An Extensive Study of Decentralized Storage Networks Driven by Blockchain

Shweta Babu Prasad[1], Ashok Kumar A R[1], Rajini V Honnungar[2]

Department of Computer Science, R V College of Engineering, Bengaluru, India[1]
Department of Electronics & Communication, RNS Institute of Technology, Bengaluru, India[2]

Corresponding author: Shweta Babu Prasad, Email: shwetababup@rvce.edu.in

The concept of blockchains, a novel method for building distributed networks, was first introduced in 2008. Allowing users to send their files safely over a decentralized network with end-to-end encryption eliminates the risk of data loss associated with centralized data control. A security, integrity, and privacy flaw needs to be fixed even though numerous studies in this particular field have focused mostly on storage capacity and efficiency. This paper first provides an overview of blockchain-based storage systems and explain show they operate. It then compares these systems to cloud-based storage networks and surveys the different decentralized storage networks that are currently on the market, including SIA, File coin, and Storj. The benefits and drawbacks of blockchain-based storage are then covered. In conclusion, we look at the security issues with decentralized storage networks and consider future research directions and possible fixes.

Keywords: Blockchain, decentralized network, smart contract, IPFS, Storj.

in: *Advancements in Communication and Systems*. Ed. by Ashish Kumar Tripathi and Vivek Shrivastava. Computing and Intelligent Systems, SCRS, India., 2023, pp. 457–470. DOI: https://doi.org/10.56155/978-81-955020-7-3-40

1. Introduction

Blockchain technology advancements in recent years have made us re-evaluate our understanding of the Internet as a network of centralized service providers. Several blockchain networks have demonstrated the value of decentralized ledgers. Anyone can create worthwhile services without centralized management thanks to platforms like these decentralized networks. Applications for blockchain technology are numerous and include social networking, supplychain management, and financial transactions. Computers, cell phones, and cameras not only generate a lot of data every day, but they also need an increasing amount of storage for the data [1]. To address this need, a cloud storage system was developed. The Cloud, which offers storage services, is a cooperative system made up of numerous devices, numerous applications, and numerous service types. Compared to cloud storage, local storage is more expensive, less dependable, and more prone to data loss. The term "cloud storage" describes the placement of user data on servers that are secured and overseen by a third party. Instead of being kept on the owner's hardware, the data is kept in the memory of distant devices. Cloud computing is a major advancement in computing, despite the numerous security and availability issues it raises. The main issue with cloud storage is that there is no visibility or control over stored data. Users may not be aware of how their data is handled, stored, or compromised. Users and businesses need to have greater trust in one another. Due to the lack of a formal contract between service providers and users, compensation claims cannot be made. Customers also need to know if their data is sold or duplicated. Distributed storage systems are now frequently using blockchain technology. File coin offers a completely decentralized network of storage services to users and storage miners, powered by IPFS (Interplanetary File System). The miner offers its services with the ability to view matching quotes to start transactions. A space-time certificate and copying proof are used to guarantee data integrity [2]. Figure 1 shows the creation of IPFS CID using a Merkle tree. The File coin protocol consists of order books, integrity challenge answers, and blockchain records for token transactions. Smart contracts that facilitate document exchange can be created by consumers and storage providers on the Siacoin network. Customers are required by contract to submit their data storage certificates within the designated certification window. Smart contracts pay the customer's storage provider automatically if the proof is legitimate. Based on the Storj network, Storj is a peer-to-peer cloud storage service. Files are encrypted by clients before being sent to the network. Files are shielded from unauthorized access by encrypting individual blocks in distributed storage, direct encryption can safeguard data confidentiality [3]. Customers pay storage fees once their data provider certifies that the data is recoverable. According to a recent analysis, a large chunk of transactions are written more than an hour after release of transaction. Solutions for distributed storage that have issues with transaction latency are not competitively advantageous. It also demonstrates how crucial it is to update middlemen and the system agreement to maintain the system's viability. A hard fork is unavoidable because the block created by new protocol node is invalid. Ethereum experienced a hard fork, dividing the population in half. Two new currencies are created during the Ethereum hard fork process: ETH, and ETC Currency. The decentralized system has certain drawbacks, such as slow updating and maintenance challenges. Regarding distributed networks, [4] discusses several definitions. One of their primary features is the sharing of resources, including content, storage, CPU power etc. There are many benefits to a distributed file system, such as fault tolerance, availability, scalability, and performance. The aforementioned advantages can only be realized by managing multiple servers and carrying out user application tasks. In distributed storage file systems, data replication lowers data loss and boosts data availability. Although quick and simple, this method has some drawbacks, like high storage overhead. To ensure effective replication, files must be distributed across various domains in a way that prevents multiple data failures [5]. Another technique that reduces computation complexity and addresses the issue of large overheads is erasure code. Being based on blockchain technology, distributed file storage systems do not require a centralized entity to maintain network control. These storage techniques therefore have a higher level of security than other kinds [6]. Despite recent improvements to security solutions, the built-in security laws in the Cloud cannot be fully fixed. Users on decentralized blockchain storage networks have the option to rent out their unused space to other users in need [3]. The topic of our survey is blockchain technology's application to storage networks. The supplier will get crypto currency in exchange for renting out this storage. Any client can access free storage from the system if they require it [4]. All of the data about storage availability, client-provider

contracts, and free storage with each provider is kept in a decentralized ledger. With little to no central control, this technique can create an autonomous storage network.

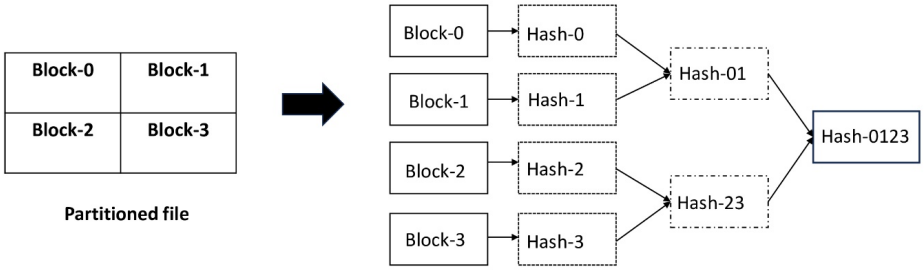

Figure 1. Creation of IPFS CID using Merkle Tree

This paper is organized into three sections. The paper's first section outlines a thorough literature search. The second section summarizes previous research on important subfields and subdomains of decentralized storage systems and provides implementation case studies. On the other hand, the third section looks at the benefits, drawbacks, restrictions, and possible issues related to decentralized storage. A sizable research gap is present concerning the security adoption in decentralized storage systems, even after taking into account the novelty and contribution of this survey. Numerous studies have been done in this area, but the majority concentrate on topics other than security and privacy. After a thorough comparison of several storage systems, the paper discusses potential security concerns with decentralized storage systems, along with potential fixes and constraints.

2. History and Associated Work

Scholars in [6] have examined several prior blockchain-based studies. According to research in this field, a large percentage of all research done on IoT is focused on data storage and sharing. The IoT and data storage are the two most talked-about blockchain topics. Blockchains, both public and private, seek to address the two primary issues with the current system: data manipulation and single points of failure. Blockchain-based storage networks offer a more secure and efficient means of storing data than customary centralized storage systems that rely on centralized servers. User data, system-related information, and personal information about users are some of the contents that are saved. The theories that have been explored to enhance current blockchain-based storage networks and leverage blockchain technology to leverage centralized systems to improve current ones are examined in the section that follows. By storing hashes of zone files, the blockchain-based decentralized Domain Name System (DNS) suggested in [7] can be used to stop data misuse. Additionally, it has several parsing nodes in case one of them fails to prevent the system from collapsing. The authors of [8] have proposed comprehensive reporting mechanism called PingER (Ping End-to-End Reporting). Actual files are kept off-chain by Distributed Hash Tables (DHT), and each file's metadata is kept on a private blockchain.

2.1 Literature Survey

This system eliminates the centralized party to measure Internet performance globally. A keyword search service can be used to solve search problems in storage blockchains [8]. A node (storage provider) will receive the encrypted data after it has been encrypted and sent over the system's network for storage. Even though keywords are kept on the blockchain, nodes still give each other permission. As such, the blockchain can be searched by permissioned nodes and data owners. A system for stopping data fraud has been suggested to be put into place [9]. The provenance of data is tracked by a package called Data Provenance, which contains metadata about the data's acquisition, ownership, and

transformation. It will be impossible for data to be maliciously altered as long as the majority of participants are reliable [10]. The privacy issues raised by conventional cloud storage systems have been addressed in part by the proposal of attribute-based encryption (ABE). Consequently, SPOF can occur in traditional cloud storage systems. Blockchain technology, according to [2], offers a single point of failure for decentralized storage. This model can achieve decentralized storage thanks to ABE technology, Ethereum blockchains, and inter planetary file systems. This platform also has a keyword search function that permits the data's owner to determine who must have access to the encrypted information. To upload digital data to a private blockchain, Block House, a private blockchain, has been developed. Private blockchains are used by blockchain-based storage networks to recoup costs associated with idle hardware. At regular intervals, the redundancy and availability of the data are verified. In this network, two intelligent contracts manage payment, logging, and storage security. In this project, the PoR algorithm is used to reach a consensus. Small and medium-sized businesses have the option to retain vital data on their network rather than in the cloud. Data distribution systems in organizations were looked at, per a study done in [1].Due to the requirement for data extraction, there is a worry that these systems could cause a mishap. Artificial intelligence and blockchain technology were used to create a decentralized data storage system. Artificial intelligenceis combined with blockchain to comprehend, generate, and retrieve knowledge. You can save money and safeguard your data simultaneously with blockchain. The Internet of Things (IoT) is expanding daily, and so is the volume of data it produces. There are trust concerns because third-party storage spaces hold a lot of important data. To address this issue,[11] suggests putting in place smart contracts, encryption, consensus algorithms, and blockchain-based multi-center storage systems. Objects with artificial intelligence (AI) are those who comprehend their surroundings and make choices that improve their chances of accomplishing predetermined objectives.

2.2 Cloud-Based Storage Structure

Since cloud storage offers many benefits to businesses when used properly, its use has become standard across corporate and industry sectors. Hard drives are essentially where cloud data is stored, much like how data is handled in general [21]. Cloud data is kept on servers of major corporations rather than on personal devices. Through the internet, the user can access this data. Including storage in current infrastructure is required to meet the growing demand for storage due to the volume of digital content available online. It will be necessary to buy pricey servers to accommodate this, which are difficult to maintain and require expensive configuration. Data migration to fail-over server is also expensive. Meeting the ever-increasing demand for data storage requires a substantial undertaking and a large budget Figure 2 shows the hierarchy in cloud-based storage where users are placed at the lowest level, and next, come the servers at the data centre at the highest level.

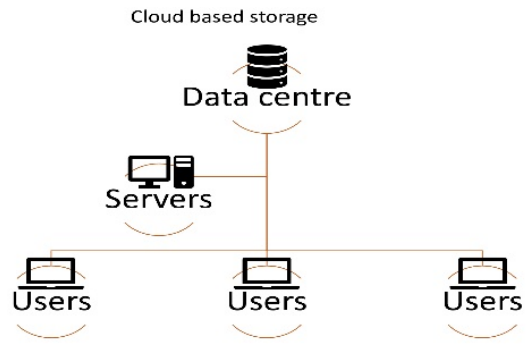

Figure 2. Hierarchy in cloud-based storage

2.3 Blockchain-Based Storage Structure

Peer-to-peer transactions are recorded on public ledgers called blockchains. Networks can scale and function without a central server thanks to peer-to-peer architecture, even in the event of an unpredictable node population during a network failure. Since blockchains store the complete history of transactions, it is very difficult to change them. Genesis blocks are the first building blocks that are parentless [12]. A mining process is used to verify transactions by resolving a computationally challenging puzzle and locating a unique nonce. A new block can only be made by blockchain users voting on a common group. Figure 3 shows the Merkle Tree structure in blockchain transactions.

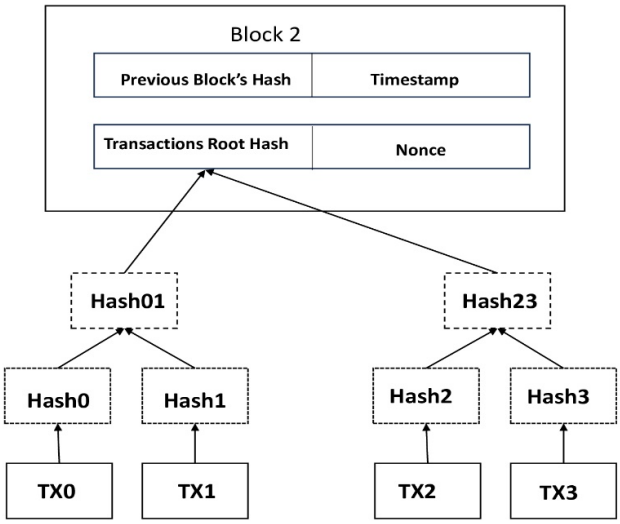

Figure 3. Blockchain Transactions in a Merkle Tree

2.4 Decentralised Storage

Although blockchain technology is not new, there have been advancements in decentralized storage networks in recent times. Numerous attempts have been made in the industry to integrate blockchain concepts, such as filecoin.io and Storj.io, to create a decentralized storage network. For instance, File coin seeks to facilitate blockchain data storage by miners through the use of a cutting-edge idea called Proof-of-Spacetime. In the blockchain system, this "Proof of Spacetime" takes the place of the traditional "Proof of Work" [6]. Without wasting time on calculations, new blocks can be mined rapidly. Rather, they extract blocks quickly by storing information in the network. Native tokens are utilized in transactions between providers and clients. By creating an incentive layer atop the blockchain and utilizing native tokens, one can add another layer of incentive functionality. Information regarding storage capacity and contracts between suppliers and customers can also be kept in a smart contract.

2.5 What is the Need for Decentralization?

Both individual consumers and major corporations have moved their data to centralized servers due to the ease of use of cloud storage. Giants in the tech industry such as Amazon, Microsoft, IBM, and Google now control most of the large data centers that have arisen as a result of economies of scale [7]. Although companies compete to offer users a wide range of service providers, the kind of offerings is frequently seen as a could be censored or improper use of personal information. Experts also note that

as cloud storage is shifting to the cloud, there will be a 71% increase in legitimate data breachesbetweennowand2020.
Numerous approaches are being investigated to use decentralized storage networks to upend the current cloud market. These networks will probably function initially as open markets with free markets. Anyone can join the network in this way, and data is replicated across several nodes to avoid relying on a single point of failure. Additionally, a natural companion to blockchain integration is public-key cryptography. Before data is stored, a host typically encrypts it, limiting who can decrypt it to the original owner and any parties they have consented to share it with. The procedure renders all data acquired via a hack worthless to a hacker and increases the resistance of these services to censorship and manipulation [8].

2.5.1 Elevated Values of Speed

P2P systems depend on P2P technology instead of centralized storage. During periods of high traffic, data transmission via the central server does not occur. The fact that multiple copies of the data are kept in various places speeds up downloads.

2.5.2 Distribution of Load

Hosts can store data locally to prevent frequent requests to the server. Both the load on the server and the amount of network traffic are reduced. The server can streamline data and reduce blockages in the main network.

2.5.3 Honest Valuation

As decentralized storage systems reach millions of nodes, perfect competition results. Nodes cannot charge more on their own. As a result, all nodes have the same prices. This market makes sure that only nodes of the highest caliber endure and contend.

2.5.4 Improved Security and Privacy

The main benefit of decentralized data storage systems is the high degree of security they offer. Shared data is encrypted copies of the original data that are shared amongst each other, and is divided into smaller portions using hashes or public-private keys. By protecting the data from nefarious actors, the entire process is secure. Furthermore, unlike centralized systems, no owner information is present in any stored data.

2.6 Benefits of A Decentralized Repository

P2P storage is being preferred as an alternative to centralized storage. These are some benefits that a decentralized storage system offers.

2.6.1 Increased Credibility

The decentralized network uses multiple hosts to store and distribute data. The data is stored in redundant copies, removing the possibility of a SPOF. There will be backup copies accessible in the event of a hardware malfunction or loss. Every piece of shared data is also given a unique hash value [9]. Data security is increased by including this additional layer of security.

2.6.2 Reduced Prices

Hardware and storage costs are greatly decreased with a decentralized data storage system. Decentralization lowers the requirements for machine performance, which lessens the need to spend a lot of money on high-performance hardware and software. Furthermore, the decentralized network can store data on millions of nodes. As a result, there is substantially more storage space available. By continuously using all of the available empty storage space, this system minimizes waste and eliminates the need to purchase new storage. The total cost of storage is much less than that of centralized cloud storage. Table 1 below provides a brief comparison of decentralized and centralized storage networks.

Table 1. Comparison of decentralized and centralized storage networks.

PROPERTIES	CENTRALIZED	DECENTRALIZED
SPOF	Inaccessible due to SPOF	Comprise of backup and redundant servers to mitigate risks
Scalability	Incur additional costs and hardware upgrades and the scaling process is complex	Scaling is seamless and can be done by adding nodes to the network
Database failure	Affects all users	This does not affect all users, some databases can still be accessed
Data consistency	Provides the user with a complete data view	Does not affect all users
Ease of updating data	Easier to maintain and update as there is only a single database	Managing and updating is tasking as multiple databases are involved

2.7 P2P Networks

Clients and servers are two roles that P2P nodes can play. File sharing is the goal of P2P networks, also known as decentralized public storage [10]. By using its search tools, users can locate and download desired files from other computers. A specific network offers tools and methods for this purpose. Upon installation of a client program on their PC, users are granted access to their files within a peer-to-peer (P2P) network. They can also search for and download files. When members of a network download a file simultaneously from multiple sources, the downloaded portions of the file become instant sources for other users. BitTorrent is a well-known example of a P2P network that guarantees high bandwidth.

P2P networks typically link nodes from various administrative domains. P2P networks are dynamic, meaning users can join and exit them frequently. P2P nodes that coincide with Internet nodes and store data about multiple other nodes form a virtual overlay network on top of the Internet. Every link in the overlay P2P network corresponds to a physical link in the core network. Queries and answers in a P2P storage system need to be routed appropriately, and data needs to be fault-tolerantly stored and effectively searched for. To satisfy these needs, various infrastructure types and algorithms are developed. P2P networks can be categorized using overlay network topologies, data search methods, and distribution control techniques.

It is a common misconception that P2P networks are fully decentralized; however, some may be more centralized than others. The network's central resource register and other data are stored on an isolated P2P network with a single central server. Users on the network can find desired files by requesting their address from the central registry server. These peer-to-peer networks are ill-scaled, resulting in a single point of failure. Both fully decentralized and hybrid systems fall under this category, depending on their features. Nodes in the network each have a distinct function. Under fully decentralized systems (e.g., Gnutella and Chord), nodes differ from one another. In hybrid networks, specific nodes help other regular peers process search requests. We refer to these nodes as super peers or dominating nodes. In peer-to-peer networks, there is often heterogeneity in terms of computing power, stability, and net quality. A fully decentralized system can't take advantage of the heterogeneity of a hybrid system.

Caching and dynamic indexing of files in smaller regions of overlay network is the responsibility given to the super peers. They take on the role of proxy servers, indexing and conducting searches on behalf of the regular nodes that are linked to them. All queries are therefore sent to super peers first [11]. A single point of failure and bottlenecks can be prevented by carefully selecting dominating nodes. According to their structure, the decentralized P2P networks can be classified as either structured or unstructured. This occurs automatically the majority of the time. Structured P2P networks are well-defined in terms of their architecture and data allocation.

Through the efficient routing of queries to a node that contains content and the assurance of data correspondence, a distributed hash table (DHT) makes content identification easier. and whereabouts.

These networks allow for the high scalability of systems. Despite offering effective message routing in a medium having fixed number of nodes, their drawback lies in the intricate network management architecture. P2P networks have unstructured network topologies and unregulated data storage locations, with no set of rules governing them. Queue flooding and route indexing are two straightforward search mechanisms that flood queries to find required data. Much more difficult to handle problems with unstructured networks including availability, reliability, and scalability.

Conversely, networks with fluctuating nodes are more appropriate for unstructured systems. Data search/allocation and overlay topology are probabilistic processes that rely on specific presumptions. Variations in data distribution and retrieval properties can be obtained by combining the two properties. While the locations of the data are precisely determined, the overlays are defined using a probability approach. Networks that are in between structured and unstructured are known as weakly structured. There are additional characteristics that can be used to classify P2P systems. The existence of a hierarchy, for example, can be used to distinguish between hierarchical and non-hierarchical overlay networks. Overlay networks are generally non-hierarchical and flat in fully decentralized systems.

Every hybrid system and some fully decentralized systems are hierarchical. Non-hierarchical systems inherently require stability and load balancing. A network's hierarchical structure makes use of the heterogeneity of its nodes to enhance performance, routing, and scalability.P2P networks that are decentralized can also be categorized according to various attributes. The illustration in Figure 4 provides a sample peer-to-peer network.

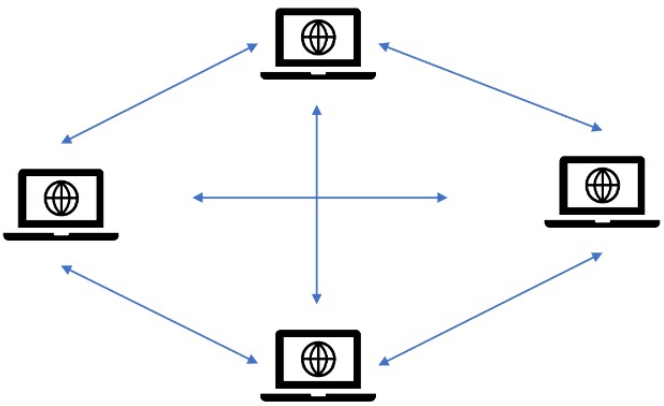

Figure 4. Peer-to-peer network

Blockchain technology is used in many different fields. Known researchers gave a brief overview of their studies on several decentralized storage-related topics, including the technology utilized for decentralization and issues and their resolution in decentralized storage. Some addressed issues and challenges for the future. The technology offers independent financial settlement, audit, and reconciliation mechanisms with increased transparency through fraud prevention, which has the potential to completely transform storage networks.

3. Storage Repository Dependent on Blockchain

Multiple storage repositories are present that are dependent on blockchain. A few of them are illustrated below:

3.1 Storj

Based on P2P and remote technologies, Storj is a freely available cloud storage network. A combination of decentralized and centralized architectures has been used in the design of the hybrid network Storj. Since content on the web is divided and shared among numerous peers, it is thought of as decentralized from a storage standpoint. For communication control, Storj dependson centralized servers [12].Peer storage nodes can store encrypted files and segments of them, and a centralized server handles user authentication and exchanges between them. The Storj network is made up of multiple units. Among these are a bridge, a renter, or a provider. Users can rent space on the Storj network. The Storj Client application lets users communicate with the network by allowing them to upload and download files. To communicate with the network, the renter must first engage with the bridge. After that conversation, a bridge allows the renter to send and receive files from providers. The focal point of the network is that bridge.

Except for file transfers between renters and providers, the bridge communicated with every component of the network and delegated all communication. For both providers and renters, it serves as a gateway to the network. The bridge additionally monitors the network's health regularly by keeping an eye on all connected providers and tenants. Network users who offer storage are known as providers. They must first ask the bridge for permission to join the network. Their network integration requires approval from a bridge. After providers join, tenants can contract with them for drive space, enabling them to set up storage agreements. A peer-to-peer cloud storage provider requires several steps to upload a file. A file cannot be handled until a contract has been established between the renter and the provider. After being stored on the bridge and the required contracts have been completed, the files are queued for upload [14].

During the process, the renter splits the file into multiple shards and encrypts it. The contract specifies how the shards are divided up among the providers after they are created. Redundant copies of the shards are made and dispersed as a backup mechanism in case a provider loses, destroys a share, or experiences a service outage when a renter needs to access shards. Renters get in touch with a bridge to request files from providers when they wish to download them. Initially, the bridge ascertains whether it can reconstruct the file using the available shards, enabling the renter to download it. The bridge alerts the supplier to start providing the renter with the shards if a file can be rebuilt. The shards combine into a single file and decrypt once the renter has all the shards required to restore the file. The file is now located and saved on the renter's computer, allowing the bridge to audit the transaction. Figure 5 shows the Storj architecture and Figure 6 shows the Storj data storage sequence.

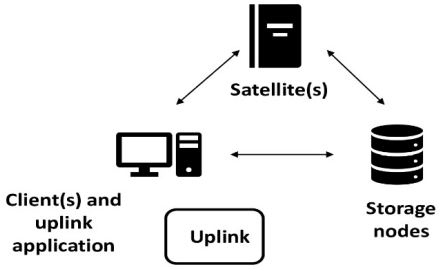

Figure 5. Storj Architecture

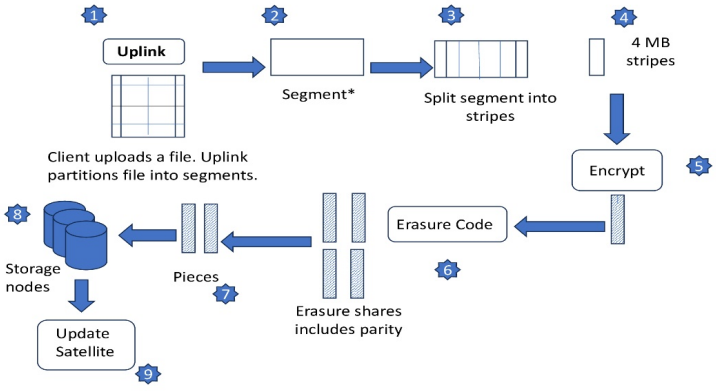

Figure 6. Storj data storage sequence

3.2 Filecoin

In the P2P Filecoin network, users and storage providers are essential components. Computers with internet access and space are available for rent by miners in the Filecoin Network. By providing storage to the network in return for FILs, miners are the storage providers. Users are currently looking to buy storage from storage providers that are using the Filecoin protocol. Keeping important data on someone else's computer might seem risky at first. Data is broken down by Filecoin before storage to prevent hackers from accessing data stored on its network. If a malicious actor tried to access a file on the Filecoin network, they would consequently only see random bits of data. A contract is an agreement that exists between a user and a storage provider. Not to be overlooked are two categories of Filecoin deals: storage deals and retrieval deals. The terms "storage deal" and "retrieval deal" refer to the processes by which a miner receives data from a client and stores it on the network.

After a storage contract is signed, miners have ongoing obligations to demonstrate that they are providing excellent customer service by preserving their data [15]. Filecoin uses consensus mechanisms to keep outside parties out of the network. To demonstrate to the network that the storage is taking place as agreed upon in the contract between the client and the provider, Filecoin uses "proof of replication" (PoRep) and "proof of space-time" (PoST) to validate storage data. A storage provider uses PoRep to progressively encrypt data. After that, it becomes the storage provider's duty to demonstrate that the data encoding is distinct. The data has been encoded and is being stored safely if the storage provider responds promptly because the encoding sequence is gradual. The storage provider is not operating in good faith and has generated a new encoding if they do not reply right away. Miners use PoSt to demonstrate that data is continuously stored in storage after an agreement has been reached between a user and a storage provider. Random miners demonstrate the data's continued availability based on the volume of data [16].

3.3 Sia

Sia is a decentralized cloud storage platform based on blockchain technology. Skynet is a decentralized file and content delivery and sharing network built on top of Sia's cloud storage network. File-sharing, data publishing, and the infrastructure needed for apps to serve decentralized content are all added by Skynet to Sia. Any kind of data can be hosted on Skynet. Uploading files is possible via a Sia node or the Skynet Web portal. A file that is uploaded creates a Skylink, which is a 46-byte link. Regardless of whether they use Sia or not, anyone can download Skynet data using that link. To maintain the file's availability, the original uploader need not remain online. Sia performs all real-time pinning, guaranteeing fast speeds and outstanding uptime. This is advantageous for decentralized applications

because it allows them to operate with confidence. After all, their storage layer is equally decentralized as their applications. Skynet can quickly, cheaply, and easily store and distribute data. Without compromising performance or dependability, traditional infrastructureis 10 times more expensive than cloud storage, but bandwidth is 100 times less expensive. Figure 7 shows the SIA data encoding and decoding sequence.

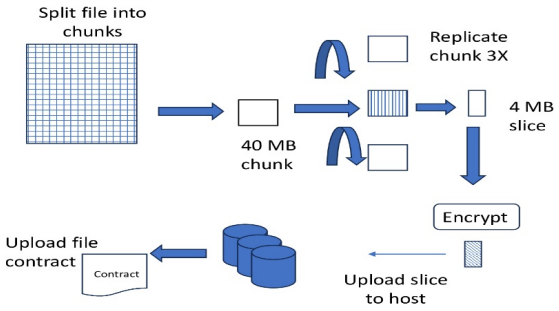

Figure 7. SIA data encoding and decoding sequence

4. Attacks and Solutions on Decentralized Repository

Numerous attacks can be launched against distributed systems. These attacks can impact any distributed storage system and are typically storage-specific.

4.1 Spartacus

Identity theft or Spartacus attacks are possible on Kademlia. By copying the Node ID and receiving a portion of the messages meant for that node, any node can pretend to be another. This technique can be used to target nodes and data. Every message needs to be signed, and ECDSA public key hashes are used to implement Node IDs. If Spartacus attackers tried to attack the system, this would stop them from signing messages or using the system.

4.2 Sybil

By putting up a large number of nodes and dropping or stealing messages, an attack known as Sybil can disrupt a network. It is challenging to carry out Sybil attacks on Kademlia since they depend on duplicate messages and a concrete distance metric. According to their Node IDs, the majority of messages are routed to a minimum of three neighbors of each node within the network. With 50% of the network under attack, Sybil attacks are only able to isolate 12.5 percent of honest nodes. Reliability and performance will decline until a significant portion of the network consists of colliding Sybil nodes, but it will still function otherwise.

4.3 Google

This attack bears similarities to the Sybil assault. Defensive against a Google attack can be difficult because it's hard to predict what the company will do. Building a network with resources comparable to the attackers' is the only way to defend against Google attacks. It would take a lot of resources to aim for the network at that level, which is not sustainable.

4.4 Eclipse Attack

By making sure that every outgoing connection reaches malicious nodes, an eclipse attack isolates a node or group of nodes within a network graph [22]. By tricking malicious nodes into thinking they are legitimate, the Eclipse attack can obscure only specific, crucial messages. By creating keypairs and

keeping this position safe from new nodes with closer IDs, the attacker can overtake the target node by locating three keys whose hashes are closer to the target's ID than the node's closest benign neighbor. As more nodes are added, the proof-of-work problem becomes more challenging because the network already has nodes. Increasing the network's node count will help defend against eclipse attacks.

4.5 Hostage Bytes

Storage-specific techniques like the hostage byte attack are used to demand more money from data owners when malicious storage providers decline to send shards or parts of shards [2]. Users of data may shield themselves from hostage byte attacks by distributing the redundant storage of shards among several nodes. The majority of the practical uses of this attack are handled by redundant storage. However, it is not a perfect solution. In reality, redundancy is exceedingly hard to overcome without several malicious nodes working together [17].

5. Decentralized Storage Network Limitations

Despite their potential, decentralized storage systems undoubtedly have drawbacks. Since the technology is still in its infancy, researchers are working to overcome its difficulties. A few difficulties with blockchain-based decentralized data storage systems are as follows:

5.1 Lack of Trust

Data is stored decentrally in a manner that evades centralized regulations by using peer-to-peer technology. Because there is no accountability in the event of lost data or transactions, businesses and consumers may find it difficult to trust the decentralized network. This lack of trust is motivating the decentralized network's developers to add the highest levels of security. Businesses may need some time to trust new technology.

5.2 Mitigation Issues

Decentralized storage technology is still in its infancy and will stay so for some time. Businesses and consumers do not quickly adopt decentralized storage systems because of performance issues. Early adopters need to take advantage of this strategy before it becomes a mainstream technology. Developers are already addressing issues related to performance [18].

5.3 Security Issues

Despite its resilience, the network is susceptible to hacking attempts by malevolent nodes that can initiate hub attacks, cause disruptions, and potentially compromise the system as a whole. Currently being developed to thwart these attacks are decentralized storage systems based on blockchain [8].

6. Challenges and Advancements in Research

The problems users and organizations have with blockchain-based storage are covered in this section's discussion [19].

6.1 Security

Although blockchain networks are more secure than centralized systems, it's important to remember that security isn't always guaranteed. Decentralized networks are less likely than centralized networks to experience security problems, but they are still occasionally encountered. Due to the need to constantly decrypt and re-encrypt encrypted files, security problems may also arise when data needs to be edited or shared with a third party [20].

6.2 Access Control

Although big amounts of data should be replicated over all nodes and blockchain will always have a record of past transactions, it is not a database in and of itself. These two specifications have the potential to bloat large files stored on blockchain. Blockchain storage networks' inability to allow user file sharing is the issue. To address this problem, smart contract-based solutions have been proposed, but they are limited to IPFS [21].

6.3 Scalability

The issue here is that anyone can join a blockchain network by volunteering to become a node. However, as a network expands, it becomes more difficult to maintain network security and efficiency [22]. Blockchain networks may experience scalability issues, which could lead to delays and other issues. Many remedies have been proposed to address the scalability problems in the Bitcoin blockchain. The writers do point out that there might be more issues than just potential delays. The bootstrap time, which is expensive, is the amount of time it takes for a new node to join the network after downloading and analyzing its history.

6.4 Making a Transfer to the Blockchain

Blockchain networks don't always appear to be problematic at first glance, although there are some situations where they might appear to be. Networks based on blockchain might not always be the best option for every person or business. To ensure you make an informed choice, you must weigh the benefits and drawbacks of a blockchain before putting it into practice. Data storage for a single person is therefore more affordable and secure when done this way. As of yet, cloud storage is incompatible with data analysts and processors, so businesses should exercise caution when using it.

7. Conclusion

We expounded on the significance of decentralized storage networks, which are inherently founded on blockchain technologies, in this concise overview. Even with the abundance of research being done on blockchain-based storage, thorough examinations of each of these networks' storage networks are still desperately needed to determine whether or not they are appropriate for a given use case. Beyond that, blockchains have the potential to have a huge impact on business across a wide range of industries because of their decentralized, peer-to-peer nature. Storing and retrieving data from cloud storage is one of the most important and contentious topics of our day. Storage systems based on blockchain technology address a number of the drawbacks of conventional systems. A new method of storing data to guarantee security and privacy is covered in our poll. Scalability, data analysis, and access concerns keep blockchain-based storage in its infancy, though. Blockchain-based storage is still in the development stage, just like the other new uses for this technology. A blockchain's consensus protocols were all designed with particular objectives, like speed, in mind. To suit their requirements, organizations can create, combine, or alter protocols.

References

[1] X. Zhang, J. Grannis, I. Baggili, and N. L. Beebe, "Frameup: An incriminatory attack on Storj: A peer to peer blockchain-enabled distributed storage system," Digit. Investig., vol. 29, pp. 28–42, 2019, doi: 10.1016/j.diin.2019.02.003.

[2] S. Wilkinson, T. Boshevski, J. Brandoff, and V. Buterin, “Storj a peer-to-peer cloud storage network,” IPFS, U.K., Tech. Rep., 2016, pp. 1–37.

[3] I. Vakilinia, S. Vakilinia, S. Badsha, E. Arslan, and S. Sengupta, "Pooling approach for task allocation in the blockchain-based decentralized storage network," in Proc. 15th Int. Conf. Netw. Service Manage. (CNSM), Oct. 2019, pp. 1–6, doi: 10.23919/CNSM46954.2019.9012719.

[4] A. M. Girgis, O. Ercetin, M. Nafie, and T. ElBatt, "Decentralized coded caching in wireless networks: Trade-off between storage and latency," in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jun. 2017, pp. 2443–2447.
[5] Z. Kong, S. A. Aly, and E. Soljanin, "Decentralized coding algorithms for distributed storage in wireless sensor networks," IEEE J. Sel. Areas Commun., vol. 28, no. 2, pp. 261–267, Feb. 2010, doi: 10.1109/JSAC.2010.100215.
[6] C. Cai, X. Yuan, and C. Wang, "Towards trustworthy and private keyword search in encrypted decentralized storage," in Proc. IEEE Int. Conf. Commun. (ICC), Paris, France, Jul. 2017, doi: 10.1109/ICC.2017.7996810.
[7] N. Z. Benisi, M. Aminian, and B. Javadi, "Blockchain-based decentralized storage networks: A survey," J. Netw. Comput. Appl., vol. 162, Jul. 2020, Art. no. 102656, doi: 10.1016/j.jnca.2020.102656.
[8] A. Shah, N. Sheoran, and S. Gupta, "Decentralized storage network with smart contract incentivization candidate's declaration," Bachelor Technol. Comput. Sci. Eng., Tech. Rep., 2018.
[9] M. Aloqaily, O. Boucher, A. Boukerche, and I. A. Ridhawi, "Design guidelines for blockchain-assisted 5G-UAV networks," IEEE Netw., vol. 35, no. 1, pp. 64–71, Jan. 2021.
[10] M. Firdaus and K. H. Rhee, "On blockchain-enhanced secure data storage and sharing in vehicular edge computing networks," Appl. Sci., vol. 11, no. 1, pp. 1–21, Jan. 2021, doi: 10.3390/app11010414.
[11] N. Lipusch, "Initial coin offerings—A paradigm shift in funding disruptive innovation," SSRN Electron. J., vol. 139, pp. 1–21, Mar. 2018, doi: 10.2139/ssrn.3148181.
[12] S. Yang, A. Hareedy, R. Calderbank, and L. Dolecek, "Topology-aware cooperative data protection in blockchain-based decentralized storage networks," in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jun. 2020, pp. 1-6.
[13] Y. Zhu, G. Zheng, and K.-K. Wong, "Blockchain-empowered decentralized storage in Air-to-Ground industrial networks," IEEE Trans. Ind. Informat., vol. 15, no. 6, pp. 3593–3601, Jun. 2019, doi: 10.1109/TII.2019.2903559
[14] A. S. de Pedro, D. Levi, and L. I. Cuende, "Witnet: A decentralized Oracle network protocol," 2017, arXiv:1711.09756.
[15] S. He, Y. Lu, Q. Tang, G. Wang, and C. Q. Wu, "Fair peer-to-peer content delivery via blockchain," 2021, arXiv:2102.04685.
[16] K. Aldriwish, "A double-blockchain architecture for secure storage and transaction on the Internet of Things networks," Int. J. Comput. Sci. Netw. Secur., vol. 21, no. 6, pp. 119–126, 2021, doi: 10.22937/IJCSNS.2021.21.6.16.
[17] M. Campanelli, D. Fiore, N. Greco, D. Kolonelos, and L. Nizzardo, "Vector commitment techniques and applications to verifiable decentralized storage," Elsevier, Tech. Rep., 2020.
[18] I. Vakilinia, W. Wang, and J. Xin, "An incentive-compatible mechanism for decentralized storage network," 2022, arXiv:2208.09937.
[19] Y. Du, H. Duan, A. Zhou, C. Wang, M. Ho Au, and Q. Wang, "Enabling secure and efficient decentralized storage auditing with blockchain," IEEE Trans. Dependable Secure Comput., vol. 19, no. 5, pp. 3038–3054, Oct. 2022, doi: 10.1109/TDSC.2021.3081826.
[20] Protocol Labs, "Filecoin: A decentralized storage network," IEEE, Tech. Rep., 2017, vol. 9, no. 15.
[21] H. Kopp, M. David, F. Hauck, F. Kargl, and B. Christoph, "Design of a privacy-preserving decentralized file storage with financial incentives," in Proc. IEEE Eur. Symp. Secure. Privacy Workshops (EuroS PW), Apr. 2017, pp. 14–22, doi: 10.1109/EuroSPW.2017.45.
[22] A. P. Kryukov and A. P. Demichev, "Decentralized data storages: Technologies of construction," Program. Comput. Softw., vol. 44, no. 5, pp. 303–315, Sep. 2018, doi: 10.1134/S0361768818050067.
[23] J. C. Song, M. A. Demir, J. J. Prevost, and P. Rad, "Blockchain design for trusted decentralized IoT networks," in Proc. 13th Annu. Conf. Syst. Syst. Eng. (SoSE), Jun. 2018, pp. 169–174.