Surfing the chain. An analysis of blockchain-based data storage solutions to the data availability problem of decentralized social networks

Abstract

Decentralized blockchain-based social networks (BBSNs) have emerged as a response to the lack of privacy and agency in current social media platforms. This paper addresses the most persistent challenge in current BBSNs: making data available to users at all times. Current data storage mechanisms in BBSNs are either unreliable, centralized, or lead to frequent outages. Several BBSNS have come up with innovate to address these issues. This paper discusses Cadros, Safebook, PeerSoN and Diaspora's data storage implementations, which all come with their unique set of limitations. This includes platforms being able to choose cloud service providers or letting users store data on centralized servers. These limitations generally lead to renewed centralization pressures which puts data back into the hands of a few large cooperations. A solution to this is provided by distributed data storage networks like Filecoin. On these types of platforms users are able to store their data in a decentralized way. While the prospects of implementing a data storage platform like Filecoin are enormous, automating transaction procedures or implementing new business incentives for nodes on the Filecoin network are challenges that need to be addressed first. Through the right protocol, platforms like Filecoin will provide BBSNs with a second layer decentralized data storage network (DDSN). The implementation of a DDSN has the potential to fulfill the original intention of BBSNs and provide a new and private means of navigating and creating modular metaverse spaces.

Keywords

Blockchain based social networks (BBSNS), Distributed data storage networks (DDSNs), Protocols.

Introduction

Current social media are notoriously susceptible to privacy breaches (Voorveld et al., 2018). A wave of criticism has forced Facebook to rebrand its efforts into what is now coined as the 'Metaverse'. In this immersive space, users can choose their looks and interact with one another in various worlds and in real-time. How this space is hosted in a way that maintains user privacy is an open question (Dionisio et al., 2013).

A possible solution to this is provided by blockchain-based social networks (BBSNs) (Guidi et al., 2018). Blockchain is a novel technology that builds information in blocks that are stored throughout a network (La Cava et al., 2021). The hope is that through this decentralization, emerging platforms like Mastodon, provide users with the ability to create their personalized communities, with their own rules on privacy. In a decentralized metaverse, companies and individuals would be able to provide their own spaces that other users can choose to move between. Data is owned by the individual rather than one enterprise (Dionisio et al., 2013).

However, with BBSN's increased use, new problems arise. As BBSNs are not based on a central entity, no one has access to all user data at the same time. Since a social network is characteristically subject to continuous updating, several problems emerge. Among the most prominent problems is what is referred to as 'Data Availability' (Guidi et al., 2018). When a user X be-friends a new user, user Y's profile should be made aware of this, as it may be important for them when browsing through new friends to make in the network. The data must be made available for user Y. However, unless user X and their friend user Y are online at the same, this information will not reach user Y. There is no central entity that can access and relay information whether a user is online or offline (Raman et al., 2019). This paper will provide a theoretical framework for solving the data availability issue by investigating current data storage models presented by Guidi et al. (2018), Fu et al. (2016) as well as Paul et al. (2014) and synthesizing their findings. The key objective of this research is to design a completely decentralized BBSN, that not only addresses a purely technical issue, but also a subject that is increasingly of social and legal relevance. This solution hopes to provide both robust storage and users with ownership of their data.

A synthesis of data storage options in BBSNS.

Currently, there are 3 major distinctions in how data is stored on decentralized social networks: DHT-based, SO-based, and external-resource-based (Guidi et al., 2018). Figure 1 schematically shows these three storage options.

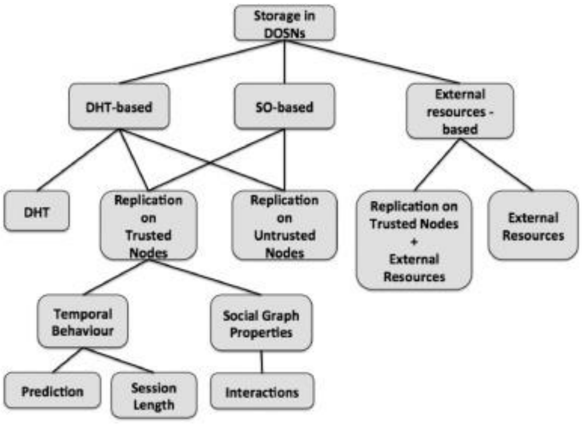
Storage in DOSNs
DHT-based
SO-based
External resources - based
DHT
Replication on Trusted Nodes
Replication on Untrusted Nodes
Replication on Trusted Nodes + External Resources
External Resources
Temporal Behaviour
Social Graph Properties
Prediction
Session Length
Interactions

Figure 1: Storage Methods for BBSNs *(Guidi et al., 2018)*

A Distributed Hash Tables (DHT) is a lookup table that can be used to bring data back from where it was stored. Ali et al. (2018) outline how DHT can be used to store all data on the blockchain. Using a DHT to store data among friends is currently the most popular approach among BBSNs. However, specifically for social media purposes this process is inefficient as the quantity of data to be stored is high. In addition, storing data on friend nodes is not reliable as those friends' nodes might be inaccessible by the network at nighttime (Guidi et al., 2018). To make data available at nighttime, social networks employ the other two data storage options outlined in figure 1: 'Social overlay' and 'External Resource Based'. Social overlays make use of untrusted nodes for the storage of data. However, with data replication on untrusted nodes, nodes quickly become overloaded and need to make use of external centralized cloud-based servers. Furthermore, a few nodes may become hubs hosting most of the other user's data in the network. This leads to renewed centralization pressures (Paul et al., 2014). Most BBSNs therefore use a hybrid approach of storing data on desktops as well as on centralized cloud servers (Shakimov et al., 2009). This however defeats the purpose of a decentralized social network as these 'external' servers are centralized.

A synthesis of different cloud-assisted approaches to BBSNs as well as their drawbacks.

To circumvent problems in using external resource-based solutions for BBSNs, various researchers have presented their own solutions. In Cadros, a scheme proposed by Fu et al. erasure coding is used to prevent the Cloud service provider from knowing the stored data's content. As a result, if the amount of data segments saved in a Cloud service provider's storage facility is fewer than a set amount, the Cloud service provider will be unable to reconstruct and know the original data. Cadros uses two techniques: complete replication which saves user x's data in the buddy circle, and erasure coding which breaks user x's data into smaller segments that are stored on the Cloud (Fu et al., 2014). However, data will remain in the hands of the platform which can selectively chose a cloud service provider.

In contrast, Cutillo et al. suggest Safebook, a peer-to-peer-based BBSN, in which each node is accessed through shells. The profile data is replicated and kept in what is referred to as the innermost shell, which is a subset of a node's direct contacts. The data retrieval process necessitates traveling the shells via a path of online nodes that are buddies with one another (Guidi et al., 2018). These methods emphasize how to store data copies such that they may be accessed even if users or specific friends of users are offline. However, they make the implicit assumption that the friends of user x will always be able to donate sufficient storage capacity to hold the duplicated data, which is not always the case (Shakimov et al., 2009).

Buchegger and colleagues proposed a solution to the problems associated with storing data on friend nodes. Their model is based on a two-tiered BBSNs architecture (PeerSoN). The DHT implementation of the first-tier acts as a look-up service. The user data is then stored on the second layer, which is made up of the network's users. Data about a user is distributed throughout the network. However, most users in a social network are not on the network to store data, but to interact with people. They do not receive rewards for keeping data on user x and as such lack the incentives to participate as a reliable source of data storage on the network (Buchegger et al., 2009).

An approach which circumvents the issues presented by storing data across the network, is presented by Diaspora, one of the biggest current BBSNs. Their major goal is to create a dependable and functional decentralized online social network. Diaspora's architecture is built on a client-server approach, with each user having their own server instance (Pod) for storage, communication, and access control. As there is no data or service replication, pods must be always available to provide dependable service. A Pod can be hosted on one's hardware or through a third-party service (cloud service). The data is saved on the pod in an unencrypted format and is safeguarded by an access control mechanism (Paul et al., 2014). However, if the user decides to store their data on one of the currently available services, such as Amazon Web Services, they again run into the risk of losing their privacy, and agency over their data. This paper therefore proposes combining the approach from either Cadros, which pertains to an element of cloud storage, or Diaspora, with the two-tier approach from PeerSoN. However, unlike PeerSoN, in which data is stored on the network's nodes, in this approach data is stored on a decentralized data storage platform, a new network layer made up of a completely different set of users. The splitting of the social interaction layer and the data storage layer are schematically shown in figures 2 and 3.

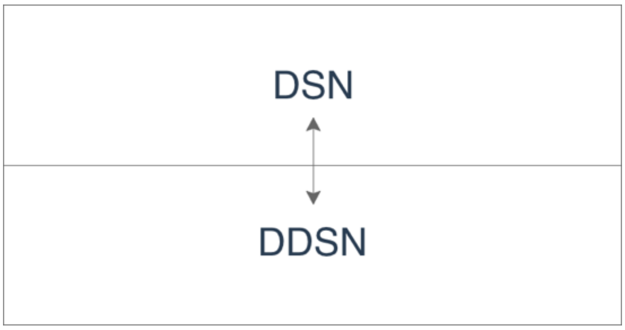

Figure 2: A decentralized social network layer (DSN) and decentralized data storage network layer (DDSN) would work hand in hand.

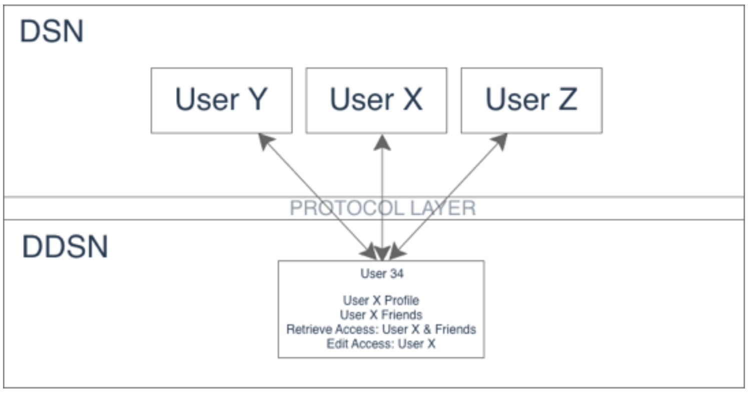

Figure 3: A decentralized data storage network layer could in theory provide data on User X at any time to anyone who is allowed to see it, providing a decentralized solution to current data availability problems.

Applying the DDSN using Filecoin

This DDSN framework provides a solution to the centralization pressures of external resource-based servers. In figure 1, such a DDSN aims to replace the external resource-based dependency of current BBSNs. An application of such a DDSN is Filecoin which is a digital platform in which users can store their data. Like Google Drive, users can upload content like videos or documents. Instead of a centralized server, this data is stored on a random node in the network, which offers up its storage space for the user. Contrasting Cadros, in which data is segmented and stored on a centralized cloud storage, with Filecoin, data is stored on a node in a decentralized network. The node can be any user who offers up their storage space. The user thereby becomes a freelance storage provider. When the user wants to retrieve their stored data, they pay the node a fee for the service. Unlike Buchegger et al.'s (2009) PeerSoN architecture, Filecoin provides economic incentives for users. It also relieves the concurrent social network layer to focus on interactions rather than data storage.

The protocol underlying a DDSN

Two of the major challenges with the implementation of Filecoin are privacy considerations for users and the DDSN's underlying fee system. With an underlying network protocol, these processes can be eased. Current BBSNs use different protocols, a set of rules used to determine not only where data is stored, but how it can be accessed. Such a protocol contains a permanent part that outlines rules spanning the entire network, and a personal part, which the user can adjust, the equivalent of a settings page on Facebook (Guidi et al., 2018). The personal part must for example specify that a user x's data can be retrieved by the user itself and friends who have reading access. This is because a user's trusted friend nodes may store some of user Xs data which can be accessed by them too (Rzadca et al., 2015). A user could then decide whether data on their profile can be added by their friends or other nodes in the network.

Furthermore, the permanent component of the protocol could provide a method to automate the detection of a node for data storage as well as the connected transaction process. Such protocol also extends to the regulation of the distribution of the data as well as its encryption methods. Such distribution of data is not limited to using only Filecoin but could also specify a combination of techniques. For example, storing data on user x on friend nodes during the day and transferring data to the DDSN at nighttime. This combination makes sure that data is constantly available. Encryption is then handled through Filecoin's exchange system. If networks want to increase security, the protocol could also include techniques such as erasure coding. A challenge then becomes making the process of retrieving and accessing data happen without latency. This is currently not possible with Filecoin. The primary reason for this is the manual-based transaction network. This is therefore the biggest limitation. To make a Filecoin like system feasible, more research needs to address a sustainable business model that can avoid charging users. Nodes providing storage require an incentive to uphold this storage in due shape. A monthly subscription model for users, or a system in which users do not pay but are shown advertisements, may be feasible.

Conclusion

To conclude, Filecoin is not yet a feasible alternative for a centralized server. It does provide the direction for new decentralized storage networks that integrate into existing BBSNs. As became clear in this investigation, current BBSNS frequently make use of unreliable data storage such as: users hosting themselves, or data replication on friend nodes, and external servers as backup. While companies such as Cadros have found creative solutions to providing users with the ability to own their data, these have come at the cost of reliability and scalability. To date solutions to the data availability problem of BBSNs are either unreliable or centralized. Hence, an underlying DDSN like Filecoin could relieve the data pressures of replication on the social layer of the network and provide a robust alternative for making data available to users in a decentralized way. With the emergence of the DDSN through the synthesis of different existing data storage models, as well as the discussion of a DDSN protocol, the technical implementation of a system like Filecoin seems feasible. However, more research needs to address the automatic implementation issues with Filecoin, which are characterized by business incentives for users in the DDSN. An approach where users pay for the privacy of their own data is realistic. Though, it may hinder platforms' ability to scale and capture most of the world's population. With the right business incentives, a DDSN has the power to make BBSNs fully decentralized. It also provides a way for users to maintain and own their privacy.

In a future metaverse, where a user may instead of interacting with their friends, interact with companies offering experiences, the user can decide what data to provide on a one-time basis, and the protocol behind the BBSN can determine in what way this data can be viewed. In the same way that users concurrently can opt-out of providing apps with the ability to track their movement throughout and beyond the platform, in a decentralized future, users will be able to do the same for experiences or services offered by companies like Facebook. If done right, the implementation of a DDSN has the potential to fulfill the original intention of BBSNs, provide a new and private means of navigating and creating modular metaverse spaces and instill a sense of hope that the war on privacy and self-agency in a world of digitalization is not yet lost.

References

Ali, S., Wang, G., White, B., & Cottrell, R. L. (2018). A blockchain-based decentralized data storage and access framework for pinger. 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE). https://doi.org/10.1109/trustcom/bigdatase.2018.00179

Buchegger, S., Schiöberg, D., Vu, L.-H., & Datta, A. (2009). Peerson: P2P Social Networking. Proceedings of the Second ACM EuroSys Workshop on Social Network Systems - SNS '09. https://doi.org/10.1145/1578002.1578010

Cutillo, L. A., Molva, R., & Onen, M. (2011). Safebook: A distributed privacy preserving online social network. 2011 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks. https://doi.org/10.1109/wowmom.2011.5986118

Dionisio, J. D., III, W. G., & Gilbert, R. (2013). 3D virtual worlds and the metaverse. *ACM Computing Surveys*, 45(3), 1–38. https://doi.org/10.1145/2480741.2480751

Fu, S., He, L., Liao, X., & Huang, C. (2016). Developing the cloud-integrated data replication framework in decentralized online social networks. *Journal of Computer and System Sciences*, *82*(1), 113–129. https://doi.org/10.1016/j.jcss.2015.06.010

Fu, S., He, L., Liao, X., Li, K., & Huang, C. (2014). Analyzing the impact of storage shortage on data availability in decentralized online social networks. *The Scientific World Journal*, *2014*, 1–14. https://doi.org/10.1155/2014/826145

Guidi, B., Conti, M., Passarella, A., & Ricci, L. (2018). Managing social contents in decentralized online social networks: A survey. *Online Social Networks and Media*, *7,* 12–29. https://doi.org/10.1016/j.osnem.2018.07.001

La Cava, L., Greco, S., & Tagarelli, A. (2021). Understanding the growth of the Fediverse through the lens of Mastodon. *Applied Network Science*, *6*(1). https://doi.org/10.1007/s41109-021-00392-5

Paul, T., Famulari, A., & Strufe, T. (2014). A survey on decentralized online social networks. *Computer Networks*, *75*, 437–452. https://doi.org/10.1016/j.comnet.2014.10.005

Raman, A., Joglekar, S., Cristofaro, E. D., Sastry, N., & Tyson, G. (2019). Challenges in the decentralised web: The Mastodon Case. *Proceedings of the Internet Measurement Conference*, 217–229. https://doi.org/10.1145/3355369.3355572

Shakimov, A., Varshavsky, A., Cox, L. P., & Cáceres, R. (2009). Privacy, cost, and availability tradeoffs in decentralized osns. *Proceedings of the 2nd ACM Workshop on Online Social Networks - WOSN '09*. https://doi.org/10.1145/1592665.1592669

Voorveld, H. A., van Noort, G., Muntinga, D. G., & Bronner, F. (2018). Engagement with social media and social media advertising: The differentiating role of platform type. *Journal of Advertising*, *47*(1), 38–54. https://doi.org/10.1080/00913367.2017.1405754