Decentralization and web3 technologies

Gaurish Korpal and Drew Scott
The University of Arizona
Tucson, Arizona 85721

***Abstract*—Today, the majority of the web's content and user data is controlled by a few large tech companies. There is a growing movement to devolve this control evenly across the entire internet, representing the transition to Web3. In order for this movement to be successful, technologies and protocols must be developed to enable web users to use the web securely without trusting any other user. That is, today's web is structured so that users must trust these companies, so trustless alternatives haven't already been developed. Broadly, this movement emphasizes developing peer-to-peer networks, blockchains, and distributed storage systems. These systems make use of cryptographic primitives to guarantee security.**

I. Introduction

This survey hopes to describe the breadth and the basics of web3 technologies, which come together to attempt to fully decentralize the web. A decentralized system is generally encouraged because of fault tolerance, attack resistance, and collusion resistance. Therefore, most of the decentralized network protocols tend to rely on distributed systems or peer-to-peer networks, which can help fight government censorship and monopoly posed by the giant tech companies. As it currently stands, the web is, however, quite centralized. Thus, in the recent past, there have been many researchers who have developed decentralized protocols and applications to work toward the goal of decentralizing the web. This survey will capture the major advancements these researchers have made.

The rest of this survey is organized as follows: section II discusses the history of the web and the motivation for the transition to Web3; section III discusses blockchains, with a focus on technical details of Bitcoin and later research that improves on Bitcoin; section IV discusses decentralized storage; section V introduces additional research directions related to Web3; and section VI concludes this survey.

II. History and Motivation

The wide variety of networks can be factored into two components: centralized (or star) and distributed (or grid or mesh). In practice, a mixture of star and mesh components is used to form a communication network. Paul Baran, one of the two independent inventors of packet switching, called such a network a "decentralized" network because complete reliance upon a single point is not always required [1] (Figure 1).

Even though the Internet was built on distributed protocols, the web needed to consolidate around a few curated service platforms in order to become practical for everyday people to use. This trend towards consolidation lead to serious implications for the two key functions of web, the web–publishing and discovery of content [2]. Therefore, in today's web, a small number of stakeholders have an outsized influence over the content the public can create and consume.

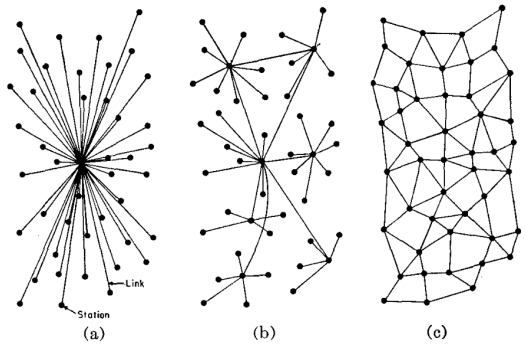

Fig. 1—(a) Centralized. (b) Decentralized. (c) Distributed networks.

Fig. 1. The variety of networks ©1964 IEEE [1].

In modern literature, the decentralized networks are the ones in which the control of infrastructure and other technologies can be distributed among contributors rather than dictated by large ("central") players. This control comes in many forms, such as having ownership of web infrastructure (e.g. servers or portals), ownership of data, influence to make decisions about the network, the power to delete contents, or decide who may access the network's shared capabilities, information, and knowledge. For example, Facebook is a centralized network where data is controlled by a central entity, and Amazon's Cloud is a distributed network where the data is stored across a grid, but still controlled by a central entity [3].

A. Re-decentralization

The excitement around the area of decentralized systems have grown in prominence over the past decade due to the popularity of the cryptocurrencies like Bitcoin. Bitcoin is a payment system that has no central points of control, and uses a novel peer-to-peer network protocol to agree on a distributed ledger of transactions, the blockchain [4]. Therefore, the blockchain technology has painted a picture of a world where untrusted networks of computers can coordinate to provide important infrastructure, like distributed storage and privacy. Advocates of these decentralized systems propose related technology as the way forward to "re-decentralize" the Web, by shifting publishing and discovery out of the hands of a few corporations, and back into the hands of users [5]. For example, Starling Lab in their "78 days" project[1] with Reuters

[1]https://www.starlinglab.org/78days/

demonstrated that a lot of information relevant to humanity can be preserved through the decentralized system. In this project they addressed the challenges of securely capturing, storing and authenticating digital photos (like identifying deep fakes). The Starling Framework used Filecoin, IPFS, Hyperledger Fabric, and GUN ecosystem (gun.eco) to securely store digital photos on a distributed network that supports authentication using the Content Authenticity Initiative (CAI) [6]. Therefore, as summarized by Zhanglong Peng in their report [7], "As the masses garner for democracy over the internet, decentralization may be the only way to achieve this. A network on its own, giving people the freedom they desire online."

However, similar to the Dark Web [8], the decentralized web will also make law enforcement nearly impossible. For example, the shadow libraries like Library Genesis (Libgen) and Sci-Hub now use IPFS-based peer-to-peer distributed library system supported by Cloudflare[2], to avoid legal attacks (domain takedowns, server shutdowns and international wom-anhunts/manhunts) [9]. Moreover, with blockchain technology and decentralized applications the enforcement of copyrights won't be possible because those buying and selling unautho-rized copies of copyrighted material on a decentralized Internet cannot be subject to court injunctions [10].

B. Self-Certifying Web Protocols

In 1989, Tim Berners-Lee and Robert Cailliau together invented the World Wide Web in which a typical user's (self-hosted) website would be made up of hyperlinked text, files, applications, and other digital objects that could be read and/or downloaded by website visitors, now referred to as "Web 1.0" or "read-only web."

The current web, referred to as "Web 2.0" (a term coined by Tim O'Reilly in 2007 [11]) or "the read-write web" (a term coined by Richard McManus in 2003) started to develop in the 2000s when new platforms emerged which allowed users to interact with content, and with one another (like Facebook and eBay). However, it doesn't capture the original vision for the Web to be a medium for the secure, decentralized exchange of public and private data [12]. Therefore, the re-decentralized web is called "Web 3.0" or "DWeb."

1) Web 3.0: The term "Web 3.0" or "post-Snowden web" was coined[3] by Polkadot founder and Ethereum co-founder Gavin Wood in 2014, referring to a "decentralized online ecosystem based on Ethereum [14]" [15]. According to Wood, "Web 3.0 is not about cryptocurrency, blockchain, or toke-nomics [16]. Web 3.0 is about decentralization, openness, and transparency" [17]. In 2017, the Web3 Foundation[4] was established, which published the 5 level (L0 to L4) Web 3.0 Technology Stack (Figure 2).

2) DWeb: The term "DWeb" or "Decentralized and Dis-tributed Web" was coined by Internet Archive founder and

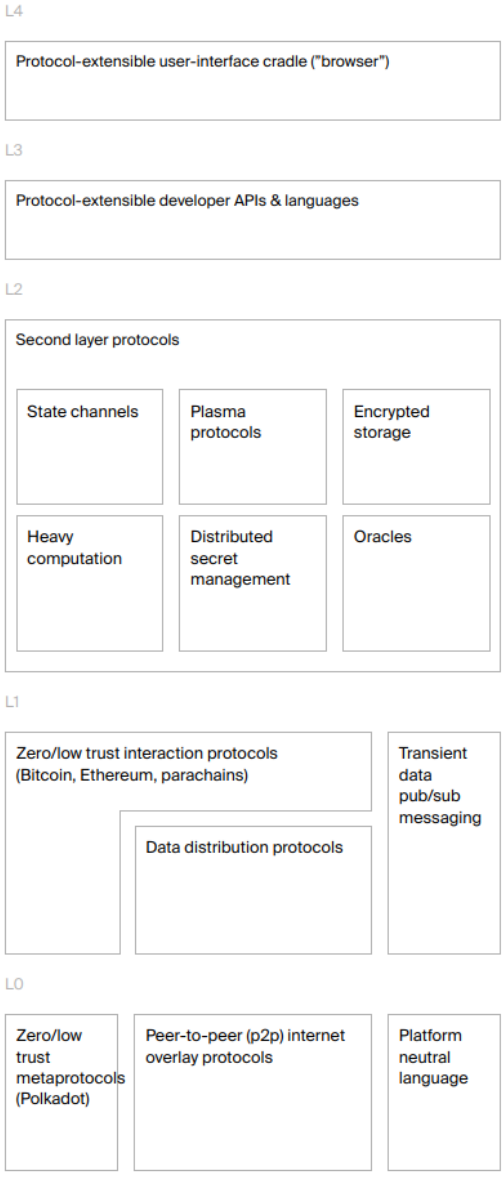

Fig. 2. Web 3.0 Technology Stack ©2022 Web3 Foundation [18]

Alexa Internet co-founder Brewster Kahle in 2015, referring to "a Web that is reliable, private and fun all at the same time built using the technologies like Maelstrom [19], blockchain (Ethereum, Namecoin, Storj), Bitcoin for payments, I2P[5], IPFS, Tahoe-LAFS and WebRTC" [20]. In 2016, DWeb com-munity [21] was formed that organizes various summits[6] and camps[7] to promote open-source development of the technolo-gies needed to build a decentralized web like decentralized storage, hosting, domain names, data, identity, data transport layers, or payments. For example, Distributed Press API [22] is used for publishing an online magazine, called COMPOST[8],

[2]https://developers.cloudflare.com/distributed-web/ipfs-gateway/

[3]The terms "Web3" and "Web 3.0" are also sometimes used to refer to Tim Berners-Lee's vision of the Semantic Web [13]. However, we won't talk about it in this paper.

[4]https://2018.web3summit.com/speakers/

[5]https://geti2p.net/en/

[6]https://www.decentralizedweb.net/

[7]https://dwebcamp.org/

[8]https://two.compost.digital/

on WWW and DWeb by seeding it to Hypercore [23] and IPFS, and can be viewed in Brave Browser [24].

However, there are many who believe that "blockchain-ification" of the Web [25] [26], using blockchain technologies like cryptocurrencies to verify transactions, pay for services, and certify content such as NFTs is a big fraud, since in their opinion, "Web3 is only about adding an additional layer of complexity in the name of justifying the underlying cryptocurrencies" [27]. In fact, there exist other projects like Named Data Networking (NDN)[9] which aim to achieve some level of decentralization by replacing the existing TCP/IP framework, without using blockchain based technologies; a comparison between NDN and blockchain based solutions can be found in [28, §II] and [13, §5.7]. Moreover, there exist projects like Beaker browser which use Dat protocol (instead of IPFS) to support P2P website hosting without using blockchains [29].

Therefore, Jay Graber [30] gave the following definitions to highlight the fact that there-is-more-than-blockchain in Web3:

- Web1 - Host-generated content, host-generated authority.
- Web2 - User-generated content, host-generated authority.
- Web3 - User-generated content, user-generated authority.

The latter is enabled by "self-certifying protocols" based on cryptographic user identifiers and content-addressed data, which helps decentralize. This definition of "Web3 = DWeb" also includes older protocols such as Git, PGP, Tahoe-LAF, and BitTorrent, and newer ones like IPFS, Polkadot, Hypercore and Secure Scuttlebutt (SSB).

III. Blockchains

Blockchains have facilitated the development of Web3 by providing public and immutable stores of data.

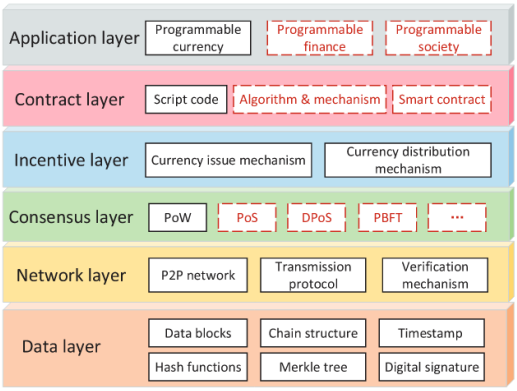

Fig. 3. The architecture of Blockchain, where the contents of the red dotted were not part of the original Bitcoin design ©2019 ACM [31].

A decentralized system without a similar structure wouldn't properly function because its users wouldn't be able to verify the system's state. In such a scenario, the system's users would, in the best case, lack confidence the system effectively carries out their requested actions. Obviously, a system shrouded in uncertainty is ineffective. Recently, there have been many advancements in blockchain technologies, most of which are based on critiquing and improving upon the first blockchain, Bitcoin. Thus, a discussion of Bitcoin's fundamental features follows, after which improvements on Bitcoin will be acknowledged and briefly discussed. However, since it is impossible to capture the true essence of the role of blockchain in this short paper, we would encourage the readers to also read the classic papers like [32], [33], [34], [35], [36], and [37], short expository articles like [38] and [39], and books like [40] and [41].

A. Bitcoin

The Bitcoin blockchain was created for a single application: decentralized currency. Decentralization, in this context, means that governments have no control over the supply or state of the currency and banks are not needed to complete transactions. Creating a currency that does not require any centralized entity is a difficult task for one main reason: a payee must be able to verify that a payer truly has enough money to complete a transaction. More specifically, (1) the payee must be able to verify that the payer is paying them with real money, (2) the payee must be able to verify that the payer is paying them with money the payer owns, and (3) the payee must be able to verify that the payer hasn't already spent the money that they are sending to the payee. In a centralized system, banks can easily track how much money their customers have and can verify said money is real via a central bank. Then, this information can be used to guarantee the integrity of transactions. But a decentralized system, which doesn't use centralized third parties to complete transactions, must enable its users to verify transactions on their own. This can only be done with an accurate and publicly available history of transactions, so that payees can verify they are paid with real money. This is why a blockchain is used for Bitcoin.

A blockchain is a database that is shared across a network of computers [42]. It is designed to be an immutable and append-only structure. Without this property, past transactions could be erased or forged, resulting in a completely dysfunctional system. Implementation details of the blockchain's structure follows to explain why users can trust its immutable property. Blockchains group transactions into "blocks", where each new block is connected to the previous block to form a "chain". This chain is created by including the cryptographic hash of the previous block in the data of the new block. Thus, if a malicious actor attempts to change the state of a block (i.e. modify a previous transaction), the cryptographic hash of the block will change, so its link to the next block will be broken. So, in order to modify a previous block, a malicious actor must update each subsequent block's hash link to create a valid chain. Therefore, a blockchain's security is dependent on the difficulty of creating and changing blocks. So, the Bitcoin blockchain uses a "proof of work" protocol, which involves repeatedly computing the cryptographic hash of two parameters, the block itself and a generated number, until the hash value is smaller than a threshold [43]. By establishing a sufficiently low threshold, the network can assume a significant

[9]https://named-data.net/

amount of work was put in to find that number. Thus, a malicious actor must spend a great deal of time recreating the entire chain after the block they want to change. For this reason, the blockchain is treated as an immutable data structure (Figure 4).

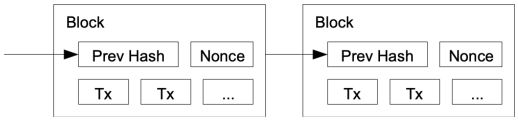

Fig. 4. Overview of a blockchain's structure, including the hash link, the proof of work nonce and transactions [43].

Each Bitcoin transaction is a transfer of ownership of a digital coin from the payer to the payee. To transfer a coin, a hash of the payee's public key and the coin's previous state is computed. Then, the payer signs this hash using their private key to confirm the transfer of ownership [43]. Thus, an outsider is able to look at a coin's previous state to verify the current state was created using a valid public/private key pair (Figure 5).

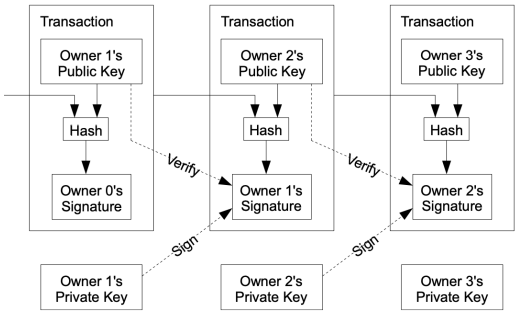

Fig. 5. A visual representation of a coin's state as it transfers between users [43].

B. Diverse Blockchain Applications

As previously mentioned, Bitcoin was created to support one service: financial transactions. Web3, however, refers to decentralizing the entire web, not just a small subset of its applications. So, soon after Bitcoin was introduced, another blockchain, called Ethereum [44], was created to serve a broad array of applications. The Ethereum blockchain natively supports smart contracts, which allow users to write code that can be stored on the blockchain to be executed when needed. Therefore, users can write any application they want and have it run on the Ethereum blockchain. For example, smart contracts can be written to create sub-currencies, Decentralized Autonomous Organizations, reputation systems, and much more (Figure 6).

C. Scalability

Bitcoin has two primary weaknesses in regard to its ability to scale. First, due to the large overhead associated with storing and verifying transactions, Bitcoin cannot handle the quantity

```
from = msg.sender
to = msg.data[0]
value = msg.data[1]

if contract.storage[from] >= value:
    contract.storage[from] = contract.storage[from] - value
    contract.storage[to] = contract.storage[to] + value
```

Fig. 6. A sample Ethereum smart contract for implementing a sub-currency [44].

of transactions required to support the global economy. Second, since Bitcoin uses a computationally intensive and naive proof of work protocol, mining power can centralize as a result of skyrocketing energy costs. A discussion of both of these problems follows.

1) Bitcoin Lightning Network: Bitcoin can support around 7 transactions per second, while Visa can handle around 47,000 transactions per second [45]. Clearly, the Bitcoin blockchain must be able to handle more transactions to become a viable global system. One solution, the Bitcoin Lightning Network [45], tackles this problem by only processing a small subset of transactions directly on the blockchain, which allows for more transactions in the system than before. While it would seem that processing transactions away from the blockchain would break the security guarantees of the system, it is possible to design this system to be functional. The Bitcoin Lightning Network establishes off-blockchain "payment channels" between two Bitcoin users. A payment channel is a pool of shared money where the portion each user owns is able to change over time. Only two blockchain transactions are needed for payment channels: (1) opening a channel (users input their amount of the shared pool) and (2) closing a channel (users liquidate their share of the pool). All other transactions between users with a shared payment channel don't need to be processed on the blockchain; these transactions only go through the payment channel. Each time a transaction is made through a payment channel, the users generate a new public/private key pair to sign it. Then, if the payment channel is closed, they can claim their money using the corresponding key on the blockchain. So, if two users are frequently making transactions, establishing a payment channel between them significantly cuts down the number of blockchain transactions used.

One security concern associated with these channels is that a user may maliciously close the channel using a previous state (i.e. a state where they had a larger portion of the shared pool than they do now). To solve this problem, whenever a new state of the channel is created, each user reveals their private key of the previous state, so that either user can claim the currency if a previous state has been maliciously revealed [46] (Figure 7).

The problem of scalability, however, isn't solved if each user has to have a payment channel with each other user. Thus, payments must be able to flow through payment channels between third parties. So a mechanism, called a Hash Time Lock Contract (HTLC) [46], is used to guarantee the two end

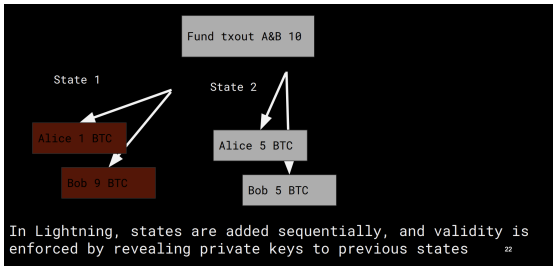

Fig. 7. A state transition of a payment channel between Alice and Bob [46].

users that payments are not stolen by the forwarding parties. First, the payee generates a random number and calculates its hash. Then, the hash is directly communicated to the payer. Then, the payer forwards the payment to a forwarding party, but the payment is locked using an HTLC corresponding to the hash created by the payee. The forwarding party may unlock the HTLC only if they can show they know the random number that generated the hash. So, they will forward money via other payment channels to eventually reach the payee, who will reveal the random number once their payment is received. Then, each forwarding party can unlock their HTLC to claim the money (Figure 8 and Figure 9).

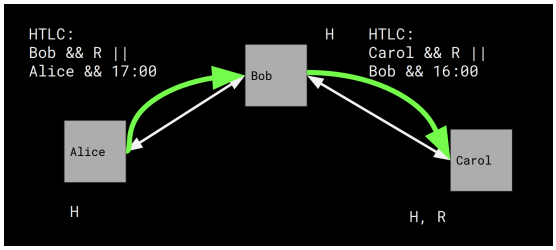

Fig. 8. Alice forwards a payment to Bob (untrusted) to go to Carol. R is the random number and H is its hash [46].

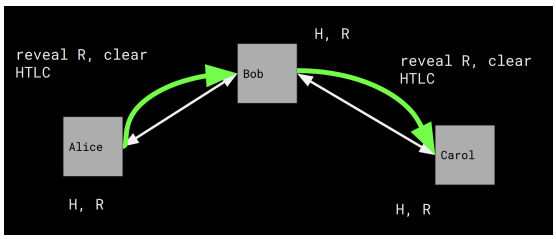

Fig. 9. Bob forwarded the payment to Carol, so Carol can reveal R, so Bob can claim the money from his payment channel with Alice [46].

2) Plasma: Similar to the Lightning Network, Plasma [47] is a series of contracts which runs on top of an existing blockchain (e.g. Ethereum) to ensure enforcement while ensuring that one is able to hold funds in a contract state with net settlement/withdrawal at a later date. It is composed of two key parts of the design: Reframing all blockchain computation into a set of MapReduce functions, and an optional method to do Proof-of-Stake token bonding on top of existing blockchains with the understanding that the Nakamoto Consensus incentives discourage block withholding. Plasma is not designed to reach assured finality rapidly, even though transactions are confirmed in the child chains rapidly, it requires it to be finalized on the underlying root blockchain. However, with Lightning (including Lightning on top of Plasma), it's possible to do incredibly rapid updates with reasonable sense of localized finality.

3) Polkadot: Polkadot [48] is designed to be a fully extensible and scalable blockchain development, deployment and interaction test bed. The usual blockchain implementations focus on providing a single chain of varying degrees of generality over potential applications. However, Polkadot is a scalable heterogeneous multichain designed to provide no inherent application functionality at all. It is considered to be scalable because it is a set of independent chains (e.g. the set containing Ethereum, Namecoin and Bitcoin), called "parachains," with pooled security and trust-free interchain transactability. There are four basic roles in the upkeep of a Polkadot network: validator, nominator, collator, and fisherman. Loosely speaking, the validators are similar to the mining pools of current PoW blockchains; the nominators are similar to the miners of the present-day PoW networks; and the fishermen are independent "bounty hunters" motivated by a large one-off reward with low resource requirement and bandwidth commitment, similar to "full nodes" in present-day blockchain systems. The collators maintain a "full-node" for a particular parachain, i.e. they retain all necessary information to be able to author new blocks and execute transactions in much the same way as miners do on current PoW blockchains. The precise nature of the relationship between collators, nominators and validators is expected to change over time. There is some similarity between the design of Polkadot and Plasma; instead of a structure with "fishermen" validators ensuring block accuracy, Plasma constructs a series of child blockchains which enforce state via Merkle proofs [47].

4) Proof of Work Shortcomings: As previously mentioned, Bitcoin operates using a simple hash-based proof of work protocol. This has two primary weaknesses that harm Bitcoin's ability to effectively scale up. First, the proof of work is simple to compute; that is, it only requires a number and the basic information of the current block. This has led to the popularity of "mining pools" where nodes may contribute to finding the proof of work without independently verifying the transactions [44]. So, only the leaders of the pools are deemed responsible for validating transactions, which is a form of centralization. One solution, proposed by Ethereum, uses a proof of work protocol that forces miners to include more data of the transactions, thus leading miners to independently verify transactions. Second, proof of work is, by its nature, computationally and energy intensive. As the network grows, proofs of work become harder and harder, thus requiring much more energy. So, other protocols such as proof of stake [48], proof of authority [48], and proof of storage [49] have been proposed to cut down on energy usage.

D. Blockchain Privacy

It is important to note that Bitcoin doesn't provide built-in, fully anonymous transactions. That is, since transactions must be publicly broadcast, it is possible to trace Bitcoin transactions back to their owners. Users, whether they are individuals or organizations, often do not want the public to know where their money is going or where it is coming from. Additionally, in a system where transactions can be traced through time to users with poor reputations (consider a terrorist group or sanctioned foreign government), some coins may not be accepted as payment (i.e. coins are no longer fungible), which is undesirable. Zerocoin [50] and Zerocash [51], which builds on Zerocoin, approach this problem by creating an automated and secure laundering protocol for Bitcoin. These protocols are based on the concept of zero knowledge proofs, wherein a laundry user is able to prove to others that they own the laundered coins without revealing their identity [31].

IV. Decentralized Storage

Decentralized storage can be visualized as a network of P2P (peer-to-peer) servers that store data across a global network of storage nodes. Thus, in a truly decentralized storage system both storage location and storage management are decentralized [3].

A. Distributed File System

The InterPlanetary File System (IPFS) is a peer-to-peer distributed file system that seeks to connect all computing devices with the same system of files. It combines the important P2P properties like Distributed Hash Tables (DHTs), filesharing system (BitTorrent), version control system (Git), and Self-Certified File Systems (SFS). The central IPFS principle is modeling all data as part of the same Merkle Directed Acyclic Graphs (DAGs)[10] [52]. A Directed Acyclic Graph (DAG) is a type of graph in which edges have direction and cycles are not allowed. A Merkle-DAG is a DAG where each node has an identifier and this is the result of hashing the node's contents (any opaque payload carried by the node and the list of identifiers of its children) using a cryptographic hash function like SHA256 [53]. A blockchain-based extension to IPFS, called acl-IPFS [54], has been proposed which can provide access control by leveraging Ethereum smart contracts to handle the access control list (users can register files, and grant or revoke access to them).

B. Decentralized Storage Network

Decentralized Storage Networks (DSNs) aggregate storage offered by multiple independent storage providers and self-coordinate to provide data storage and data retrieval to clients. Coordination is decentralized and does not require trusted parties: the secure operation of these systems is achieved through protocols that coordinate and verify operations carried out by individual parties. Following are the three popular DSN providers:

- Filecoin [49] works as an incentive layer on top of IPFS, which can provide storage infrastructure for any data. Clients pay a network of miners for data storage and retrieval; miners offer disk space and bandwidth in exchange for payments. Miners receive their payments only if the network can audit that their service was correctly provided using Proof-of-Replication (PoRep) and Proof-of-Spacetime (PoSt) schemes.
- Storj [55] is an Amazon S3 compatible DSN that assumes the Ethereum-based STORJ token as the default mechanism for payment. It also supports a reference architecture that backs an IPFS node[11] backed by their decentralized network self-hosted using services like Nextcloud.
- Sia [56] is a variant on the Bitcoin protocol (Proof of Work) that enables decentralized file storage via cryptographic contracts. These contracts can be used to enforce storage agreements between clients and hosts. After agreeing to store a file, a host must regularly submit Proof of Storage (PoS) to the network. The host will automatically be compensated using Siacoin for storing the file, regardless of the behavior of the client. It also supports IPFS via Skynet[12].

C. Personal Data Stores

Today, data is a valuable asset in our economy, and we all reap the benefits of a data-driven society. However, there is a growing public concern about user privacy. Personal data stores (PDS) enable individuals to easily reuse their data by providing more control and transparency on how this data is stored and shared. For example, openPDS presents a model for autonomous deployment of a PDS which includes a mechanism for returning computations on the data instead of the raw data itself [57]. In fact, PDS are a key technical component of the *Solid* project (*so*cial *l*inked *d*ata) started by Tim Berners-Lee for decentralizing the WWW [58]. Moreover, blockchains can be used to design a decentralized PDS which acts like an automated access-control manager that doesn't require trust in a third party [59].

D. Decentralized Communication

An open, decentralized messaging platform is essential for the success of the decentralized web. Therefore, several decentralized instant messaging (IM) protocols have were developed in recent decades and one of the most popular one among them is the Matrix protocol. Matrix consists mainly of home servers and clients. Inside these home servers are something called rooms, which are a virtual environment where messages get sent between the participating clients. The decentralization of the Matrix comes from the fact that each room is stored on every single home server participating in the same room. This creates a federated network of home servers and clients where no central server exists [60]. Another example is Status, which provides a messaging platform via the Whisper protocol and

[10] https://dag.ipfs.io/

[11] https://www.storj.io/blog/ipfs-now-on-storj-network

[12] blog.sia.tech/supercharge-your-ipfs-apps-with-homescreen-6ecf147eb4cc

also a mobile interface to interact with decentralized applications (DApps) that run on the Ethereum Network [61]. Because the Ethereum protocol also acts as a large distributed key-store, one can migrate user accounts, credentials, and reputation on-blockchain with the help of the Whisper communication protocol [14].

V. Research Directions

Let's look at some important research directions related to decentralization and web3 technologies.

A. Security and Privacy of Smart Contracts

Smart contracts, as previously discussed, allow developers to build new applications on top of a blockchain. Thus, smart contract research is quite popular, especially in topics which relate to security and privacy. Hawk [62], similar to the previously discussed Zerocoin and Zerocash privacy protocols for Bitcoin, works to make smart contract transactions untraceable on the Ethereum blockchain by creating a framework for smart contract programmers to ensure privacy. DEFIER [63] is a tool developed to analyze the history of the Ethereum blockchain to discover and analyze patterns of attacks. It detected hundreds of thousands of exploit attempts and can be used by smart contract owners to identify if their system is vulnerable. Tools like oyente and Manticore analyze smart contract code to find vulnerabilities, and a tool called MAIAN is even able to generate input to create an exploit for an identified vulnerability [64].

B. Cyberattacks

The majority of successful botnet takedown operations rely on exploiting or subverting botnet command and control (C&C) infrastructures used by the owner. However, Bitcoin offers an ideal C&C dissemination mechanism because C&C communications over the Bitcoin network cannot be shut down simply by confiscating a few servers or poisoning routing tables [65]. Therefore, it is possible to create an undetectable malware based on the blockchain technology [66]. In fact, IPFS's capability to provide anonymity, persistence of the content, fast delivery and a robust network where content cannot be easily blocked provides an ideal landscape for malware authors [67]. Moreover, by using IPFS instead of blockchain to store the malware, criminals can remain offline during most procedures, with many privacy guarantees [68].

C. Scalability and Mass-Adoption

In a previous section, blockchain scalability was discussed through the lens of making it feasible for a network to handle a global scale of operations. Now, scalability problems can be viewed as a human problem: how do you create a system that many people want to participate in? A prevalent struggle in P2P networks is to incentivize peers to participate in the serving of content. Gringotts [69] provides a "Proof of Delivery" mechanism to incentivize the distribution of content across the network. FLOCK [70] is an allocation framework for jobs on a network of diverse applications (like a smart contract blockchain). It allows parties to state which jobs they prefer to process in the hopes of increasing satisfaction of participating in the network. A more complete discussion of blockchain design choices is discussed by Wohrer, et al. [71].

VI. Conclusion

In this survey, we have summarized the meaning of and the motivation for Web3, as well as its most well known applications. It should be clear that secure Web3 applications and protocols are much more difficult to design than their Web2 counterparts because decentralization requires storing data publicly, where trust is not a guarantee. Thus, the entire web must be restructured to adapt to these new assumptions. We believe our survey has conveyed the breadth and nature of the research that is occurring in the Web3 landscape. In the end, these researchers are working to build a decentralized web that has at least as strong performance, security, privacy, etc. guarantees that the current web possesses.

Acknowledgment

The authors would like to thank Sazzadur Rahaman, Yang Hong, Amy Paul, Saiful Islam Salim, and Nicolas D Winsten for their feedback on drafts of this paper.

References

[1] P. Baran, "On distributed communications networks," *IEEE Transactions on Communications Systems*, vol. 12, no. 1, pp. 1–9, 1964.

[2] C. Barabas, N. Narula, and E. Zuckerman, "The Decentralized Web," Aug. 2017. [Online]. Available: https://dci.mit.edu/decentralizedweb

[3] "Webinar Series - Imagining a Better Online World: Exploring the Decentralized Web." [Online]. Available: https://archive.org/details/dweb-webinar-series

[4] A. Gervais, G. O. Karame, V. Capkun, and S. Capkun, "Is Bitcoin a Decentralized Currency?" *IEEE Security & Privacy*, vol. 12, no. 3, pp. 54–60, May 2014. [Online]. Available: https://ieeexplore.ieee.org/document/6824541

[5] "Redigest - Monthly newsletter by redecentralize.org." [Online]. Available: https://redecentralize.org/redigest/

[6] "Image Authentication." [Online]. Available: https://www.starlinglab.org/image-authentication/

[7] Z. Peng, "Decentralized Internet," Dec. 2019, a paper written under the guidance of Prof. Raj Jain. [Online]. Available: https://www.cse.wustl.edu/~jain/cse570-19/ftp/decentrl/index.html

[8] N. Bolduc, "The Dark Web & Web3," Mar. 2022. [Online]. Available: https://our.status.im/the-dark-web-web3/

[9] shrine, "p2p Free Library: Help build humanity's free library on IPFS with Sci-Hub and Library Genesis," Oct. 2020. [Online]. Available: www.reddit.com/r/DataHoarder/comments/jb1hkn/p2p_free_library_help_build_humanitys_free/

[10] N. Vogel, "The Great Decentralization: How Web 3.0 Will Weaken Copyrights," *UIC Review of Intellectual Property Law*, vol. 15, no. 1, Jan. 2015. [Online]. Available: https://repository.law.uic.edu/ripl/vol15/iss1/6

[11] T. O'Reilly, "What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 1008839, Aug. 2007. [Online]. Available: https://papers.ssrn.com/abstract=1008839

[12] T. Berners-Lee, "SolidProject origin." [Online]. Available: https://solidproject.org/origin

[13] J. Zarrin, H. Wen Phang, L. Babu Saheer, and B. Zarrin, "Blockchain for decentralization of internet: prospects, trends, and challenges," *Cluster Computing*, vol. 24, no. 4, pp. 2841–2866, Dec. 2021. [Online]. Available: https://link.springer.com/10.1007/s10586-021-03301-8

[14] T. Gerring, "building the decentralized web 3.0," Aug. 2014. [Online]. Available: https://blog.ethereum.org/2014/08/18/building-decentralized-web/

[15] G. Wood, "DApps: What Web 3.0 Looks Like," Apr. 2014. [Online]. Available: http://gavwood.com/dappsweb3.html

[16] S. Voshmgir, *Token economy: how the Web3 reinvents the internet*, 2nd ed. Berlin: BlockchainHub, 2020. [Online]. Available: https://github.com/sherminvo/TokenEconomyBook

[17] "Polkadot's Gavin Wood on Building a Layer 0 to Underpin the Entire Blockchain-Based Economy." [Online]. Available: https://youtu.be/-avBxG3u0ik?t=4554

[18] "Web 3.0 Technology Stack." [Online]. Available: https://web3.foundation/about/

[19] D. Aleksandersen, "What happened to BitTorrent's Project Maelstrom web browser?" [Online]. Available: https://www.ctrl.blog/entry/bittorrent-maelstrom.html

[20] R. Macdonald, "Locking the Web Open, A Call for a Decentralized Web," Feb. 2015. [Online]. Available: https://blog.archive.org/2015/02/11/locking-the-web-open-a-call-for-a-distributed-web/

[21] "DWeb Community." [Online]. Available: https://getdweb.net/origin-story/

[22] "Distributed Press." [Online]. Available: https://distributed.press/

[23] M. Buus, P. Frazee, and A. Osheroff, "How the Hypercore Protocol Works." [Online]. Available: https://hypercore-protocol.org/protocol/

[24] "Brave Integrates IPFS," Jan. 2021. [Online]. Available: https://brave.com/brave-integrates-ipfs/

[25] B. Laboon, "Web3 Blockchain Fundamentals MOOC." [Online]. Available: https://www.youtube.com/playlist?list=PLxVihxZC42nF_MCN9PTvZMIifRjx9cZ2J

[26] F. F. for the Decentralized Web, "Exploring the Decentralized Web." [Online]. Available: https://www.youtube.com/playlist?list=PL37YlBYJT0nmfqDnbov6lKHUyZvRfQjap

[27] N. Weaver, "The Web3 Fraud," Dec. 2021. [Online]. Available: https://www.usenix.org/publications/loginonline/web3-fraud

[28] R. Singh, A. Donegan, and H. Tewari, "Framework for a Decentralized Web," in *2020 30th International Telecommunication Networks and Applications Conference (ITNAC)*. Melbourne, VIC, Australia: IEEE, Nov. 2020, pp. 1–7. [Online]. Available: https://ieeexplore.ieee.org/document/9315032/

[29] P. Frazee, T. Vancil, and M. Buus, "Beaker Documentation," May 2018, an analysis for COMP 117: Internet-scale Distributed Systems. [Online]. Available: https://docs.beakerbrowser.com/

[30] J. Graber, "Web3 is Self-Certifying," Dec. 2021. [Online]. Available: https://jaygraber.medium.com/web3-is-self-certifying-9dad77fd8d81

[31] R. Zhang, R. Xue, and L. Liu, "Security and Privacy on Blockchain," *ACM Computing Surveys*, vol. 52, no. 3, pp. 51:1–51:34, Jul. 2019. [Online]. Available: https://doi.org/10.1145/3316481

[32] D. L. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms," *Communications of the ACM*, vol. 24, no. 2, pp. 84–90, Feb. 1981. [Online]. Available: https://dl.acm.org/doi/10.1145/358549.358563

[33] L. Lamport, R. Shostak, and M. Pease, "The Byzantine Generals Problem," *ACM Transactions on Programming Languages and Systems*, vol. 4, no. 3, pp. 382–401, Jul. 1982. [Online]. Available: https://dl.acm.org/doi/10.1145/357172.357176

[34] R. C. Merkle, "A Certified Digital Signature," in *Advances in Cryptology — CRYPTO' 89 Proceedings*, G. Brassard, Ed. New York, NY: Springer New York, 1990, vol. 435, pp. 218–238. [Online]. Available: http://link.springer.com/10.1007/0-387-34805-0_21

[35] S. Haber and W. S. Stornetta, "How to time-stamp a digital document," *Journal of Cryptology*, vol. 3, no. 2, pp. 99–111, Jan. 1991. [Online]. Available: http://link.springer.com/10.1007/BF00196791

[36] N. Szabo, "Smart Contracts," 1994. [Online]. Available: https://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart.contracts.html

[37] "Bitcoin literature." [Online]. Available: https://nakamotoinstitute.org/literature/

[38] A. Narayanan and J. Clark, "Bitcoin's Academic Pedigree: The concept of cryptocurrencies is built from forgotten ideas in research literature." *Queue*, vol. 15, no. 4, pp. 20–49, Aug. 2017. [Online]. Available: https://dl.acm.org/doi/10.1145/3134434.3136559

[39] J. Waldo, "A Hitchhiker's Guide to the Blockchain Universe: Blockchain remains a mystery, despite its growing acceptance." *Queue*, vol. 16, no. 6, pp. 21–35, Dec. 2018. [Online]. Available: https://dl.acm.org/doi/10.1145/3305263.3305265

[40] P. Champagne, *The book of Satoshi: the collected writings of bitcoin creator Satoshi Nakamoto.* E53 Publishing LLC, 2014, oCLC: 1027652160. [Online]. Available: https://www.bookofsatoshi.com/

[41] A. Narayanan, J. Bonneau, E. Felten, A. Miller, and S. Goldfeder, *Bitcoin and cryptocurrency technologies: a comprehensive introduction*. Princeton: Princeton University Press, 2016. [Online]. Available: https://bitcoinbook.cs.princeton.edu/

[42] M. Murray, "Blockchain explained - Reuters Visual Guide," Jun. 2018. [Online]. Available: http://graphics.reuters.com/TECHNOLOGY-BLOCKCHAIN/010070P11GN/index.html

[43] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," Oct. 2008. [Online]. Available: https://bitcoin.org/bitcoin.pdf

[44] V. Buterin, "Ethereum Whitepaper," 2014. [Online]. Available: https://github.com/ethereum/wiki/blob/f83c4692be242ad350bef0c5f8757b7b3c27b2d9/%5BEnglish%5D-White-Paper.md

[45] J. Poon and T. Dryja, "The Bitcoin Lightning Network: Scalable Off-Chain Instant Payments," Jan. 2016. [Online]. Available: https://lightning.network/lightning-network-paper.pdf

[46] T. Dryja, "Lecture 13: Payment Channels and Lightning Network," Cambridge, MA, 2018. [Online]. Available: https://www.youtube.com/watch?v=Hzv9WuqIzA0

[47] J. Poon and V. Buterin, "Plasma: Scalable Autonomous Smart Contracts," Aug. 2017. [Online]. Available: http://plasma.io/plasma-deprecated.pdf

[48] G. Wood, "Polkadot: Vision for a Heterogeneous Multi-chain Network," Oct. 2016. [Online]. Available: https://polkadot.network/PolkaDotPaper.pdf

[49] P. Labs, "Filecoin: A Decentralized Storage Network," Jul. 2017. [Online]. Available: https://filecoin.io/filecoin.pdf

[50] I. Miers, C. Garman, M. Green, and A. D. Rubin, "Zerocoin: Anonymous Distributed E-Cash from Bitcoin," in *2013 IEEE Symposium on Security and Privacy*, May 2013, pp. 397–411, iSSN: 1081-6011.

[51] E. Ben Sasson, A. Chiesa, C. Garman, M. Green, I. Miers, E. Tromer, and M. Virza, "Zerocash: Decentralized Anonymous Payments from Bitcoin," in *2014 IEEE Symposium on Security and Privacy*. San Jose, CA: IEEE, May 2014, pp. 459–474. [Online]. Available: https://ieeexplore.ieee.org/document/6956581/

[52] J. Benet, "IPFS - Content Addressed, Versioned, P2P File System," *arXiv:1407.3561 [cs]*, Jul. 2014, arXiv: 1407.3561. [Online]. Available: http://arxiv.org/abs/1407.3561

[53] H. Sanjuan, S. Poyhtari, P. Teixeira, and I. Psaras, "Merkle-CRDTs: Merkle-DAGs meet CRDTs," *arXiv:2004.00107 [cs]*, Apr. 2020, arXiv: 2004.00107. [Online]. Available: http://arxiv.org/abs/2004.00107

[54] M. Steichen, B. Fiz, R. Norvill, W. Shbair, and R. State, "Blockchain-Based, Decentralized Access Control for IPFS," in *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. Halifax, NS, Canada: IEEE, Jul. 2018, pp. 1499–1506. [Online]. Available: https://ieeexplore.ieee.org/document/8726493/

[55] S. Labs, "Storj: A Decentralized Cloud Storage Network Framework," Oct. 2018. [Online]. Available: https://www.storj.io/storjv3.pdf

[56] D. Vorick and L. Champine, "Sia: Simple Decentralized Storage," Nov. 2014. [Online]. Available: https://sia.tech/sia.pdf

[57] Y.-A. de Montjoye, E. Shmueli, S. S. Wang, and A. S. Pentland, "openpds: Protecting the privacy of metadata through safeanswers," *PLOS ONE*, vol. 9, no. 7, pp. 1–9, 07 2014. [Online]. Available: https://doi.org/10.1371/journal.pone.0098790

[58] P. Mechant, R. De Wolf, M. Van Compernolle, G. Joris, T. Evens, and L. De Marez, "Saving the web by decentralizing data networks? A socio-technical reflection on the promise of decentralization and personal data stores," in *2021 14th CMI International Conference - Critical ICT Infrastructures and Platforms (CMI)*. Copenhagen, Denmark: IEEE, Nov. 2021, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/9663788/

[59] G. Zyskind, O. Nathan, and A. S. Pentland, "Decentralizing Privacy: Using Blockchain to Protect Personal Data," in *2015 IEEE Security and Privacy Workshops*. San Jose, CA: IEEE, May 2015, pp. 180–184. [Online]. Available: https://ieeexplore.ieee.org/document/7163223/

[60] G. C. Schipper, R. Seelt, and N.-A. Le-Khac, "Forensic analysis of matrix protocol and riot.im application," *Forensic Science International: Digital Investigation*, vol. 36, p. 301118, 2021, dFRWS 2021 EU - Selected Papers and Extended Abstracts of the Eighth

Annual DFRWS Europe Conference. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666281721000159
[61] Carl and Jarrad, "The Status Network," 2017. [Online]. Available: https://status.im/files/whitepaper.pdf
[62] A. Kosba, A. Miller, E. Shi, Z. Wen, and C. Papamanthou, "Hawk: The Blockchain Model of Cryptography and Privacy-Preserving Smart Contracts," in *2016 IEEE Symposium on Security and Privacy (SP)*, May 2016, pp. 839–858, iSSN: 2375-1207.
[63] L. Su, X. Shen, X. Du, X. Liao, X. Wang, L. Xing, and B. Liu, "Evil under the sun: Understanding and discovering attacks on ethereum decentralized applications," in *30th USENIX Security Symposium (USENIX Security 21)*. USENIX Association, Aug. 2021, pp. 1307–1324. [Online]. Available: https://www.usenix.org/conference/usenixsecurity21/presentation/su
[64] K. Karagiannis, "Advanced Smart Contract Hacking," San Francisco, CA, Mar. 2019, https://www.rsaconference.com/Library/presentation/USA/2019/advanced-smart-contract-hacking. [Online]. Available: https://www.youtube.com/watch?v=IOUnhCTw6tE
[65] S. T. Ali, P. McCorry, P. H.-J. Lee, and F. Hao, "ZombieCoin: Powering Next-Generation Botnets with Bitcoin," in *Financial Cryptography and Data Security*, ser. Lecture Notes in Computer Science, M. Brenner, N. Christin, B. Johnson, and K. Rohloff, Eds. Berlin, Heidelberg: Springer, 2015, pp. 34–48.
[66] J. Moubarak, M. Chamoun, and E. Filiol, "Developing a k-ary malware using blockchain," in *NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium*, Apr. 2018, pp. 1–4, iSSN: 2374-9709.
[67] C. Patsakis and F. Casino, "Hydras and IPFS: a decentralised playground for malware," *International Journal of Information Security*, vol. 18, no. 6, pp. 787–799, Dec. 2019. [Online]. Available: https://doi.org/10.1007/s10207-019-00443-0
[68] C. Karapapas, I. Pittaras, N. Fotiou, and G. C. Polyzos, "Ransomware as a Service using Smart Contracts and IPFS," in *2020 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*. Toronto, ON, Canada: IEEE, May 2020, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/9169451/
[69] P. Goyal, R. Netravali, M. Alizadeh, and H. Balakrishnan, "Secure incentivization for decentralized content delivery," in *2nd USENIX Workshop on Hot Topics in Edge Computing (HotEdge 19)*. Renton, WA: USENIX Association, Jul. 2019. [Online]. Available: https://www.usenix.org/conference/hotedge19/presentation/goyal
[70] N. V. Keizer, O. Ascigil, I. Psaras, and G. Pavlou, "FLOCK: Fast, Lightweight, and Scalable Allocation for Decentralized Services on Blockchain," in *2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, May 2021, pp. 1–9.
[71] M. Wohrer and U. Zdun, "Architectural Design Decisions for Blockchain-Based Applications," in *2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*. Sydney, Australia: IEEE, May 2021, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/9461109/