This document is an early draft of an unfinished idea; feedback is appreciated.

Emerald: Leveraging Blockchain in a Heterogeneous Service Network for Imageboards

Alexander Sellström
alexander@sellstrom.me

January 29, 2025

Abstract

This paper introduces a new way to build peer-to-peer social media platforms. Leveraging blockchain technology and zero-knowledge cryptography to overcome the limitations of existing decentralized social media, such as censorship, privacy, and scalability. Imageboards are uniquely suitable for such a decentralized application due to their bounded storage requirements. The technical challenges of creating such a decentralized application lie in the storage, distribution, and retrieval of content throughout the network, as well as the incentive mechanisms and governance rules for the network participants. This thesis proposes a novel protocol that combines content-addressable networks and proof-of-stake consensus to achieve a secure, scalable, and censorship-resistant decentralized imageboard platform.

1 Introduction

The advent of mainstream social media platforms has brought about a revolution in communication, enabling individuals to connect, share ideas, and engage with content on an unprecedented scale. However, the centralized nature of these platforms has given rise to a host of social challenges. Centralized platforms have wielded significant power over the flow of information, leading to concerns surrounding censorship, privacy breaches, and the legal responsibilities of these platforms. Moreover, the centralized architectures introduce technical constraints and vulnerabilities, such as single points of failure and susceptibility to various cyberattacks.

In response to the limitations of centralized social media, there is a burgeoning movement toward decentralized social media platforms. These emergent solutions aim to circumvent the pitfalls of centralization by distributing control and infrastructure across a peer-to-peer network. By empowering users and communities to govern their own digital spaces, decentralized platforms hold the promise of fostering greater freedom of expression, privacy, and resilience against censorship.

However, the transition to decentralized social media is not without its own set of challenges. Scalability and performance are two key areas where decentralized platforms often struggle to match the responsiveness and reliability of their centralized counterparts. The protocols powering these decentralized solutions must grapple with the technical complexities of efficiently routing content and managing distributed storage across a network of nodes.

In the realm of decentralized social media, blockchain-based projects [1, 2] have emerged as a promising alternative. These platforms leverage the distributed ledger technology of blockchains to create censorship-resistant and transparent social networks. Yet, the protocols underpinning these blockchain-based social media projects often lack adequate incentives for nodes to reliably distribute content to end-users, resulting in performance issues. In addition, these protocols are plagued by frustratingly poor performance that can be attributed to their reliance on a Distributed Hash Table (DHT) to find nodes that can share the data the client is looking for. To address these shortcomings, some platforms have opted to integrate centralized gateways to facilitate content delivery, potentially reintroducing the very issues that decentralization sought to eliminate, such as single points of failure and control.

Anonymity stands as a cornerstone feature of imageboards, serving as a safeguard against the suppression of free speech and providing protection against legal repercussions for illicit expressions. The blockchain domain has seen the development of protocols like Z-cash [3] and Monero [4], which leverage zero-knowledge cryptography to ensure transactional anonymity. Similarly, the Waku messaging protocol employs zero-knowledge cryptography to maintain user anonymity and implement rate limiting. The potential application of such cryptographic advancements to anonymous forums is an area ripe for exploration, with the possibility of enhancing privacy and freedom of expression in the digital realm.

2 Blockchains

The concept of blockchain technology emerged from the Bitcoin white paper published in 2008 [5]. A blockchain is a type of distributed ledger that records data in blocks, which are sequentially connected to each other by cryptographic hashes. The data stored in the blocks usually consists of financial transactions of a cryptocurrency, such as the identities of the sender and receiver, and the amount transferred. The blockchain grows by appending new blocks through a consensus mechanism that depends on the specific blockchain design. Blockchains are mainly used for decentralized financial systems today, but they can also store any kind of information, enabling a wide range of applications to be developed using the technology.

2.1 Consensus

Consensus is the mechanism by which new blocks are added and thereby advancing the state of the application(s) that runs on the blockchain. There are different types of consensus algorithms, such as Proof of Work (PoW) and Proof of Stake (PoS), that have different advantages and disadvantages.

PoW is the consensus algorithm used by Bitcoin [5] and other blockchain networks. It requires nodes in the network to solve a computationally expensive puzzle to be allowed to author a new block. This process is referred to as mining, and the nodes that carry it out are referred to as miners. Miners are rewarded for mining a block and collect additional rewards in the form of transaction fees paid by the users. Due to the limited space in a block, miners are incentivized to include the transactions that pay the highest fees. PoW provides a high level of security and decentralization, but it also consumes a considerable amount of energy that leaves a devastating carbon footprint [6].

PoS algorithms aim to address some of the drawbacks of PoW. They do not require miners to solve puzzles, but instead select validators based on how many coins they stake or lock up as collateral. The way in which validators take turns proposing blocks depends on which specific PoS algorithm is being used, but is typically either round robin scheduling, pseudo-random selection, or proposer-builder separation.

3 Decentralized Storage

One of the main challenges of blockchain technology is scalability, which refers to the ability of a system to handle a growing number of transactions or users. Storing large amounts of information on a blockchain can negatively affect its scalability, as it increases the size of the blocks and the network bandwidth required to propagate them. For these reasons it is best to store and deliver larger amounts of data using specialized decentralized storage protocols and only store hashes of files on the blockchain.

One such protocol is the InterPlanetary File System (IPFS) [7], which is a distributed file system that uses content-addressable storage as its primary data structure. Content-addressable storage means that data is identified by its content rather than its location, using a unique identifier called a Content Identifier (CID). This allows IPFS to retrieve data from any node in the network that has it, without relying on centralized servers or intermediaries. IPFS also uses a Distributed Hash Table (DHT) to store and lookup CIDs and their corresponding network addresses, enabling routing and discovery of data. A DHT lookup must be done for every file to be downloaded, which can take several minutes[citation goes here]. This added latency makes for a jarring user experience. In addition, IPFS does not have a built-in incentive mechanism to encourage nodes to store and share data, which may limit its availability and reliability.

Other protocols, such as Swarm [8], aim to address these limitations by providing a decentralized storage and communication service that is compatible with the Ethereum [9] blockchain. Swarm uses a similar content-addressable storage model as IPFS, but with some differences in how CIDs are generated and stored. Swarm also introduces an incentive system based on peer-to-peer accounting and payments, which rewards nodes for storing and serving data, and penalizes them for failing to do so. Swarm also offers features such as encryption, erasure coding, mutable resources, and feeds, which enhance the security, resilience, and functionality of the system. Swarm also suffers from high latency due to long lookup times.

4 Imageboards

Imageboards are online platforms where users can post and discuss images anonymously. They are different from other types of forums or social media, as they do not require user registration, and do not store user profiles or histories. Imageboards are characterized by their topical boards, which focus on specific themes or interests, such as anime, video games, politics, hobbies, etc. Users can create new threads by uploading an image and adding a comment, and other users can reply with images or text. Imageboards have a finite number of threads that can be active simultaneously. When a user creates a new thread, the oldest inactive thread is deleted. This deletion mechanism makes imageboards dynamic and transient in nature, but also reduces the storage needs compared to most other types of social media. With such modest storage requirements, in the order of two hundred gigabytes, an entire board could easily fit on a single node in a peer-to-peer network.

5 General Structure

On Emerald, the service providers (back-end servers) can be consumer-grade desktop computers belonging to any private person, as opposed to a corporate data center. The blockchain itself only stores the hashes of media posted on Emerald due to bandwidth constraints inherent to blockchain consensus algorithms. The media itself is stored by service provider nodes that make the content available to end-users for a small fee.

Emerald distinguishes itself from most other blockchain projects in that it has an application-specific blockchain. Other blockchains are often general-purpose blockchains that allow smart contracts to be deployed on-chain, enabling all kinds of decentralized applications to be built on top of them. Smart contracts increase complexity and attack surface, which exposes users to risks such as fraudulent or buggy smart contracts.

6 Posts and Threads

Posts and threads are identical in the way they're represented on-chain, but the first post with a certain thread identifier is treated as the original post by the application logic. A post is a transaction type that simply contains a thread identifier and a content identifier (CID).

The content identifier is a hash of the combined hashes of the text and any attached media files, also known as a root hash. A root hash of a hash list is used instead of the individual hashes to save precious bandwidth during the consensus process.

7 Content Storage and Delivery

Media in posts, such as text and images, are not stored on the blockchain. The blockchain merely stores content identifiers for the content which is handled by a content-addressable decentralized content delivery network (dCDN) similar to IPFS, which also uses libp2p. Blockchain full nodes that also want to be able to receive and broadcast media must also be network participants in the dCDN.

The dCDN differs from IPFS in that it does not use a distributed hash table (DHT) to find nodes. DHT lookup is a slow process involving finding a node that possesses a file so that it can voluntarily share it with you. Emerald's approach makes each node in the dCDN store and serve all of the files of a board. Clients only need to find service nodes once. When new CIDs are added to the blockchain, the client can simply request the corresponding files from the service nodes they're connected to.

The network is divided up into sub-networks, one for each board so that file nodes are not forced to handle data they are not interested in. This allows nodes with limited disk space to provide services. A defining characteristic of Imageboards is that threads disappear when new ones are created. A limit on the maximum number of active threads allows for very modest storage space requirements for service nodes in the range of 50-150 gigabytes for one board. In addition, nodes can opt out of hosting boards that are known to contain illegal or otherwise objectionable content.

After a post has been included on the blockchain, the node(s) that originally broadcast the transaction can start broadcasting the corresponding files to their neighbors that are also part of that board's dCDN subnetwork. Nodes are kept honest by the threat of being blocked by their neighbors if they do not at least offer to transmit a file that their neighbors have seen.

8 Service Contracts

If a client wishes to employ a service node a service contract is created. The service contract is cryptographically signed by both parties prior to being broadcast to show that both parties are consenting to the deal. The contents of the contract include the public keys of both parties, an expiration date, and a number of tokens held in escrow by the contract itself.

The client will send cryptographically signed messages, stating that they have received service, to the service node at regular intervals throughout the duration of the contract. The application logic dictates that the service node can broadcast these signed messages to collect the balance held in escrow. Each message works like a signed check for a fraction of the balance. All of the "checks" can be cashed using a single transaction after the contract is over.

The application logic forbids anyone other than the service node from withdrawing the tokens held in the contract to prevent the end-user from taking back the money. If the end-user's client decides that the node is no longer performing acceptably they can simply stop sending them "checks". The remaining balance in the contract is burned some fixed time after the expiration date of the contract.

With this system, malicious end-users are financially penalized if they enter service contracts that they don't intend on paying for. A rating function that takes unfulfilled service contracts into account can be used to avoid selecting malicious service nodes.

When interacting with other blockchains like Ethereum, users that cannot run their own node must typically rely on centralized API services. This defeats much of the purpose of using a decentralized network like Ethereum. In August 2022, the US Treasury put a ban on the Tornado Cash smart contract on Ethereum. This caused US-based Ethereum API service providers like Infura to block users all over the world from using it. Such censorship would not work very well on Emerald, since the user's client can simply rent a new service node if one starts acting up.

9 Decentralized Moderation

A complete lack of moderation is unlikely to result in a usable board, but appointing administrators with absolute authority to silence users reintroduces the very same problems inherent to centralized platforms that Emerald is supposed to solve. A solution to this would be to have a permissionless jury that functions in a similar manner to the consensus mechanisms that power the blockchain.

Users willing to help moderate the board and earn (governance) tokens for doing it can stake some tokens and vote on the legality of reported posts. These jurors will cast their encrypted vote and reveal the encryption key after the voting period has ended. Votes are encrypted during the voting period to prevent bots from simply copying the majority. The majority of the voting power decides the fate of the reported user and the post in question.

If a post is marked as deleted, the service nodes are no longer obligated to serve the file to their neighbors and clients. Jurors that voted against the majority are financially penalized by having a percentage deducted from their stake. This mimics how a proof-of-stake blockchain deals with malicious validators. The protocol awards jurors.

Scalability issues arise when the number of jurors increases. One way to address this is to only let the top n stakers vote, mirroring how proof-of-stake works. Another solution would be to pseudo-randomly select jurors for each trial. The seed for pseudo-random selection can be derived from some globally deterministic seed, such as the hash of a recent block.

10 Rate Limiting Nullifier

A defining characteristic of imageboards is that all users are anonymous and do not have accounts. If users use the same identity for many posts they may end up "doxing" themselves. Simply generating a new key pair and transferring the tokens to the new address would leave an obvious paper trail leading back to the original one. A solution to this is to use a rate limiting nullifier[10] in place of a transaction fee.

References

[1] A. Labs, “Frequency,” accessed: 2024-03-31. [Online]. Available: https://www.frequency.xyz/

[2] Subsocial, “Subsocial,” accessed: 2024-03-31. [Online]. Available: https://subsocial.network/

[3] D. Team, “Dat protocol,” accessed: 2023-11-24. [Online]. Available: https://dat-ecosystem.org/

[4] M. C. T. Nicolas van Saberhagen, “Monero,” accessed: 2023-12-14. [Online]. Available: https://www.getmonero.org/

[5] S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system,” Dec 2008, accessed: 2024-02-12. [Online]. Available: https://bitcoin.org/bitcoin.pdf

[6] A. De Vries, U. Gallersdörfer, L. Klaaßen, and C. Stoll, “Revisiting bitcoin’s carbon footprint,” *Joule*, vol. 6, no. 3, pp. 498–502, 2022, accessed: 2024-02-12. [Online]. Available: https://www.cell.com/joule/pdf/S2542-4351(22)00086-1.pdf

[7] P. Labs, “IPFS,” accessed: 2023-11-24. [Online]. Available: https://ipfs.tech/

[8] S. Foundation, “Swarm,” accessed: 2023-11-24. [Online]. Available: https://www.ethswarm.org/

[9] E. Foundation, “Ethereum,” accessed: 2024-02-13. [Online]. Available: https://ethereum.org/

[10] A. Revuelta, S. Tikhomirov, A. Challani, H. Cornelius, and S. P. Vivier, “Message latency in waku relay with rate limiting nullifiers,” Cryptology ePrint Archive, Paper 2024/1073, 2024. [Online]. Available: https://eprint.iacr.org/2024/1073