sirius
f01137150	2819	-	-
f0131822	2816	-	-
f0154294	2785	China	MakerST

Table 1: Top 10 miners by blocks created.

label that says who owns the miner (Tag), and where the node is situated. In particular, all the miners for which we were able to gather the Tag and nation, are situated in China. This is due to the fact that the Chinese government banned the mining of Bitcoin in the country because of the high energy cost[5], encouraging the involved companies to find other business opportunities, who found in Filecoin a perfect scenario because of its reduced cost for mining a block, because the mining of a Filecoin block boils down to a Proof of Stake, where the storage committed is the stake. On top of that, we can see some important Tag, such as "LDPOOL" and "DCPool-Sirius", which are two mining pool companies, while many other Tags are linked to cloud storage companies. The fact that the most important miners are linked to cloud storage companies highlights that Filecoin is far from being concretely decentralised market storage because these companies are dominating the storage market and monopolising the mining operations. Big companies are investing in Filecoin, and they are able to run multiple nodes in the system, as testified by the fact that some Tags occur multiple times, easily cuts out nodes owned by private people from the possibility to mine new blocks. If we compare the top mining nodes in our dataset with the top ten nodes by storage space committed to Filecoin according to Filfox (see Table 2), we can see that the first 9 of them are matching, confirming our claim regarding a low level of decentralisation of Filecoin and the impossibility for small nodes to come into play in the storage market and mining process. The Table also shows the storage power of storage miners and the ratio of storage power owned by the storage miner with respect to all Filecoin. In short, we recall that storage power is a measure that indicates how much storage a storage miner has committed to Filecoin, and therefore indicates the probability that the miner will be elected as a block creator. The data reported in the Table shows the impressive amount of resources committed to Filecoin by these nodes, which a simple user can hardly match considering the infrastructure required to manage such massive storage space. It is also important to notice that only 10 storage miners hold almost 7% of all the storage power. This is a clear sign that block rewards will be mostly assigned to a few selected users, and therefore create a degree of centralisation in the Filecoin storage market and blockchain.

Miner id	Storage power	Ratio
f0688165	147.66 PiB	0.92%
f0127595	132.00 PiB	0.83%
f0142720	129.79 PiB	0.81%
f049911	116.76 PiB	0.73%
f0123261	115.11 PiB	0.72%
f0135467	106.34 PiB	0.66%
f0134867	98.95 PiB	0.62%
f0131822	89.52 PiB	0.56%
f0427688	84.48 PiB	0.53%
f01137150	84.42 PiB	0.53%

Table 2: Top 10 nodes by storage offered, obtained through Filfox on the 23rd of February 2022[6]

	Gini coeff.
Bitcoin block production 2019 [8]	0.7
Ethereum block production 2019 [8]	0.92
Filecoin block production 2022	0.589

Table 3: Comparison of the Gini coefficient in some blockchain-related environments.

We also provide a comparison of the Gini coefficient [5] of the Bitcoin and Ethereum block production with the Filecoin block production obtained from our dataset. As we can see from Table 3, Filecoin block production achieves a high value of the Gini coefficient, although not as high as in other blockchain systems. The values reported in the Table indicate that the Filecoin's block creation process is mostly dedicated to a selected chaste of storage miners, adding a degree of centralisation to the system.

4.2 Message senders and receivers

We now put our attention on the messages sent on the Filecoin network, focusing on the senders and receivers. We recall that receivers must be mining nodes in case a method

[5] https://www.bloomberg.com/news/articles/2021-09-26/chinese-regulators-are-serious-about-banning-crypto-this-time

[6] https://filfox.info/en/ranks/power

of the Filecoin VM should be invoked. We show in Figures 1 and 2 the distribution of the top addresses by the number of sent and received messages respectively. The figures show that the two sets have no common address, however, by inspecting closely their addresses, we were able to discover 5 addresses that send their messages almost always to the same miner. This activity can be ascribed to the fact that each pair of sender and miner are owned by the same entity, that tries to make their miner work as much as possible so as to increase the amount of storage offered by the miner and increase its storage power, and therefore the chances it will mine a block. The fact that none of the miners who received more messages appear in the list of the top miners per number of mined blocks highlights the fact that there is some sort of way for miners to specialise their behaviour, based on the resource at their disposal.

Sender id	Miner id
f01731277	f01731371
f01716450	f01716466
f01687967	f01688066
f01245941	f01245980
f01652949	f01653068

Table 4: Sender and associated miner addresses

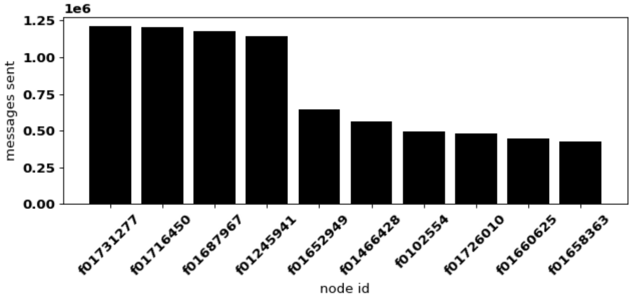

Figure 1: Top ten accounts by messages sent.

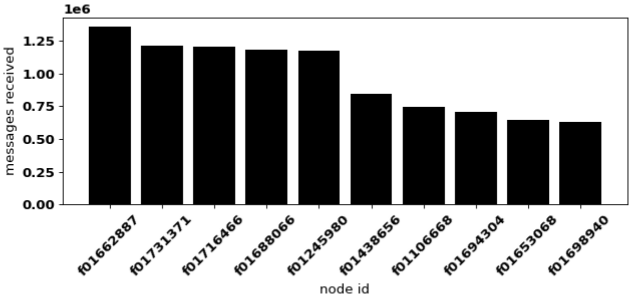

Figure 2: Top ten storage miners by messages received.

Name	Frequency	Gas used
ProveCommitSector	45%	74%
PreCommitSector	43%	21%
SubmitWindowedPost	10%	4%
other	2%	1%

Table 5: Frequency and gas used by methods

4.3 Methods and Gas used

In this Section, we analyse more in detail what is the method call embedded in a message sent, as so to understand what kind of activity is the most popular on Filecoin. As shown in Table 5, ProveCommitSector and PreCommitSector are the most used methods with a very similar number of occurrences, followed by SubmitWindowedPost, while all the other methods combined only make for 2%.

The first two methods are used within the Proof of Replication respectively to seal a sector and to prove that the data is being stored in the sector. The third is used by the Proof of Spacetime to prove that data contained in a sector has been stored continuously. The occurrences of the first two methods are similar, according to the standard procedure to store data in Filecoin and happen more frequently because it is the way for miners to store more data and therefore have more chances to mine new blocks. The number of occurrences of SubmitWindowedPoSt is much less because with a single call one can test a whole partition (2,349 sectors).

For what concerns gas usage, we can see a different situation. Indeed, the method ProveCommitSector consumes much more gas because of the many internal calls and the numerous steps required to seal a sector.

4.4 Sealed storage

Figure 3 shows the amount of new storage sealed in sectors during our observed period. As we can see, there is a significant and steady daily increase of storage in Filecoin, either it being committed or as a result of deals. In particular, we see that in the very first few days, more than 25 Pebibyte (Pib)[7] were added daily, most of which in the form of 32GiB sectors. This initial expansion phase can be ascribed to the fact that storage miners were benefiting from an increased number of deals from verified clients (+4.8% in the previous week). Sealing deals with verified clients is very profitable, indeed storage containing such deals are worth up to 10 times in storage power, which increases the chances for storage miners to be selected as block creator. On the other hand,

[7] 1 Pib ≈ 1,126,126 GB

the newly added storage saw a slight decrease, in particular towards the end of February.

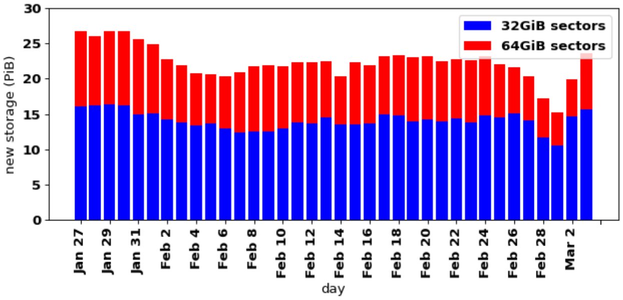

Figure 3: Storage sealed in sectors.

This effect could be an effect of an upcoming update of the protocol (version 15[8]). This update introduced two important novelties. To begin with, verified clients can now be stripped of their roles in case they adopt some incorrect behaviour, and can now see their deal limit removed. On a further note, the update introduced the possibility to update the sectors with committed capacity into sectors containing actual data, without needing to discard the old sector and then create and seal a new one. This is a time-consuming process, in particular considering the fact that Filecoin has more than 16 Eib[9] available storage, and that 90% of it is in the form of committed storage. So, to prevent an enormous amount of computation to rebuild all the involved sectors, the new update lets storage miners simply upgrade the sectors. This update clearly favours cloud computing companies that are encouraged more than ever to pledge big portions of storage to the system, even if it is only as committed capacity, as the sectors can be upgraded effortlessly.

5 CONCLUSION

In this paper, we studied the decentralisation of Filecoin, a Web3 fully decentralised storage market implemented on top of IPFS. In Filecoin, clients rent storage space from storage miners through deals stored on the blockchain. Storage miners are also eligible for mining blocks for the blockchain proportionally to the storage pledged to the system. However, this incentive mechanism favours storage miners with massive storage capacity, such as cloud computing companies, provoking the reduction of decentralisation of not only the storage market of Filecoin and the block creation process of its blockchain but also of IPFS.

We provide an evaluation of the decentralisation of Filecoin through a dataset that includes approximately 47M messages. Our analyses showcase that the mining process is monopolised by a few storage miners who are also the ones with the highest storage committed to Filecoin. The main motivation behind this result is the fact that the rewarding system lets nodes commit their storage space even if it does not contain deals, while still benefitting from an increased chance to mine new blocks. This effect was further amplified by a recent update that lets storage miners easily accommodate deals in committed storage effortlessly.

In future works, we plan to revise the reward system to encourage small storage miners to join the Filecoin market to increase the decentralisation of the system. Since the blockchain requires massive resources, we also plan to experiment with layer-2 solutions to reduce the system requirements to run a Filecoin node, so that the whole system is accessible to the largest pool of users possible.

REFERENCES

[1] Andreas M Antonopoulos. 2014. *Mastering Bitcoin: unlocking digital cryptocurrencies.* " O'Reilly Media, Inc.".

[2] Erik Daniel and Florian Tschorsch. 2021. IPFS and Friends: A Qualitative Comparison of Next Generation Peer-to-Peer Data Networks. *CoRR* abs/2102.12737 (2021).

[3] Sammy de Figueiredo, Akash Madhusudan, Vincent Reniers, Svetla Nikova, and Bart Preneel. 2021. Exploring the storj network: A security analysis. In *Proceedings of the 36th Annual ACM SAC*. 257–264.

[4] Yuefeng Du, Huayi Duan, Anxin Zhou, Cong Wang, Man Ho Au, and Qian Wang. 2020. Towards privacy-assured and lightweight on-chain auditing of decentralized storage. In *IEEE 40th ICDCS*. 201–211.

[5] Frank A Farris. 2010. The Gini index and measures of inequality. *The American Mathematical Monthly* 117, 10 (2010), 851–864.

[6] Ben Fisch. 2018. Poreps: Proofs of space on useful data. *Cryptology ePrint Archive* (2018).

[7] Barbara Guidi, Andrea Michienzi, and Laura Ricci. 2021. Data persistence in decentralized social applications: The ipfs approach. In *2021 IEEE 18th CCNC*. 1–4.

[8] Qinwei Lin, Chao Li, Xifeng Zhao, and Xianhai Chen. 2021. Measuring decentralization in bitcoin and ethereum using multiple metrics and granularities. In *2021 IEEE 37th ICDEW*. 80–87.

[9] Zhixin Liu, Jing Han, and Xiaoming Hu. 2011. The proportion of leaders needed for the expected consensus. *Automatica* 47, 12 (2011), 2697–2703.

[10] Petar Maymounkov and David Mazieres. 2002. Kademlia: A peer-to-peer information system based on the xor metric. In *International Workshop on Peer-to-Peer Systems*. 53–65.

[11] Tal Moran and Ilan Orlov. 2019. Simple proofs of space-time and rational proofs of storage. In *Annual International Cryptology Conference*. 381–409.

[12] Maksym Petkus. 2019. Why and how zk-snark works. *arXiv preprint arXiv:1906.07221* (2019).

[13] Yiannis Psaras and David Dias. 2020. The interplanetary file system and the filecoin network. In *2020 50th Annual IEEE-IFIP DSN-S*. 80–80.

[14] Michael Szydlo. 2004. Merkle tree traversal in log space and time. In *Eurocrypt*. 541–554.

[15] Mirko Zichichi, Stefano Ferretti, and Gabriele D'Angelo. 2020. On the efficiency of decentralized file storage for personal information management systems. In *2020 IEEE ISCC*. 1–6.

[8] https://github.com/filecoin-project/lotus/discussions/7898

[9] https://dashboard.starboard.ventures/dashboard