Identifying Roadblocks in Building Decentralized Apps

Tianyuan Yu
UCLA
Los Angeles, USA
tianyuan@cs.ucla.edu

Adam Thieme
UCLA
Los Angeles, USA
adam@cs.ucla.edu

Lixia Zhang
UCLA
Los Angeles, USA
lixia@cs.ucla.edu

Abstract

Bluesky is one of the well-known decentralized social apps with millions of users. However, measurement studies show that it may not have fully achieved the goal of decentralization, that is, putting all the control power of the app in the hands of end users. This paper takes Bluesky as a case study to investigate the reasons for the observed centralization and identifies two technical roadblocks in building fully decentralized apps.

CCS Concepts

• **Social and professional topics** → **Centralization / decentralization**; • **Information systems** → **Social networking sites**.

Keywords

Bluesky, Decentralized Social Networks

ACM Reference Format:
Tianyuan Yu, Adam Thieme, and Lixia Zhang. 2025. Identifying Roadblocks in Building Decentralized Apps. In *Applied Networking Research Workshop (ANRW 25), July 22, 2025, Madrid, Spain.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3744200.3744778

1 Introduction

As pointed out in a recent workshop report [4], Internet centralization started with the consolidation of apps. In recent years, many efforts have been made to develop decentralized applications, and some of them, such as Bluesky [5], have attracted millions of users. However, measurement studies [1] show that Bluesky may not have achieved their goal of being a fully decentralized social app, that is, putting all the control power of the app in the hands of the end users.

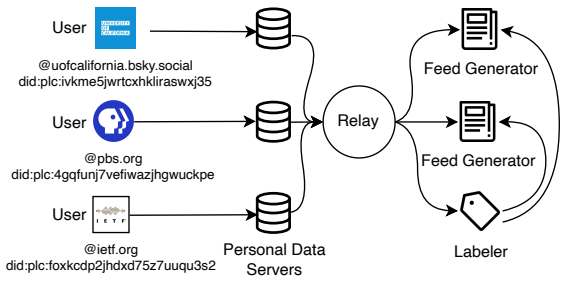

Figure 1: A Simplified Bluesky System Diagram

Taking Bluesky as a case study, this paper looks into the specifics of the Bluesky design to understand which parts of the app are not decentralized, and why. We first give a brief overview of the Bluesky design by describing the five major components of the Bluesky design in Section 2, and then discuss why two of these components are centralized in Section 3. In Section 4, we articulate the technical roadblocks of decentralization revealed by examining Bluesky's design and operations in building truly decentralized Internet apps.

2 How Bluesky Works

Fig. 1 shows the five major components of Bluesky.

DID for identity, DNS for display: Each Bluesky user is uniquely identified by a Decentralized Identifier (DID) [8] which is associated with a DNS name for display purposes (also known as *handle* in Bluesky). For users who do not have a DNS name (which are the majority), Bluesky assigns them one under its own domain *.bsky.social*, which the Bluesky PBC, an organization responsible for the development and maintenance of Bluesky, controls. Bluesky PBC operates a directory to map DIDs to their DNS names. However, [1] shows that 98.9% of the total 5,077,159 user handles are under *.bsky.social*, and Bluesky manages the binding between DNS and DID for most Bluesky users. There do exist operators (*e.g.*, swifties.social) who delegate domain names to Bluesky users as handles and manage their binding with DID, but they handle no more than a few hundred registered domain names. For example, *swifties.social* has 256 domain names

ANRW 25, July 22, 2025, Madrid, Spain

ACM ISBN 979-8-4007-2009-3/25/07
https://doi.org/10.1145/3744200.3744778

registered by Bluesky users. The study observed that most non-Bluesky user handles are registered with a few registrars. The top seven registrars provide more than 68% of the non-Bluesky users. NameCheap and Cloudflare have 20.94% and 11.46% of the total user handles, respectively.

PDS to store signed user data: One PDS can host multiple user repositories. Repositories store user data, such as posts and likes, in data structures. The user data is signed by the private key associated with the DID, and named with the URI scheme[1]. Personal Data Servers (PDS) manage user keys and sign the data the users produce. The Bluesky architecture is open to self-hosting PDSs, with PBC operating default PDSs for users. Since PDSs manage users' login credentials (keys) and personal data, users who don't want Bluesky to have the ability to lock them out of their account or delete their data have an incentive to host their own PDSs. We were not able to query the Bluesky relay, *bsky.network*, for details about the PDSs it subscribes to. Some anecdotes [6] suggest that Blueksy operates around 20 PDSs and there are more than 300 self-hosted PDSs[2].

Relays to aggregate PDSs: Bluesky uses a Relay to aggregate user data by gathering data from all connected PDSs to publish *Firehose*, a real-time stream of all user activities; it also keeps the data for three days. The architecture allows for multiple alternative Relays without a defined solution for interconnecting them. To date, Bluesky PBC operates the only Relay, resulting in only one Firehose which Feed Generators and Labelers pull from.

Feed Generators to create custom feeds: Any Bluesky user may create a Feed Generator by self-hosting a server that consumes the publicly-accessible Firehose to curate the feed with their content selection policies (*e.g.,* only include activities that contain particular hashtags). This curated feed can be accessed via a URL which is put in a special record in their repository. Other users can then subscribe to this publication from the Blusky website. After being introduced in May 2023, the number of Feed Generators has been steadily increasing. On 30 March 2024, Bluesky had 3,782 active Feed Generators, and analysis shows that they are contributed by diverse communities with different interests.

Labelers for content filtering: Any Bluesky user with a DID can serve as a labeler, which takes input from Firehose, labels individual user activities, and suggests the default action to be taken by other Bluesky clients. For example, *nogifsplease.bsky.social* is a Labeler that labels GIFs found in the Firehose. Bluesky initially launched one Labeler in 2023. Since 15 March 2024, when the Bluesky platform began allowing community Labelers, there has been rapid growth in the number of Labelers. In May 2024, the authors of [1] discovered 62 unique Labelers.

[1]For example, *at://did:plc:z72i.../app.bsky.feed.post/3k43tv4rft22g*

[2]No information about how many users are covered by the self-hosted PDSs.

3 Why Some Components are Centralized

Identity Management Centralization Although Bluesky defines DID as its official user identifiers, in reality, the app uses user handles, which are DNS names. Bluesky uses DNS names for user handles because it wants users to keep their handles and social graphs when migrating between PDSs; user handles should be globally unique and server-independent, which DNS names provide.

Since most Internet users do not have their own DNS names, Bluesky has to assign a DNS name to most users under its own domain "`.bsky.social`" and manages the names for the users. If all Internet users had their own DNS names, Bluesky would not need to be this central manager. Email addresses might seem an alternative to DNS names, since they uniquely identify Internet users, but they are subject to the control of email providers. Unlike DNS, changing your provider means that your email address also changes.

Relay Centralization Bluesky has seen no more than one relay thus far. Compared to PDSs, self-hosting PDSs offers users an incentive: control of their signing keys. There is no similar incentive for hosting a Relay, which is also far more resource demanding; it is the Feed Generators and Labelers that differentiate content for clients, not the Relay. The centralized Relay seems an inevitable outcome: there is currently no easier way for Feed Generators and Labelers to discover the contents of all PDSs; the reality defaults to taking the easiest way out. One potential incentive for deploying multiple Relays is to reduce the control power of a single Relay, which is the motivation for the design of Nostr [3] and Secure Scuttlebutt [2]. On the other hand, the Bluesky community has not had enough incentive to move away from a central Relay.

4 Looking Forward

This paper investigates the reasons for the observed centralization in some parts of the Bluesky operations. We noticed that the observed centralized operations reflect two technical obstacles in today's networking support that create roadblocks in developing decentralized apps: the lack of app-independent user names that are readily usable, and decentralized platforms where data consumers can meet with data producers without an intermediary. Permissionless (or anonymous) blockchains rely on voting, with voting power controlled by resource ownership (*e.g.,* PoW, PoS), and their anonymous nature complicates law enforcement. IPFS does not provide an app platform, but a storage layer for anonymous content publication [7]. Mitigating these two technical obstacles could help us understand how to build more-decentralized apps.

References

[1] Leonhard Balduf, Saidu Sokoto, Onur Ascigil, Gareth Tyson, Bjorn Scheuermann, Maciej Korczynski, Ignacio Castro, and Michal Krol. 2024. Looking at the blue skies of bluesky. In *Proceedings of the 2024 ACM on Internet Measurement Conference*. 76–91.

[2] Secure Scuttlebutt Community. 2025. Secure Scuttlebutt: An Off-Grid Social Network. https://ssbc.github.io/scuttlebutt-protocol-guide/. SSB Protocol Guide.

[3] fiatjaf. 2025. NIP-01: Basic protocol flow description. https://github.com/nostr-protocol/nips/blob/master/01.md

[4] Christian Huitema, Geoff Huston, Dirk Kutscher, and Lixia Zhang. 2023. Report of 2021 DINRG Workshop on Centralization in the Internet. *ACM SIGCOMM Computer Communication Review* 53, 2 (2023), 31–39.

[5] Martin Kleppmann, Paul Frazee, Jake Gold, Jay Graber, Daniel Holmgren, Devin Ivy, Jeromy Johnson, Bryan Newbold, and Jaz Volpert. 2024. Bluesky and the at protocol: Usable decentralized social media. In *Proceedings of the ACM Conext-2024 Workshop on the Decentralization of the Internet*. 1–7.

[6] Gergely Orosz. 2024. Inside Bluesky, the Decentralized Twitter Competitor. https://newsletter.pragmaticengineer.com/p/bluesky. The Pragmatic Engineer Newsletter.

[7] Ruizhe Shi, Ruizhi Cheng, Bo Han, Yue Cheng, and Songqing Chen. 2024. A closer look into IPFS: Accessibility, content, and performance. *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 8, 2 (2024), 1–31.

[8] Manu Sporny, Dave Longley, Markus Sabadello, Drummond Reed, Orie Steele, and Christopher Allen. 2025. Decentralized Identifiers (DIDs) v1.1. https://www.w3.org/TR/did-1.1/. W3C Recommendation.