D. Trautwein, M. Schubotz, and B. Gipp. "Leveraging Node Heterogeneity to Improve Content Discovery and Content Retrieval in Peer-To-Peer Networks". In: *CoNEXT Student Workshop 2021 (CoNEXT-SW'21), December 7, 2021, Virtual Event, Germany.* Virtual Event, Germany: ACM, Dec. 2021. DOI: 10.1145/3488658.3493781

Leveraging Node Heterogeneity to Improve Content Discovery and Content Retrieval in Peer-To-Peer Networks

Dennis Trautwein[1], Moritz Schubotz[2], Bela Gipp[1]
[1] University of Wuppertal, Germany ({last}@gipplab.org)
[2] FIZ-Karlsruhe, Germany ({first.last}@fiz-karlsruhe.de)

ABSTRACT

Distributed content-addressed networks like the Interplanetary File System (IPFS) have gained popularity in recent years. It is used as a data storage layer for Blockchain applications or generally as a censorship-resistant file-sharing network. In the past, many traditional peer-to-peer networks eventually implemented *super-peer* sub-topologies within their otherwise nonhierarchical network to increase the efficiency of various system tasks. While it can be observed that IPFS also holds such tendencies, no formalism supports these developments. This article will present a research project that will investigate if and how a *super-peer* architecture can complement the IPFS network from a content discovery and retrieval perspective. Further, first preliminary results on peer churn rate in the IPFS and Filecoin networks will be discussed.

KEYWORDS

ipfs, filecoin, peer-to-peer, super-peers, churn

1 INTRODUCTION

In recent years the peer-to-peer (P2P) application architecture has gained popularity due to the emergence of crypto-currencies such as Bitcoin or Ethereum. This development motivated efforts to decentralize other applications as well. The IPFS network [1] is a prominent example of a P2P system that provides public data storage services and introduces content-addressing as a primitive to ensure data integrity and verifiability – core properties relevant to facilitate decentralization. There are, however, many challenges that arise from cutting down central service providers that need to be overcome for it to gain further adoption.

A promising approach to compete with the existing performance characteristics of Internet applications like the Web is based on the idea of relaxing the decentralization requirement, acknowledging the existence of differently capable network participants, and off-load network tasks to the subset of such potent peers – knowing that this shift is tantamount to a light form of centralization. Hence, ensuring permissionless entry to this subset of peers is a key property. Traditional P2P networks from the early 2000s eventually converged on leveraging this heterogeneity in their network topology by introducing so-called *super-peer* sub-topologies [2].

This research project will focus on investigating if and how a *super-peer* architecture can complement the IPFS network from a content discovery and content retrieval perspective. The architecture should respect the users' privacy by providing anonymization similar to what Virtual Private Networks (VPNs) do in the Internet, properly incentivize disproportionally contributing network participants, and ensure low latency access to content. It can be observed that the IPFS network already holds tendencies to exploit heterogeneity due to the emergence of platforms like *Pinata*[1] or *Infura*[2] that provide *pinning services* to address storage persistence. Protocol Labs[3], the supporting company behind IPFS, made this topic a research priority by publishing the `RFC|BB|L1/2-05`[4] in November 2020, further emphasizing the relevance and importance of this research area.

The remainings of this article are structured as follows: Section 2 will give a concise research objective. Afterward, in section 3, preliminary findings are presented before section 4 gives an outlook over the next steps.

2 KNOWLEDGE GAP

Based on previous research and the motivation in the previous section, the following research objective is formulated:

Propose, implement, and evaluate an architecture that leverages heterogeneity in peer-to-peer networks while incentivizing highly capable peers, minimizing centralization forces, and preserving users' privacy.

Previous work suggests that P2P networks consist of highly heterogeneous peers with vastly varying capabilities [6]. It was observed that "bandwidth, latency, availability, and the degree of sharing vary between three and five orders of magnitude across the peers in the system" [6, p. 13]. An approach that introduces some sort of hierarchy or *gradient topology* [5] into the system naturally necessitates some degree of centralization. Further, if a smaller set of peers becomes responsible for certain tasks, those peers could be a threat to users' privacy. E.g., usage patterns could be derived as fewer peers are hit with more traffic. Lastly, peers who choose to contribute more resources to the system as the average participant needs to be compensated adequately.

In order for the decentralized Web to succeed, it must become on par with the current architecture of the Web from a performance perspective. Exploiting the heterogeneity of peers to drive this shift while making reasonable compromises to the above criteria is a promising approach for that goal.

While the above objective is formulated generically, the research project will primarily target the IPFS network. As IPFS is built upon `libp2p`[5] it is not far that many of the research results, tools and techniques can be applied to other related networks like Filecoin[6].

The next section will give a brief outline of the anticipated work packages and present first network measurement results.

[1] https://pinata.cloud/
[2] https://infura.io/
[3] https://protocol.ai/
[4] https://github.com/protocol/beyond-bitswap
[5] https://libp2p.io/
[6] https://filecoin.io/

3 METHODOLOGY

The research project will consist of five work packages. First, existing literature on *super-peer* architectures will be reviewed in a comprehensive and systematic manner. The literature review is a crucial step in the research process. It is important to understand state-of-the-art and previous approaches in order to transfer the knowledge to the research problems at hand. The key deliverable of this work package is a comprehensive and systematic literature review on super-peer overlay networks.

Second, a detailed view of the IPFS network is necessary for the development of a complementary architecture. This can be achieved with the novel, self-developed network analysis tool `nebula-crawler`[7]. It is specifically designed to take measurements in P2P networks that are built on top of `libp2p`, which includes IPFS. Figure 1 shows first results of the peer churn rate in the IPFS and Filecoin networks, which has not been measured before to the best of the author's knowledge. The measurements were conducted in a two weeks time window in July 2021. The data shows that 80 % of the peers in the IPFS network stay online for three hours or less and 80 % of the peers in the Filecoin network for nine hours or less. These results are, e.g., relevant for network-wide parameters like k-bucket size in the Kademlia [3] distributed hash table protocol employed in both networks. It further motivates to leverage the heterogeneity of the peers and off-load important network tasks to the set of stable and highly capable participants.

A holistic understanding of the state-of-the-art research on *super-peer* networks via the first work package and the detailed network view from the second work package allows devising a concept of a *super-peer* architecture. It should take privacy, incentivization, and centralization into account tailored to the content-addressable nature of the IPFS network. It should also ensure permissionless entry and carefully consider load distribution among the *super-peer* nodes. For the latter, random re-peering as used in the BitTorrent network could be a viable solution.

The fourth and fifth phases will focus on the implementation and evaluation of the proposed architecture. I plan to employ tools like testground[8] to evaluate the performance of the implementation and compare it to existing benchmarks that were derived from other approaches to speed up content routing performance like *Bitswap* [4].

4 CONCLUSION

This research project proposes to introduce the exploitation of heterogeneity in the IPFS network to speed up content discovery and content retrieval. The measured churn rate shows a need to shift focus to more reliable peers and leverage their stability and performance. Devising a suitable architecture that carefully considers the trade-offs regarding peer incentivization, users' privacy, and centralization concerns are at its heart. The research will provide promising grounds for further research on the IPFS and related networks, e.g., due to the development of measurement tools.

[7] https://github.com/dennis-tra/nebula-crawler
[8] https://docs.testground.ai/

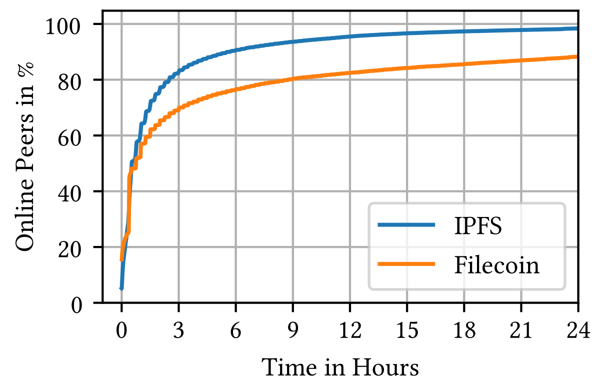

Figure 1: The churn rate of the IPFS and Filcoin networks as cumulative distribution functions (CDFs) based on 270 k **and** 100 k **observed peer sessions respectively from 2021-07-09 to 2021-07-23.**

REFERENCES

[1] J. Benet. "IPFS - Content Addressed, Versioned, P2P File System". In: 2014.

[2] B. Beverly Yang and H. Garcia-Molina. "Designing a Super-Peer Network". In: *Proceedings 19th International Conference on Data Engineering (Cat. No.03CH37405).* 19th International Conference on Data Engineering. Bangalore, India: IEEE, 2003, pp. 49–60. DOI: 10/b8jsg9.

[3] P. Maymounkov and D. Mazières. "Kademlia: A Peer-to-Peer Information System Based on the XOR Metric". In: *Peer-to-Peer Systems.* Ed. by P. Druschel, F. Kaashoek, and A. Rowstron. Red. by G. Goos, J. Hartmanis, and J. van Leeuwen. Vol. 2429. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 53–65. DOI: 10.1007/3-540-45748-8_5.

[4] A. de la Rocha, D. Dias, and Y. Psaras. *Accelerating Content Routing with Bitswap: A Multi-Path File Transfer Protocol in IPFS and Filecoin.* 2021.

[5] J. Sacha. "Exploiting Heterogeneity in Peer-to-Peer Systems Using Gradient Topologies". PhD thesis. Dublin, Ireland: University of Dublin, Trinity College, 2009. 218 pp.

[6] S. Saroiu, P. K. Gummadi, and S. D. Gribble. "A Measurement Study of Peer-to-Peer File Sharing Systems". In: *Multimedia Computing and Networking 2002.* Vol. 4673. SPIE, Dec. 10, 2001, pp. 156–170. DOI: 10.1117/12.449977.

Listing 1: Use the following BibTeX code to cite this article

```
@inproceedings{Trautwein2021a,
 author       = {Dennis Trautwein and Moritz Schubotz and Bela Gipp},
 title        = {Leveraging Node Heterogeneity to Improve Content Discovery and Content Retrieval in
     Peer-To-Peer Networks},
 booktitle    = {CoNEXT Student Workshop 2021 (CoNEXT-SW'21), December 7, 2021, Virtual Event, Germany},
 doi          = {10.1145/3488658.3493781},
 isbn         = {978-1-4503-9133-7/21/12},
 year         = {2021},
 address      = {Virtual Event, Germany},
 month        = {Dec.},
 publisher    = {ACM},
 topic        = {misc}
}
```