Persistent Addressability in Alternative Social Networks

PhD Thesis Exposé
by
Paul Fuxjäger

Supervisor:
Univ. Prof. Dipl.-Math. Dr. Peter Reichl
Faculty of Computer Science
University of Vienna

Abstract

Federated social media systems offer the promise of a more transparent and democratically legitimated distribution of moderation power (and responsibility) than their centrally managed counterparts. Motivated by this promise, numerous projects based on the paradigm of an open peer-to-peer network of independent server instances have been developed within the global open source community in recent years.

Evidence collected in these experiments suggests that loss of social cohesion due to polarisation effects, or the abuse potential due to amplification of dis-information and demagogy, do in fact seem to be significantly less prevalent in these alternatively structured media ecosystems. However, these alternative networks have not been able to reach significant adoption outside of niche communities yet. This PhD thesis aims at analysing reasons for that and providing some conceptual and technological solutions, including the following contributions:

1. **Structural analysis of the main barriers towards mainstream adoption of federated social media services.** Based on the observations collected within the most successful project within this space to date - the social microblogging network called 'Mastodon', which inherited core concepts from Twitter - we present and substantiate the hypothesis that relevance outside of fairly closed cultural niches depends on *long term stability* of federated social media identities (DNS based addresses and established associations between them).
2. **Definition of a functionally separated architecture.** Since individual server instances are short lived in the absence of stable revenue from third parties, we propose to make use of a dedicated public ledger that maintains a registry of decentralised identifiers (DIDs) as a common reference layer - to keep DNS-based addresses/associations up-to-date whenever an instance fails to keep up or moderation conflicts between instances occur.
3. **Definition and proof-of-concept implementation of a message broadcast protocol** to maintain consensus on a permission-less ledger with *local scope:* In order to gain write access on the ledger (containing *locally registered* DIDs), all inhabitants within the given population area are invited to personally maintain a 'locally cooperating public node' (similar to a node in a citizen science sensor network). Each node periodically generates and provides attestations of over-the-air broadcast transmissions from neighbouring nodes, thereby providing a *proof-of-local-cooperation*, which serves as the basis for the consensus mechanism for the local public DID ledger.
4. **Performance analysis** of such a novel public DID registry network in the city of Vienna, based on the current SOTA of long range unlicensed wireless communication standards (e.g. LoRa). Drawing from the lessons learnt during 15+ years of voluntary activity within community operated wireless mesh networks (funkfeuer.at and battlemesh.org), we will try to validate whether the proposed concept (of using separate networks for identity hosting and content hosting) is suitable to increase long term stability of established relations within federated social networks.

Introduction

Interpersonal communication is increasingly moving away from shared public spaces and editorially supervised ‘legacy’ media systems - towards algorithmically optimised online platforms. It seems very likely that political opinions of future generations are predominantly shaped by experiences within such new media ecosystems [1].

However, the vast majority of currently established online social media platforms are detached from democratically legitimised institutions and discourses [2], and are often based on a manipulative business model: opaque filter algorithms which are applied to maximise dwell time. More and more evidence continues to mount that suggests that this reinforces polarisation and radicalisation effects within networked societies overall [3].

Regulatory measures (e.g. the EU Digital Services Act) constitute an important measure to limit the damage to social cohesion - but the oligopoly of global ‘attention shopping malls’ cannot be broken by regulatory initiatives alone. In order to get a sustainable grip on the current ‘crisis of trust’, a new kind of public infrastructure, based on alternative socio-economic model must be developed in parallel to regulatory interventions.

Lessons learnt from previous experiments in the area of federated social microblogging (e.g. Diaspora, StatusNet, Identica, Appdotnet) indicate that ‘software development and maintenance’ of such an alternative public infrastructure *no longer* constitutes the dominant requirement for achieving mainstream adoption. Software libraries have emerged, whose level of maturity is on par with the dominant advertising-revenue-dependent platforms. The majority of the maintenance effort is carried out by a large group of voluntary contributors within the global open source ecosystem.

The most successful example in that realm is the Mastodon network, which inherits fundamental UI concepts from Twitter. This network is based on a federated protocol called ActivityPub which is maintained by the World Wide Web Consortium (https://www.w3.org/TR/activitypub/), and as of 2022, this network comprises of ~4000 federated server instances hosting ~4 million active accounts worldwide (https://mastodon.fediverse.observer/stats). The mastodon stack (servers and clients) is currently maintained by around 800 developers, based on the AGPL License. Since the advent of federated social microblogging around 15 years ago, networks of that magnitude have not been observed before in that space.

However, relative to the dominant social microblogging platform operated by Twitter, the Mastodon network has not gained significant relevance outside of special interest niches of our society, yet. Hence, in this thesis I will explore key reasons for that and establish technical solutions which will resolve some of the key issues.

Problem Definition

Federated social media systems are fundamentally based on an open communication network paradigm:

‘Anyone shall be able to create an account on any server instance and be able to interact freely with all existing accounts on all other server instances within the network. Hierarchical relations between accounts and servers shall only be introduced when abuse cannot be handled in any other way.’ [4]

However, the recent history of federated microblogging networks tells us that mainstream relevance can only be achieved if the resources that are available to coordinate to limit abuse of the network (e.g. SPAM, Dis-Information campaigns) *scale with the number of active participants*.

In the starting phase of a network the open communication paradigm mentioned above works because the required moderation effort is minimal. This inevitably changes when the number of active users continues to increase: since no targeted advertising revenue (or VC funding that expects third party funding eventually) is available, the resources available for moderation work are limited to voluntary contributions of ‘cognitive surplus’ and individual financial donations to the moderation team.

The resulting problem:

If the moderation load on a given server cannot be handled by the moderation team responsible for that instance, all other instances that are receiving posts from offending accounts registered on the under-moderated instance will start to add it to the list of blocked servers (aka ‘de-federation’). As a consequence, **all users on a blocked instance need to find a new home** if they wish to stay connected to all other users within the network.

This causes established (DNS based) links to break!

I am aware that conflicts caused by differing moderation policies are inevitable. What I am addressing here is the problem of continuous address changes due to *lack of moderation resources.* In the current federated ecosystem, a small number of accounts (generating content that is perceived as offensive) can cause large communities of people to be continuously being cut off from each other. We argue that this presents a fundamental limit to the adoption rate, since utility of a network is growing nonlinearly with the number of addressable users (also known as Metcalfe’s Law [5]).

Furthermore, a small group of maliciously acting users is able to cause a partitioning of the whole network with low effort, this vulnerability to denial of service attacks is present in all social networks that are based on a federated model.

The problem of missing long term stability of DNS based addresses within the current network architecture was acknowledged by leading Mastodon developers during a panel discussion on the potential of integrating decentralised identity APIs within ActivityPub based networks [6]:

‘if you really care about your own identity … you have to operate your own server’

This statement seems undeniably true, as each new user that signs up in a federated social network needs to first make a decision about which instance to join, and all subsequent activities are subject to the moderation policies which are governed by the admin of that instance.

In such a scenario, the effort necessary to ‘operate your own server’ should to be as low as possible under current technical constraints in order for the network to be perceived a public utility. Ideally, it should be as low as operating a client device (e.g. keeping a mobile phone OS and installed applications up-to-date).

This motivates a ‘functionally separated’ architecture which splits the responsibility of hosting the social graph (which does not need to be moderated) from the responsibility of hosting the actual social media content.

Hence, we will investigate whether the application of such a separation principle is suitable to increase long term stability of individual social media addresses in the context of federated networks, thereby increasing the societal relevance beyond niche groups of society which are bound by common special interests.

Currently, due to predicted change of ownership of the dominant Twitter platform [7] - which is expected to result in a significant change of moderation policies - the popularity of the Mastodon network is growing faster than ever before (see https://mastodon.fediverse.observer/stats).

In order to get closer to a future in which alternatively governed social media networks are perceived as public infrastructure, ways to cope with current moderation dynamics need to be developed in order to be able to offer long term address stability that is on par with centralised platforms. This is the underlying motivation for the following research questions.

Research Questions

RQ 1: How to minimise effort to maintain long term control over one's own personal public social media address?

- Is a separate network - based on decentralised ledger technology - suitable to maintain a *persistent* public registry of associations between public keys (controlled by users) and DNS addresses (controlled by server admins), thus increasing long term stability of relations within a federated social network?

- Which of the already implemented DID methods (https://w3c.github.io/did-spec-registries/#did-methods) are suitable for mapping links between long term public keys to and short term DNS-based instance addresses?

<PublicKey> → DID Resolving Method → Name@<current-DNS-Address>

- Which DID methods offer private key rotation (in the event of loss/compromise) and what are the requirements on the DLT network to be able to support such schemes in a secure way?

RQ 2: Which mechanisms are available for maintaining consensus on a local DID registry?

- Which mechanisms (other than Proof-of-Work, Proof-of-Stake, Proof-of-Space-Time, or Proof-of-Authority) are available to maintain consensus on a public DID registry network that consists of nodes that are in physical proximity and thus able to receive wireless broadcast transmissions neighbouring nodes?

- Can a generic 'proof of local cooperation' be derived from mutual attestations of local radio broadcast transmissions within the local network?

- How can inclusivity be optimised when write access on the local DID ledger depends on the personal capacity of operating a node within the local network?

- Under which conditions shall local residents be able delegate broadcast transmissions to local node operators, enabling them to register a DID without personally controlling a node?

RQ 3: Which conditions are required to maintain robust consensus on a local DID registry that is based on proofs of local cooperation?

- What is the minimum density of honest participants needed to maintain consensus on the common state of the DID registry?

- What maximises resistance to SPAM and sybil attacks when node densities are critically low?

- In case of abuse: what motivates node operators to sustain a minimal amount of local moderating effort that is necessary in order to maintain consensus on list of banned DIDs?

- Which metrics are suitable for evaluating the robustness of the resulting governance model within a small scale local testbed?

Related Work

In key related work [8] an overview of socio-economic models for moderation of digital public spheres is presented. The authors argue that a digital communication service is generally perceived as ‘a public utility’ only if the governance model is legitimated by direct participation of users in a fully transparent fashion.

The concept proposed in this PhD project is closely related to the quest for robust and open methods to provide so called *proofs of personhood* [9] - which is about giving all real people inalienable digital participation rights independent of identity, including protection against erosion of their democratic rights through identity loss, theft, coercion, or fakery.

In fact, the aim of this project is to develop a method that is equivalent to organising large scale physical pseudonym parties, which are identified in [9] as being the only method capable of supporting the four pillars of digital democracy without compromise:

- Inclusiveness: Any human person should be able to participate, regardless of nationality, wealth, race, gender, connections, education, or expertise.
- Equality: All participants must be treated equally for democratic deliberation and decision-making purposes: i.e., ‘one person, one vote.’
- Security: Digital personhood must protect both individuals and the democratic collective from compromise in the digital and physical domains.
- Privacy: Digital personhood must guarantee each participant’s freedom to communicate, associate, and express their true intent in democratic processes.

Currently, most permissionless DLT networks are based on consensus mechanisms that are violating one or more of these fundamental requirements, and therefore do not satisfy the conditions necessary to be considered a public utility.

The term ‘health’ is often used in related work on governance processes, but it often remains undefined what ‘health’ actually means. A more detailed specification that seems useful for operationalisation in the context of this project is found in [10], and it is based on the concept of ‘sense of coherence’, which is derived from empirical studies of groups that have shown extraordinary robustness in the presence of stress.

The suitability of unlicensed low power transmission modes for long range machine to machine communication in urban contexts is empirically validated in [11]. In high node density scenarios, interference ultimately limits broadcast capacity, an empirically validated model for interference in such license-free multiple access networks is presented in [12].

A recent paper on conflict-free replicated data types (CRDT) [13] indicates that such a data structure may be suitable to maximise Byzantine fault tolerance for keeping a common list of banned DIDs.

Related projects (Helium [14] and FOAM [15]) make use of physical proximity and direction communication between network nodes to derive *proofs of location* as a foundational mechanism to maintain a decentralised, privacy preserving, censorship resistant infrastructure. This indicates that physical proximity and local radio communication links are relevant building blocks for network concepts with global scope that are able to support the four pillars identified above.

The groundwork for the concept presented in this PhD exposé has been developed while working on two software development projects in the area of decentralised identity and personal data stores. Both of these projects have received funding from NetIdee [16].

Summary

The goal of this PhD project is to define, implement and experimentally validate a generic over-the-air message broadcast protocol that enables an open group of locally cooperating nodes (embedded in a dense population area) to form a network that is able to maintain consensus on a ledger containing decentralised identities (DIDs) that have been *registered locally*.

In other words, we aim to conceptualise and build a proof-of-concept of a *public infrastructure* that is able to unambiguously answer DID-to-DNS resolution requests via public IP in the following from: “Has DID:<local coordinate>:<public key string> been registered during a given time window - and if so - what service endpoints (DNS address of AcitivityPub actor) are currently associated with that DID?”

Robustness against maliciously acting nodes is achieved by deriving voting privileges from each nodes individual capacity to provide (independently verifiable) evidence of having contributed to the maintenance of local consensus on the common history of locally observable broadcast events.

This project is based on the idea of ‘local permissionlessness’ which is derived from concepts that have already been successfully applied in community operated wireless mesh networks: *transparency, unlicensed transmission modes and local cooperation* (as formalised in the Pico-Peering Agreement https://www.picopeer.net).

Work Plan

	2022				2023				2024				2025	
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2
Literature Review	■	■	■	■	■	■	■	■	■	■	■	■		
Protocol Design		■	■	■	■									
Testbed Setup				■	■	■	■							
Collecting Data							■	■	■	■	■	■		
Evaluation & Postprocessing								■	■	■	■	■	■	
Publication and Dissemination			■			■				■			■	■

Publication Plan

2022/Q3: Conference paper on the functional separation concept

2023/Q2: Conference paper on the details of the local broadcast message protocol

2024/Q2: Conference paper on preliminary results from the proof-of-concept testbed

2025/Q1: Journal paper on the complete architecture, including final results collected from testbed experiments

2025/Q2: Submission of dissertation

References

[1] Christian Fuchs, “Culture and Economy in the Age of Social Media", Routledge, 2015 https://doi.org/10.4324/9781315733517

[2] Philip M. Napoli, “Social Media and The Public Interest: Governance of News Platforms in the Realm of Individual and Algorithmic Gatekeepers”, Telecommunications Policy, Volume 39, Issue 9, 2015, https://doi.org/10.1016/j.telpol.2014.12.003

[3] Shoshana Zuboff and Karin Schwandt, “The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power”, New York: PublicAffairs, 2019.

[4] Lucia La Cava et al, “Understanding the Growth of the Fediverse through the Lens of Mastodon”, Applied Network Science, 2021, https://doi.org/10.1007/s41109-021-00392-5

[5] Zhang et al, “Tencent and Facebook Data Validate Metcalfe’s Law”. Journal of Computer Science and Technology, Volume 30, 246–251 (2015). https://doi.org/10.1007/s11390-015-1518-1

[6] First International ActivityPub Developers Conference, APCONF 2019, https://redaktor.me/apconf/

[7] Mike Isaac and Lauren Hirsch, “Elon Musk Agrees to Buy Twitter”, The New York Times, https://www.nytimes.com/2022/04/25/technology/musk-twitter-sale.html

[8] Jillian C. York and Ethan Zuckerman, “Moderating the Public Sphere” ,Human rights in the age of platforms, 2019, https://doi.org/10.7551/mitpress/11304.003.0012

[9] Bryan Ford, “Identity and Personhood in Digital Democracy: Evaluating Inclusion, Equality, Security, and Privacy in Pseudonym Parties and Other Proofs of Personhood”, arXiv preprint, 2020, https://arxiv.org/pdf/2011.02412.pdf

[10] Lenneke Vaandrager and Lynne Kennedy, “The Application of Salutogenesis in Communities and Neighbourhoods”, The Handbook of Salutogenesis, Springer, 2017, https://www.ncbi.nlm.nih.gov/books/NBK435839/

[11] Michael Rademacher et al, “Path Loss in Urban LoRa Networks: A Large-Scale Measurement Study”, 2021, https://doi.org/10.48550/arxiv.2109.07768

[12] Paul Fuxjäger and Stefan Ruehrup, "Validation of the NS-3 Interference Model for IEEE802.11 Networks," 2015 8th IFIP Wireless and Mobile Networking Conference (WMNC), 2015, pp. 216-222, https://doi.org/10.1109/WMNC.2015.40

[13] Martin Kleppmann, “Making CRDTs Byzantine Fault Tolerant”, 9th Workshop on Principles and Practice of Consistency for Distributed Data (PaPoC ’22), 2022, https://doi.org/10.1145/3517209.3524042

[14] Amir Haleem et al, “Helium - A Decentralized Wireless Network”, Helium White Paper, 2018, http://whitepaper.helium.com

[15] Ryan John King et al, "FOAM – The Consensus Driven Map of the World”, FOAM White Paper, https://foam.space/publicAssets/FOAM_Whitepaper_May2018.pdf

[16] https://www.netidee.at/sovereignid and https://www.netidee.at/mastodonid