A Strategic Analysis on Decentralization Methodology and Data Privacy Transferability from Centralized Data Privacy

Ecem Sözeri, Furkan Özgültekin
Faculty of Computer Engineering
Koç University
Istanbul, Turkey
esozeri21@ku.edu.tr, fozgultekin19@ku.edu.tr

***Abstract*— Online social networks have reached enormous user numbers in recent years. The main reason behind such popularity is social network's ability to provide users with a platform to connect with family, friends, and colleagues. Users interact with each other by sharing their personal and sensitive information on social networks. Sensitive data shared on social networks spreads very quickly, which makes these platforms very attractive for attackers to gather information about users. This situation raises privacy and security concerns. In this paper, different privacy and security measures taken by social networks are examined comparatively. This paper presents the unique security and privacy design methods and challenges of OSNs. We examined the centralized method, which is one of the primary methods used by social networks, with its advantages and disadvantages. We also inspected the industry standards and new proposals in the area of decentralization for social media applications, to establish semi-congruent trust to users along with comparisons on how theoretical data privacy and security is currently employed for centralized platforms and how they can be transferred to decentralized platforms, both cryptographically and logically. In the final, we summarized the article with the problems and solutions of decentralized methods applied to solve the problems of centralized strategies.**

I. Introduction

A social network consists of entities and connections between entities. Entities in social networks represent people or institutions, and connections define their relationships [5]. Social networks such as Facebook, Twitter, Instagram, and LinkedIn can be used for personal or business reasons. These networks allow people to talk to each other, share information, and socialize. As of December 30, 2020, there are approximately 4 billion users on the internet. However, there are 2.7 billion monthly active users on Facebook, 330 million on Twitter, and 320 million on Pinterest [4]. The information shared by social network users has recently brought up data privacy and security issues. Because OSNs have such extensive sensitive data, it has attracted the attention of attackers. Over time, attackers began to see OSNs as platforms to obtain users' sensitive data and use them for marketing advertising or threats. For this purpose, different methodologies have been put forward for data privacy and security of social networks.

With the inception of social media applications in the last couple of years, individuals have grown more keen on publishing and sharing their personal data to singular hosted services which employ data storage techniques to establish privacy and security. But in terms of the current conditions one major unestablished gray area that hasn't been addressed to the masses has been trust. Whilst there are many theoretical employments of establishing provable trust, the major setback has been in both implementation and usability. Decentralization, both partial and totalitarian ways have been proposed for establishing trust within social media and social networking protocols and some applications have been implemented but still have a long way to go. In this article we try to address the positives and negatives of both centralized and decentralized social networking; taking into consideration up to date proposals and implementations of decentralization for establishing trust, Older techniques for establishing data privacy and security on centralized platforms and how they can possibly be transferred to decentralized social media and what the industry standard is currently at in terms of "decentralized" social media.

The rest of this paper is organized as follows. We first discuss the requirements and weaknesses of centralized methods that address privacy and security issues in social networks in Sec. 2. We then elaborate on the emerging decentralized techniques to eliminate the shortcomings of centralized methods and discuss the problems that decentralized methods also have.We examine solutions in the literature for the issues of decentralized methods in Sec. 3. Finally, we conclude remarks and future research.

II. Problem

The fact that social networks have a lot of users and these users share a lot of sensitive data with each other has made these platforms attractive for attackers. Therefore, ensuring the privacy and security of users' data has been one of the critical problems for OSNs [5]. Different methods are developed to ensure data security and privacy in OSNs. Confidentiality is essential to ensure privacy and security in social networks and encryption is a primary method for protecting confidentiality and securing data [8]. Users' shared data should be encrypted

using a secure key for each relation. Also, the key must be securely shared among the connections with minimal cost. Various encryption schemes can be used for confidentiality and security of data. We can group these methods into three categories: centralized, decentralized and distributed. Today's most used social networks, like Twitter, Facebook, and LinkedIn, are centralized social networks [10]. All functionalities, like storage, maintenance, and access to OSN services, are enabled by OSNs. This architecture has some advantages. They are straightforward and easy to implement. With this method, also called client-server architecture, users can search for their lost connections again or, for example, can easily find their old-school friends. Because the central architecture can quickly bring other users, since all data is stored in one place, it will be straightforward to find similar people with methods such as data mining.

On the other hand, it has some disadvantages as all data is in one place, and all controls are in OSN providers. This architecture can easily suffer from all the drawbacks of centralized systems like single point of failure, denial-of-service (DoS) attacks [3]. In centralized social networks, users do not have full control over their data. Users have to trust social networks fully [2]. The current problem in centralized social media is trust, while there are theoretically proven applications to secure a users' data, there isn't an established trust mechanism, except for open sourcing code which whilst is transparent, is a risk on some grounds. The prevalence of this statement can be seen with incidents such as Facebook & Cambridge Analytica accusations, which has damaged the already minimal trust that mass social media providers present to users. Whilst legislation and regulations do exist on corporations to mitigate the fact of low trust, it still does not incentivize or establish transparency or give users tangent grounds to trust social media providers nor prevent suspicion.

Philosophically we have come to the conclusion that decentralization is the answer to establishing trust and incentivizing trustworthiness. But the current architectures have ups and downs as well with concerns on both cryptographically heavy methodologies performance capabilities and on the other spectrum, diversity based methodologies having a potential "51% attack" risk.

We have identified three main routes on decentralized applications that are as such:

a) *Diversity Based:* Applications such as Mastodon, work in the principle of Federal Diversity, where any individual has the freedom of hosting within the Mastodon domain and sharing bulk data between different Mastodon servers via. ActivityPub. This methodology allows Mastodon to establish looser ground rules but also gives the freedom for individuals to also apply their ruleset for moderation and for data storage.

b) *Blockchain Based:* Research on Blockchain based applications allow for theoretical applications in terms of signing and distributing data storage between peers in a stricter manner, leading to a coarse distribution of data, resulting in confirmation of data validity to be by either Proof of Work or Proof of Stake which can prevent data alterations. Depending on the use case this may be a positive or a negative but it can be said that cases like user alterations, deletion of social media activity etc. can be issues on a blockchain based application.

c) *P2P Node Based:* Each application user is treated like a data Node on a Graph Database, and the application as a whole itself is the database, there are implemented libraries like Gun.js which attach to the front end of the application leading to a serverless fully decentralized application with cryptographic storage, disallowing users from accessing other users' data if stored on their device. Fault tolerance is an issue and data loss is potentially unavoidable in wide uses but with supporting owned Nodes this issue can be minimized for small to medium sized applications.

There are some concerns that need to be addressed to determine what sort of approach is suited for what use. To determine this we have used five main concerns to address the ups and downs of decentralized platforms.

Financial Concerns: For the widening of decentralized platforms there are financial concerns which may lead to small adoption or corruption of trust as a risk on platforms.

- Gas prices for Blockchain Based platforms can be a strain on the domain owner
- Server Hosting Costs for diversity based platforms can disincentivize individuals from hosting which will lead to a centralized result.
- P2P Node Applications would not necessarily be harmed in terms of financial costliness.

To address the issue on Federal Diversity applications we construct a strategic game between 2 Federally Diverse Social Media apps seeking to obtain hosts, the hosts seek to be able to host as low cost as possible, the two strategies present in this game are that the app providers can either give financial incentive to hosts, or not do so. It can also be stated that this could be the case since Mastodon, the largest decentralized social media has 10% of its instances serving more than half of its users [1].

TABLE I. FEDERALLY DIVERSE STRATEGY

	App 2	
App 1	***Give Incentive***	***Not Give Incentive***
Give Incentive	Equal Adoption	Adoption on App 1
Not Give Incentive	Adoption on App 2	Equal Adoption (lower overall)

Method Transferability: From centralized platform applications to decentralized applications, methods such as salting, differential privacy can be transferable in some architectures but possibly different methodologies may be needed to establish secure authentication and authorization.

- For Blockchain based applications, smart contracts need to be established rigorously due to them being immutable, which leads to potentially more security but if implemented the right way.
- For Diversity based applications, the issue is less present since the applications support a client-server architecture with only the option of multiple federations the user can register from.
- For P2P applications, signature based methods can allow for fast authentication but problems may exist on data migrations and loss of data, which may disallow user based applications.

Fault Tolerance: The system should be tolerant to downtime on individual data, or should mitigate for potential scenarios in which data might be lost temporarily, such as migration on peer shutdown.

Data Operations: Should be performant enough such that posting, deleting, following etc. should not be operations that take too long. Additionally the operations should relatively comply with ACID like principles on transaction management.

Ease of Adoption: The node system should be easy enough or be mounted to an already existing system such as Blockchain to be considered a fully decentralized application in practice. In some cases libraries such as Gun.js can embed the node system to the application itself which eliminates the users' requirement to be able to know how to host a server or a Blockchain node.

III. Solution

In the last couple years, research has been strong in terms of decentralized platforms, which try to establish and incentivise shared hosting of a central domain application. Which negates the need to trust only one owner's data storage methodology but does not eliminate it. Protocols such as ActivityPub and storage methodologies such as IPFS and Blockchain technologies have provided developers with language agnostic methods to implement decentralization, and with language specific libraries such as Gun.js which employs graph based P2P data storage, also provide a low cost alternative for decentralized data storage.

Examples of Work: Many proven work has been established to demonstrate small to medium size adoption of decentralized platforms. Some important work shows us that many of the protocols and theories can be applied for several use cases.

Xu et al. have demonstrated using IPFS and Blockchain to utilize immutability on authentication and use IPFS to handle posts tweet contents etc. for the purposes of a social media platform similar to twitter, they also mention the AKASHA project which is in practice which also supports a similar architectural design on the ethereum blockchain [6]. The work Xu et al. also shows how integration of different methods of decentralized protocols can work in synergy to provide a full scale application, they deploy smart contracts through the ethereum blockchain and use the Inter Planetary File System as a media content store which demonstrates strong suits of both Blockchain and IPFS (see Fig. 1).

Anandan et al. have implemented the Inter Planetary Broadcaster, a system Integrating IPFS and ActivityPub to allow easy access through Server to Server or Client to Server use cases without the need of IPFS installed to the Client Machine, the methodology uses IPFS to store files in nodes and hosts ActivityPub to communicate between owned nodes and "broadcasts" the stored IPFS data when needed to clients [7]. This sort of implementation may aid the ease of adoption on the client side, and also has the potential to be further implemented in Federal Diversity use cases, also could lower costliness for hypermedia data storage costs for any use case. The potential risk being in the implementation of the integration between ActivityPub and IPFS could potentially lead ActivityPub to be an open system and could lead potential adversaries to apply similar principles to Web Application Vulnerabilities or Network Vulnerabilities (see Fig. 2).

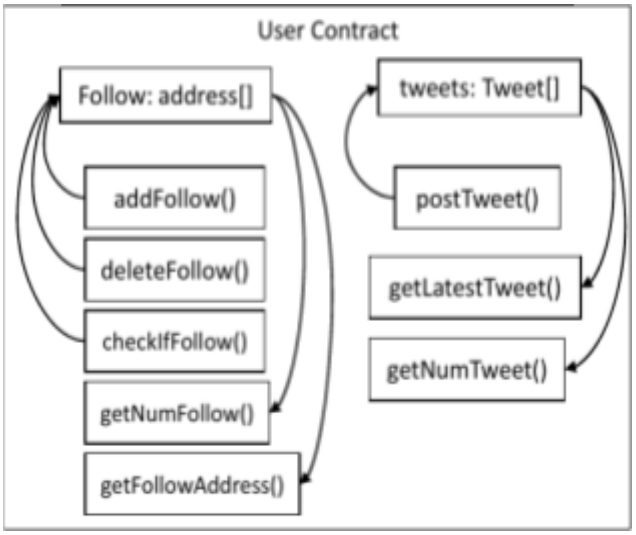

Fig 1. User Smart Contract Representation in Blockchain [2, Fig. 6]

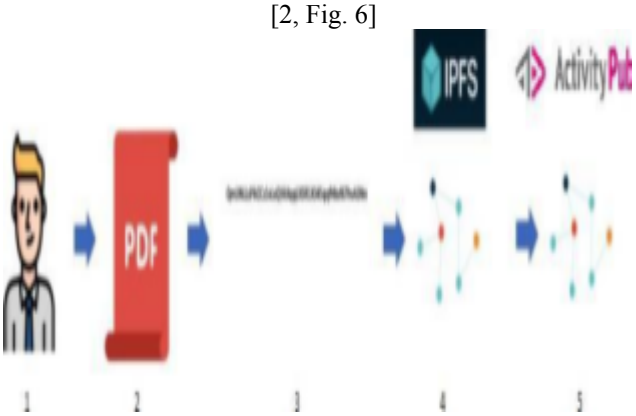

Fig 2. Work Flow Chart of Inter Planetary Broadcast [3, Fig. 1]

Seong et al. propose a more Granular Federal system when compared to ActivityHub but similar in implementation. Their proposed solution suggests personal or commercial "butler" services which regulate the access, distribution and removal of personal data on a more granular scale, with a supporting distributed database query language called SociaLite based on Datalog, which abstracts the complexities of the networking processes in distributed data sharing [9]. This granular system does allow the user to do a finer grained access control on a peer level and allows for the same web application principles to be applied in addition to potentially adding security against Request Forgery like attacks, but since a more granular access control is applied, if consideration to integrate a Client-Server architecture instead of a full Peer to Peer system as a social media app with butler support is present, might be issues on subject of adapting request structure appropriate for butler based access (see Fig. 3).

The Industry Standards: Tech Giants such as Facebook, Twitter, Youtube have alternatives in the decentralized space, Diaspora, Mastodon, Peertube respectively and potentially many more under development. It can be noted that Federally diverse platforms seem to show client support approximately totaling to 10 million users at the moment and these platforms have proven to be valid solutions to the current problem of trust with a couple problems, the main one being partial centralization (see 1st citation on page 1) [11].

Raman et al. have done a case study on the user and administrator behavior along with uptime and downtime analysis on the social media platform Mastodon, their results show that the trend line of Mastodon instances tend to fall into a plateau but the user popularity seems to exceed the existing instance amount (see Fig. 4) [1].

Fig 3. PrPI, Butler based Application Architecture [4, Fig. 1]

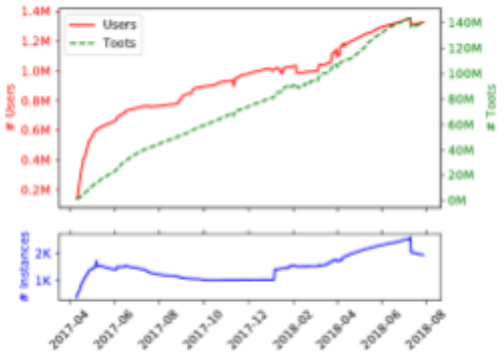

Fig 4. Work Flow Chart of Inter Planetary Broadcast [3, Fig. 1]

This sort of problem may be due to the incentivization problem we have proposed which may lead to a 51% like "monopolization" attack risk open to discussion in terms of trust but at scale may not be plausible unless Adversaries are able to obtain large financial compensation/motive to do so. They also mention and analyze this issue as a prevalence in terms of most used instances and whether or not they're being run by individuals or corporations and their properties, user activity and data being stored (see Table 2) [1].

While concerns leading to corporatizations in the decentralized space are present, individual based instances are also popularized which will potentially lead corporate instances to behave more carefully with sensitive data.

Protocols like ActivityPub have proven to be adopted at a wide scale with potential to further improve with integration to Blockchain smart contracts for more integrity and IPFS for complete decentralization. Currently speaking, ActivityPub is one of the most popular ways of supporting around 40 thousand decentralized nodes and most popular decentralized platforms seem to support this protocol to some extent for federated sharing [12].

TABLE II. USER TO INSTANCE DISTRIBUTION OF MASTODON [1, TAB. 2

Domain	Toots from Home Users	#Home Users	Users OD	Users ID	Toots OD	Toots ID	Instances OD	Instances ID	Run by
mstdn.jp	9.87M	23.2K	22.5K	24.7K	71.4M	1.94M	1352	1241	Individual
friends.nico	6.54M	8998	8809	23.3K	37.4M	2.57M	1273	1287	Dwango
pawoo.net	4.72M	30.3K	27.6K	15.4K	34.9M	1.4M	1162	1106	Pixiv
mimumedon.com	3.29M	1671	507	7510	435K	366K	420	524	Individual
imastodon.net	2.34M	1237	772	10.8K	2.37M	1.52M	711	865	Individuals (CF)
mastodon.social	1.65M	26.6K	24.8K	16.1K	30.9M	525K	1442	1083	Individual (CF)
mastodon.cloud	1.54M	5375	5209	106	7.35M	337	1198	39	Unknown
mstdn-workers.com	1.35M	610	576	12.5K	4.18M	1.98M	735	850	Individual (CF)
vocalodon.net	914K	672	653	8441	2.6M	853K	981	631	Bokaro bowl (A)
mstdn.osaka	803K	710	363	1.64K	2.68M	2.1M	561	862	Individual

Platforms leaning heavier on Blockchain to support both data and authentication like AKASHA can come at a higher cost due to gas prices which might put them on strain in the long term, but since the integration of IPFS and ActivityPub is demonstrated, it is also plausible that a potential synergy between Smart Contracts for authentication & sensitive data alterations and ActivityPub for federated sharing can lead to higher security against "51% attack" like risks since the larger node support of platforms like Ethereum and the migration from Proof of Work to Proof of Stake disincentivizing adversarial behavior on such Blockchain technology.

Additionally for granular access control if a butler like system can be adopted for a platform such as Mastodon, this can be able to further improve privacy and access control and potentially mitigate the current privacy concerns in lower diversity in terms of active users ratio to which instances most users use.

With the wide adoption of Mastodon, we have also spotted the presence of forks of the actual Mastodon repository such as Megalodon [7]. This allows for users to potentially be able to host and reinforce Mastodon by developing more secure alternatives under the same shared domain.

Use Cases: The three main ways we have identified demonstrates that some methods may be suitable for some applications while others may be more suited for other use cases.

- Blockchain based authentication and sensitive data handling can be utilized for more integrity
- Butler like granular access control can lead potential privacy concerns to be eliminated in Federally Diverse applications if integration with ActivityPub is plausible
- Fully P2P applications can work on small to medium scale with specific use cases where data loss is tolerable like private messaging, but with larger scale fault tolerance can be a concern in some use cases
- Federally Diverse applications can address some of their instance plateau issues via. incentivization (financial, community endorsements, support etc.)
- With forking, more security can be added freely to Federally Diverse applications.

Further on Security Concerns: There are still concerns over bot and spam accounts that are unestablished on the Fediverse. Since there are low regulations by its nature, If a user is known to be a spam account and is banned, they can migrate to another instance and still be under the same shared social network, or even potentially host their own instance. One such case is that of "botsin.space" wherein real users are allowed entry, but bots are encouraged to enter the federation [7]. Which in the future, may lead to a "bot safe haven" for spam purposes for Federally Diverse Social Networks.

IV. Conclusion

We took a dive into the world of decentralization through the already existing security of centralized means and took research on how applicable centralized methods can be and how capabilities of decentralization technology can further improve social network security, trust and availability. The solutions we found show promise whether it be blockchain technology for authentication, Federal diversity for establishing non authoritative data policies, eliminating the problem of a single point of failure and also many specific use cases being plausible in a peer to peer basis. Applications like fine grained access control can be addressed both in a centralized way and on a peer basis in a decentralized platform via applications such as the "data butler" system. The transferability plausibility from centralized systems to federally diverse systems seem applicable in most cases and the possible new use cases for P2P and blockchain based decentralized applications show promise in data integrity and granularity on a node basis.

Acknowledgement

This paper is an outcome of the term project of the Computer and Network Security COMP534 course at Koç University.

References

[1] A. Raman, S. Joglekar, E. De Cristofaro, N. Sastry, and G. Tyson, "Challenges in the Decentralised Web: The Mastodon Case *."

[2] Bodriagov, O., Buchegger, S. (2012). P2P Social Networks with Broadcast Encryption Protected Privacy. In: Camenisch, J., Crispo, B., Fischer-Hübner, S., Leenes, R., Russello, G. (eds) Privacy and Identity Management for Life. Privacy and Identity 2011. IFIP Advances in Information and Communication Technology, vol 375. Springer, Berlin, Heidelberg.

[3] C. Zhang, J. Sun, X. Zhu and Y. Fang, "Privacy and security for online social networks: challenges and opportunities," in *IEEE Network*, vol. 24, no. 4, pp. 13-18, July-August 2010, doi: 10.1109/MNET.2010.5510913.

[4] A. K. Jain, S. R. Sahoo and J. Kaubiyal, "Online social networks security and privacy: comprehensive review and analysis," Complex & Intelligent Systems, vol. 7, (5), pp. 2157-2177, 2021.

[5] P. Joshi and C. -. J. Kuo, "Security and privacy in online social networks: A survey," in 2011, . DOI: 10.1109/ICME.2011.6012166.

[6] Q. Xu, Z. Song, R. S. Mong Goh, and Y. Li, "Building an Ethereum and IPFS-Based Decentralized Social Network System," *IEEE Xplore*, Dec. 01, 2018.

[7] R. K. Anandan, S. K. P, S. K. Janahan, and K. Singh, "Improving Discoverability and Indexing of Interplanetary File system using Activitypub," *IEEE Xplore*, Jan. 01, 2022.

[8] F. Raji, A. Miri and M. D. Jazi, "CP2: Cryptographic privacy protection framework for online social networks," Computers & Electrical Engineering, vol. 39, (7), pp. 2282-2298, 2013.

[9] S.-W. Seong *et al.*, "PrPl," *Proceedings of the 1st ACM Workshop on Mobile Cloud Computing & Services Social Networks and Beyond - MCS '10*, 2010

[10] Sharma, N., S. Yadav, and B. Bohra. "A Review on Data Encryption Techniques Used for Social Media on Internet." *Advanced Computational Engineering and Networking* 4.9 (2016): 64-68.

[11] T. Federation, https://the-federation.info/, 2023