

The Cloud Strikes Back

Investigating Decentralization of IPFS

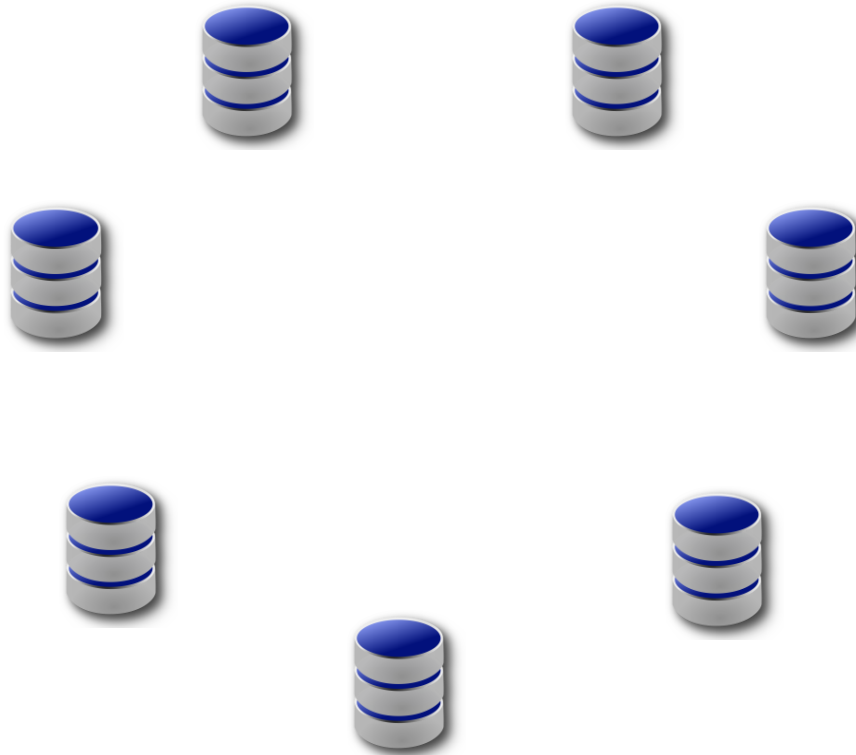
L. Balduf, M. Korczyński, O. Ascigil, N.V. Keizer, G. Pavlou, B. Scheuermann, M. Król

InterPlanetary File System (IPFS)

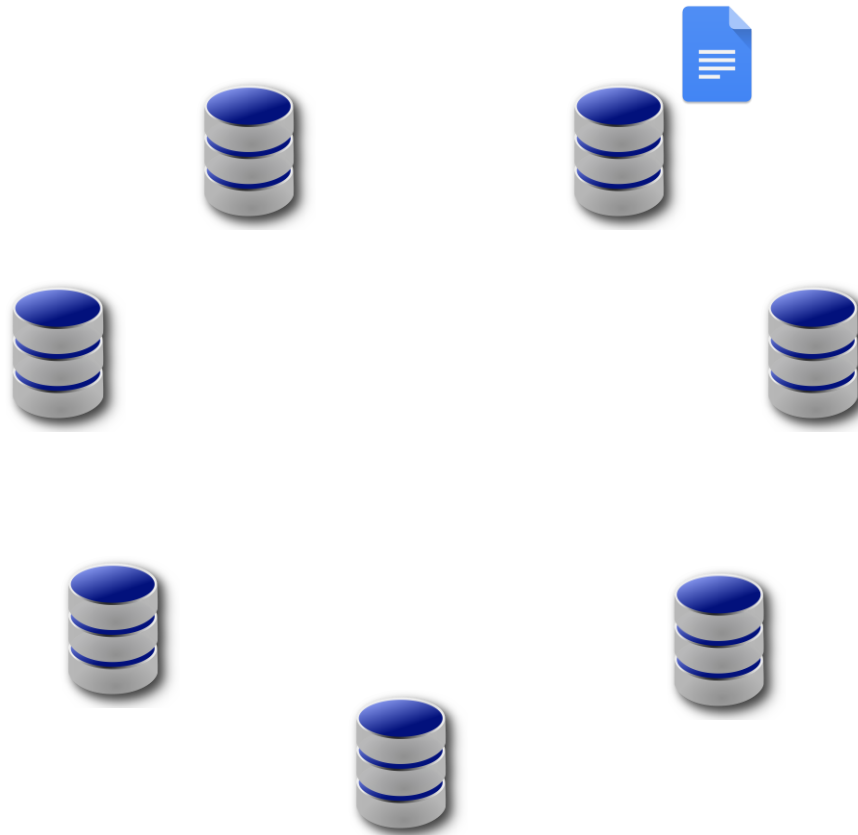
IPFS

- The largest P2P data storage/retrieval network
- Used for traditional hosting, NFTs, blockchain-referenced data, DApps.
- Growing usage: $> 10^7$ download requests/day, $\approx 30k$ nodes in DHT

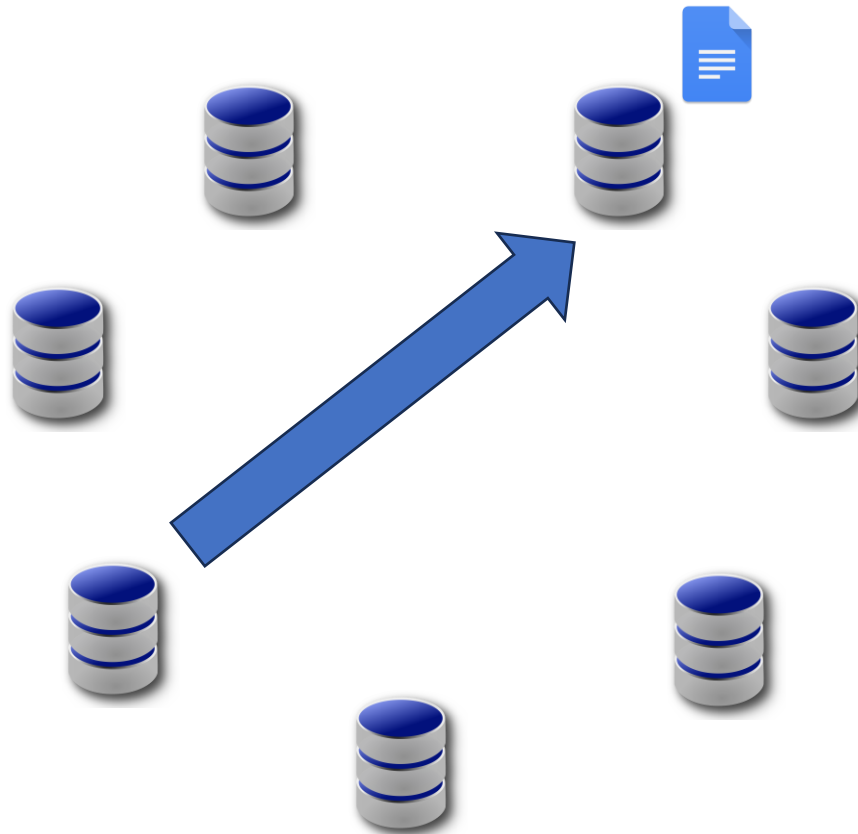
InterPlanetary FileSystem



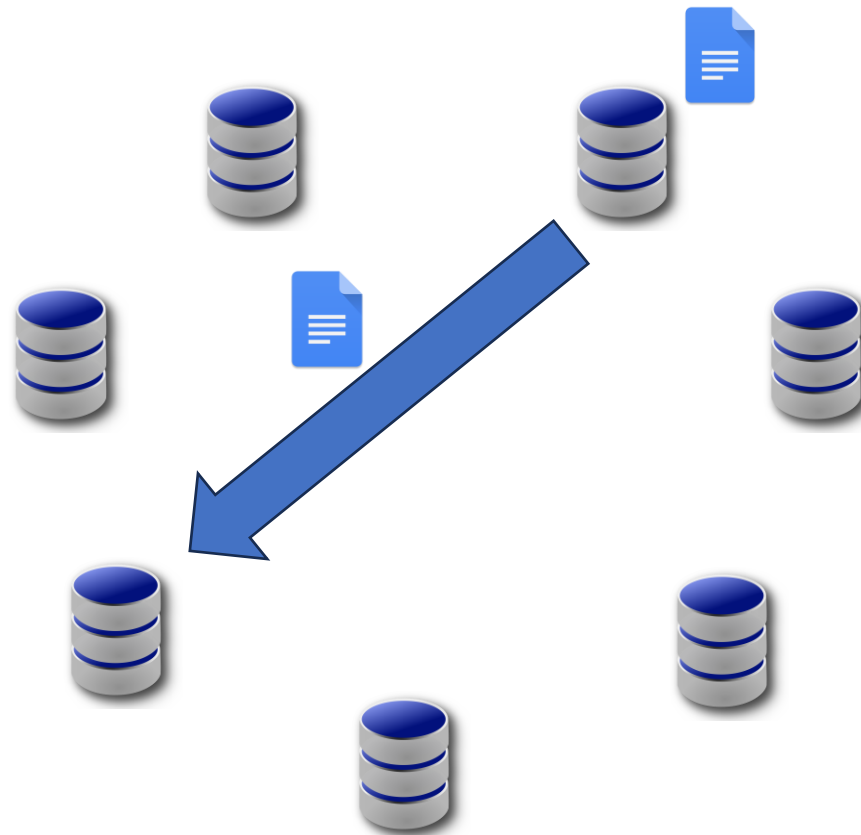
InterPlanetary FileSystem



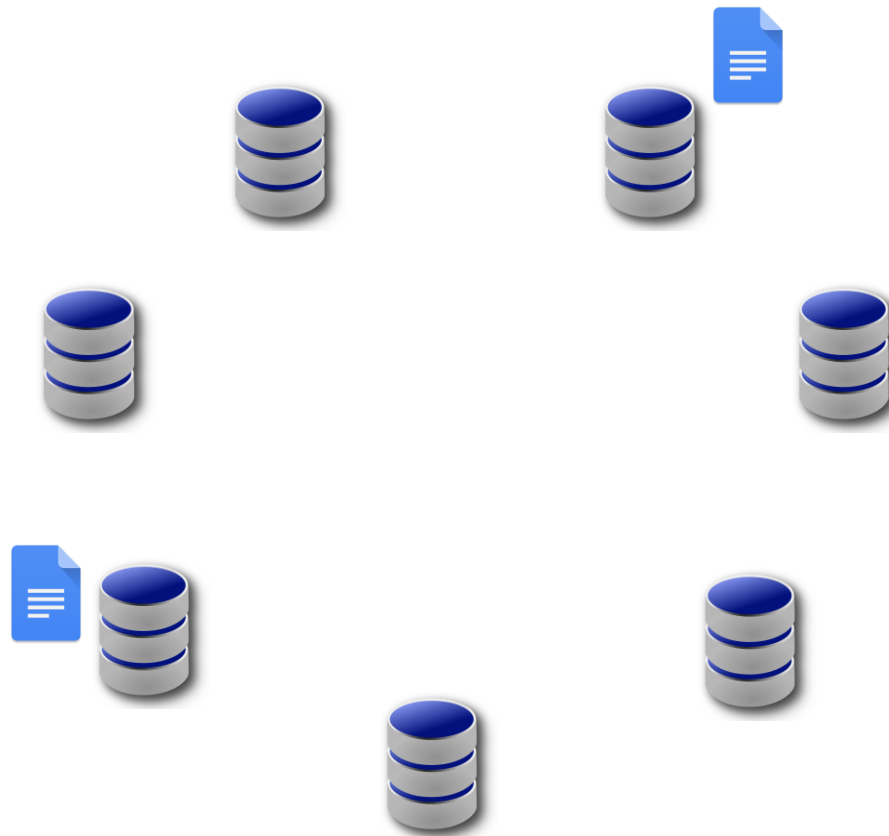
InterPlanetary FileSystem



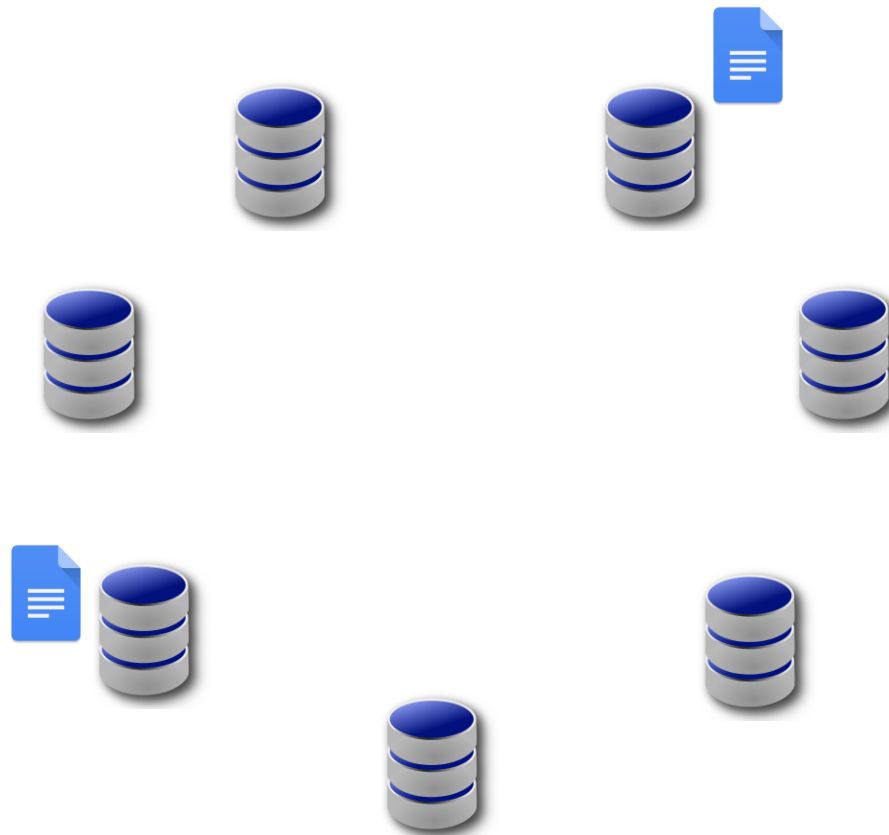
InterPlanetary FileSystem



InterPlanetary FileSystem

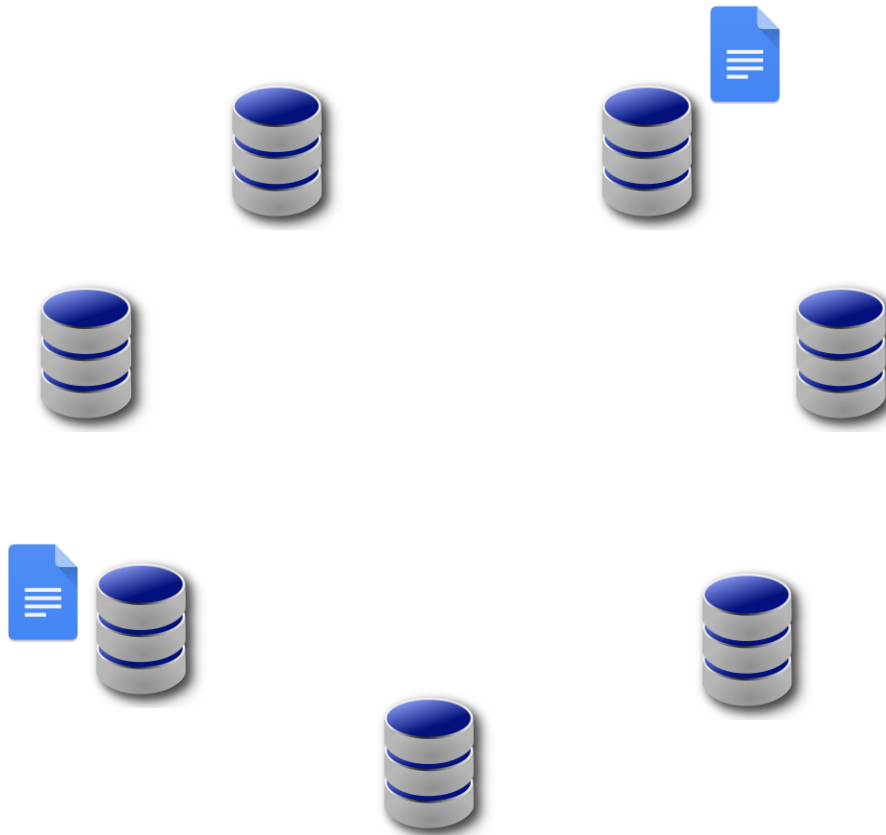


InterPlanetary FileSystem



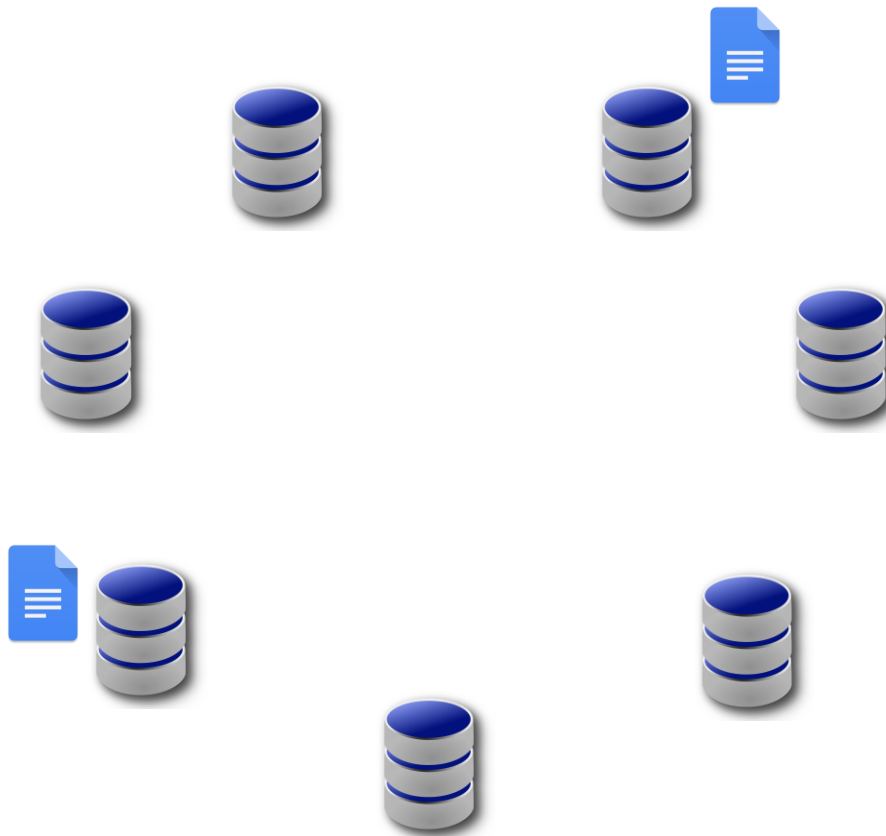
youtube.com/video.mp4

InterPlanetary FileSystem



youtube.com/video.mp4
location

InterPlanetary FileSystem



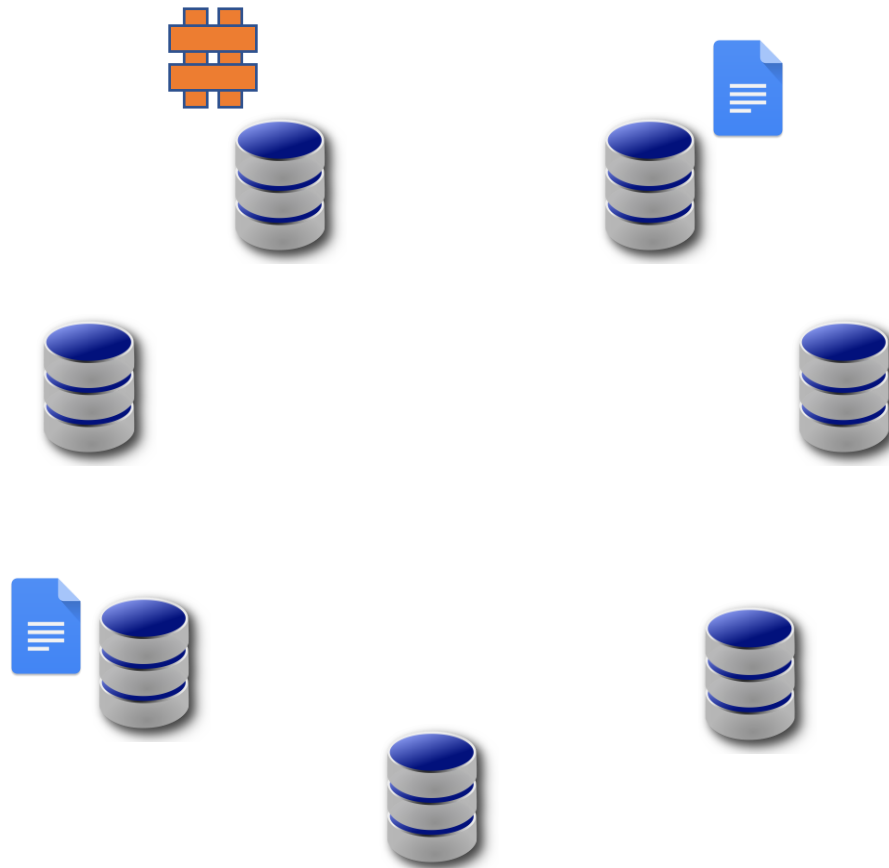
youtube.com/video.mp4
location



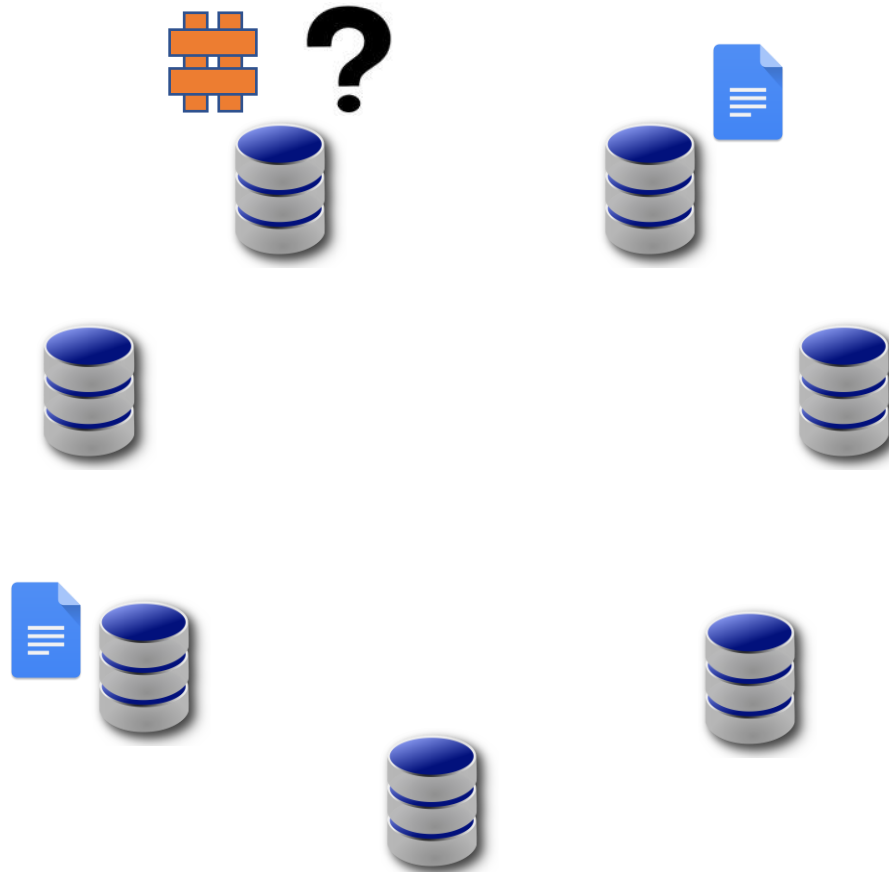
$$h(\text{document icon}) = \text{hash icon}$$

QmSAPbvtmtm22VhdQmM8Eaa4kL5u5N3

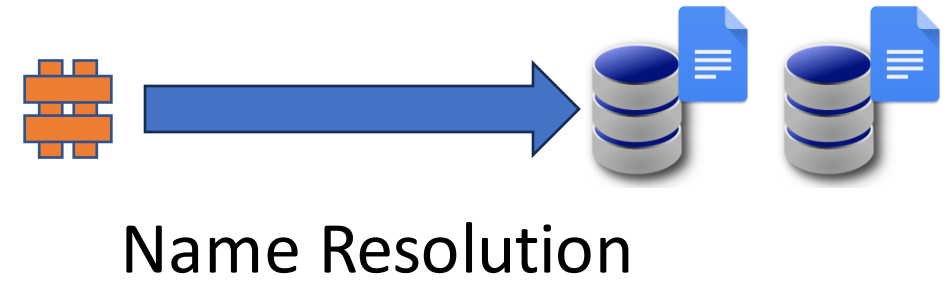
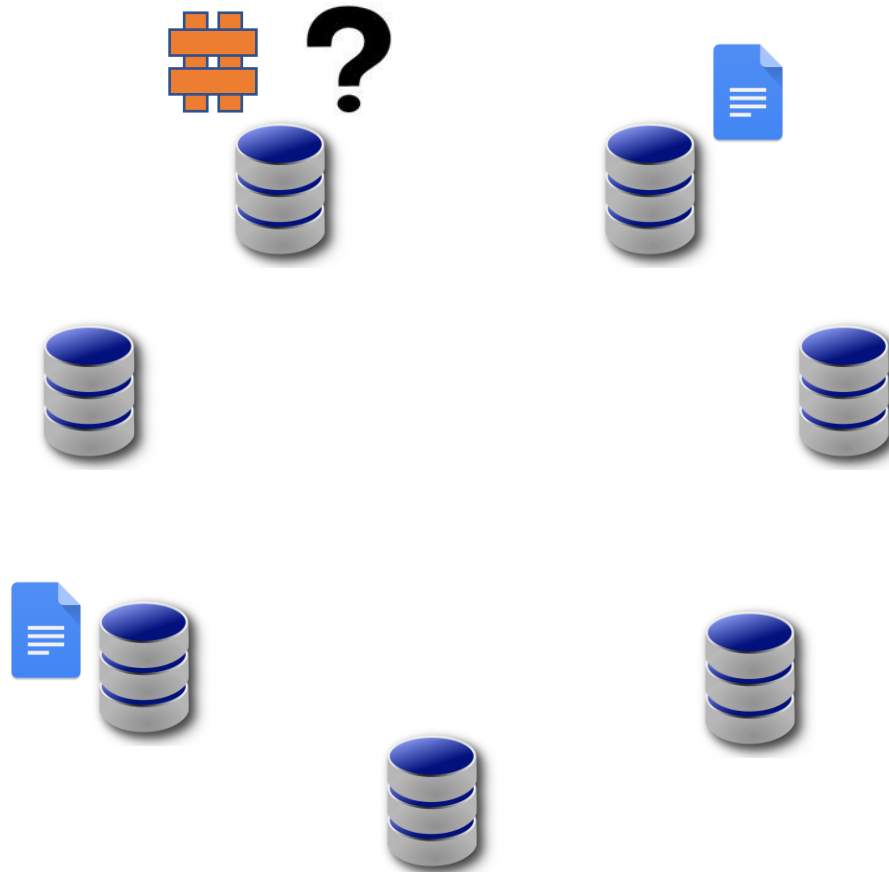
InterPlanetary FileSystem



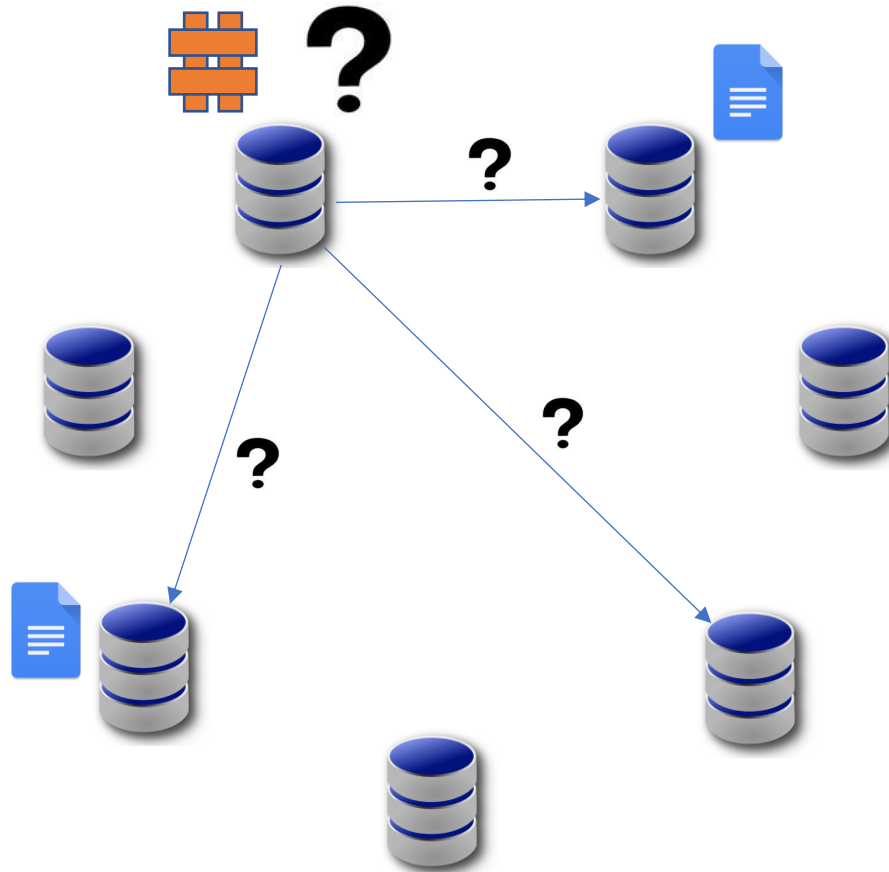
InterPlanetary FileSystem



InterPlanetary FileSystem

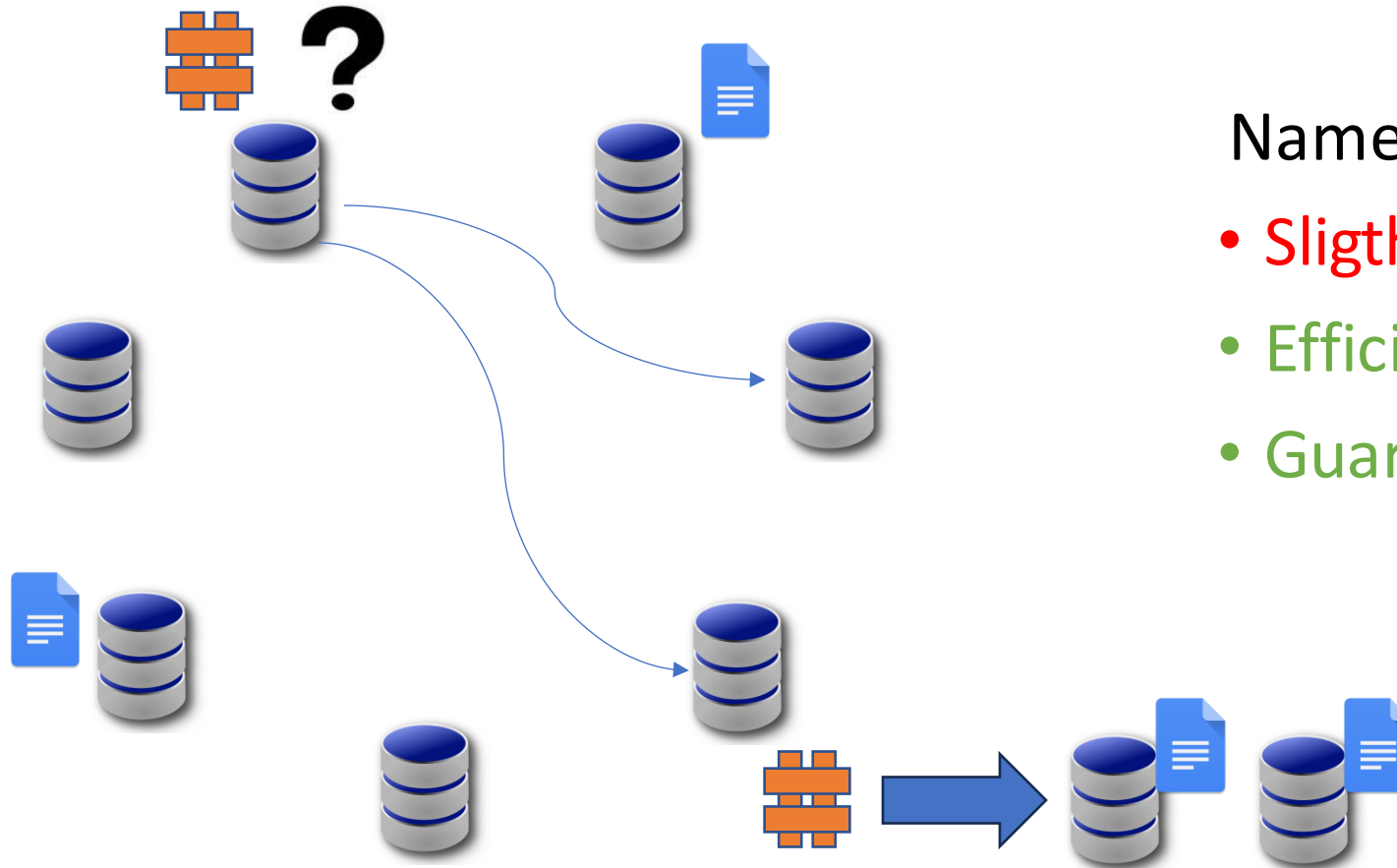


Name Resolution - Bitswap



- Fast
- Expensive (overhead)
- Not guaranteed to find

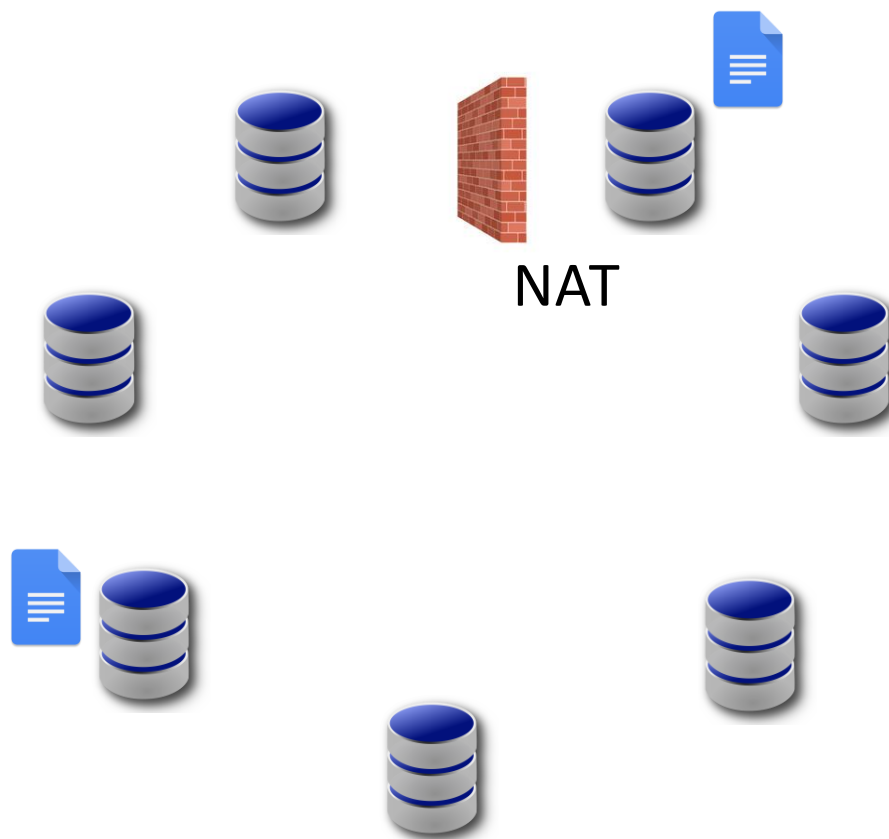
Name Resolution - Distributed Hash Table



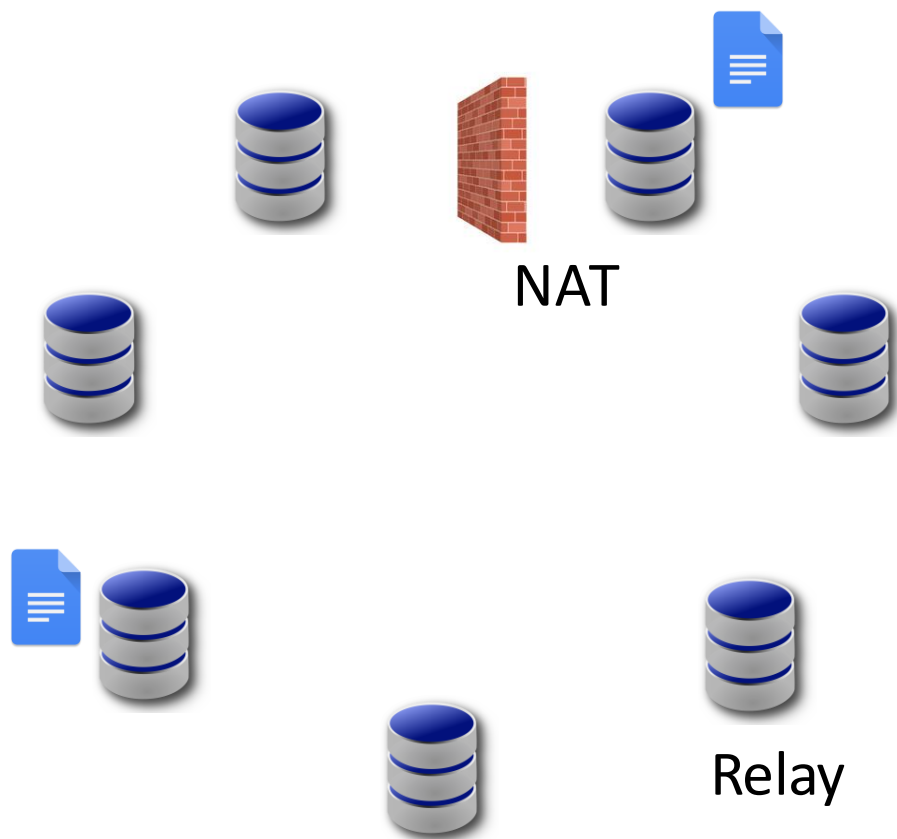
Name Resoluion

- Slightly Slower
- Efficient ($\log(n)$)
- Guaranteed to find

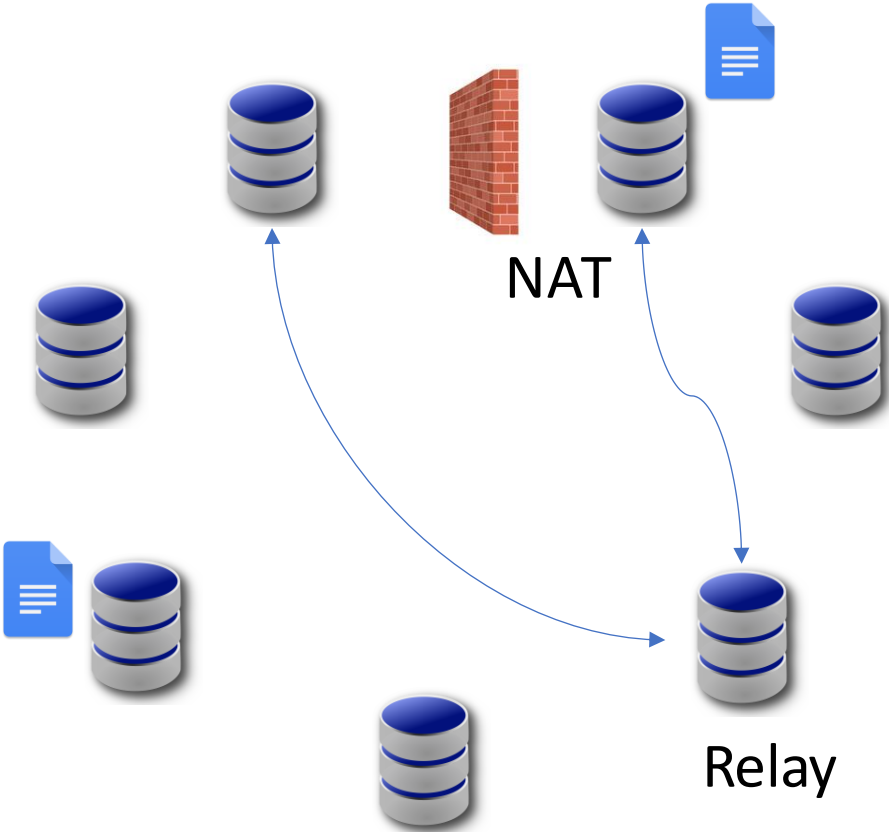
Relays



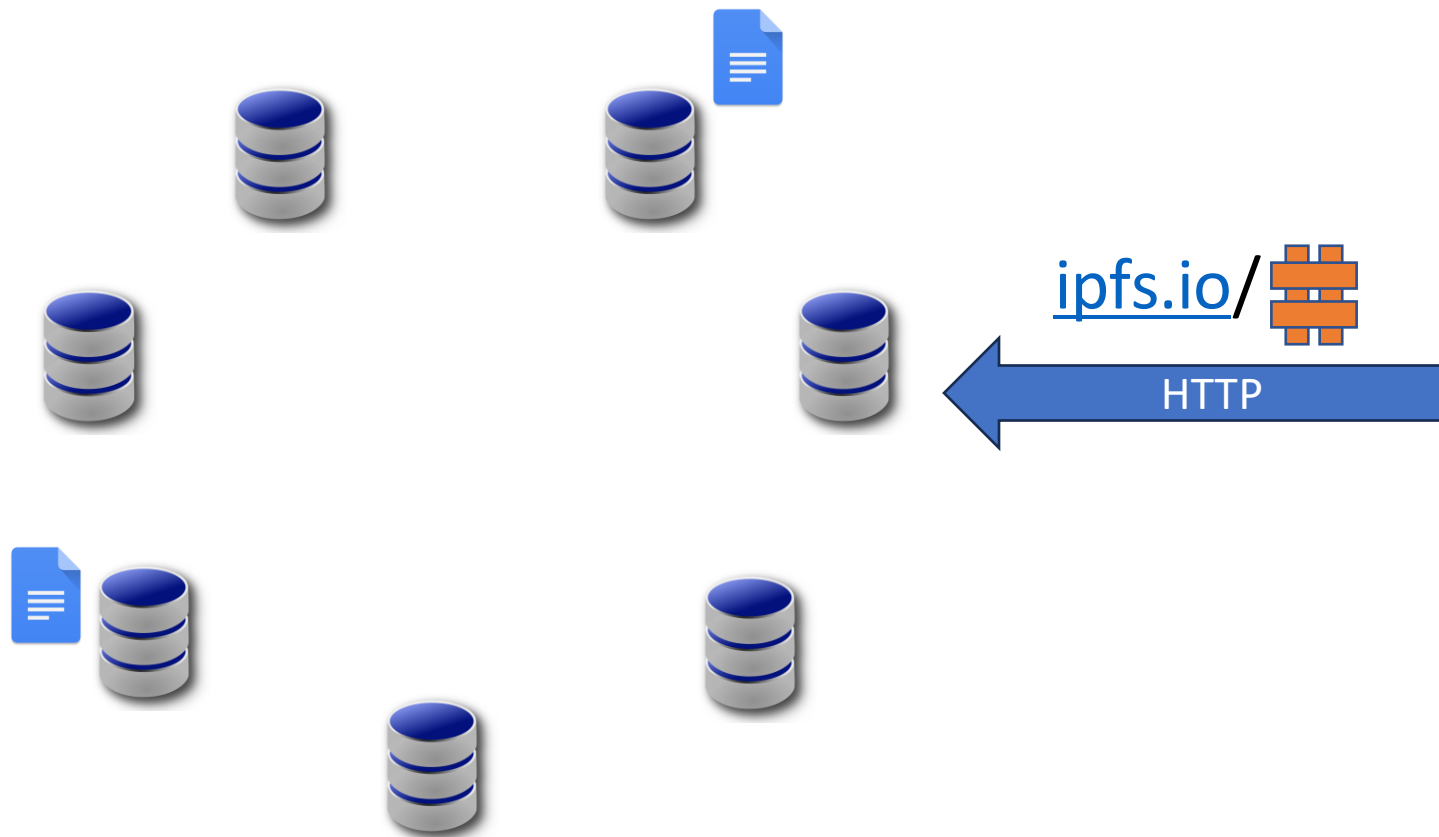
Relays



Relays



Gateways



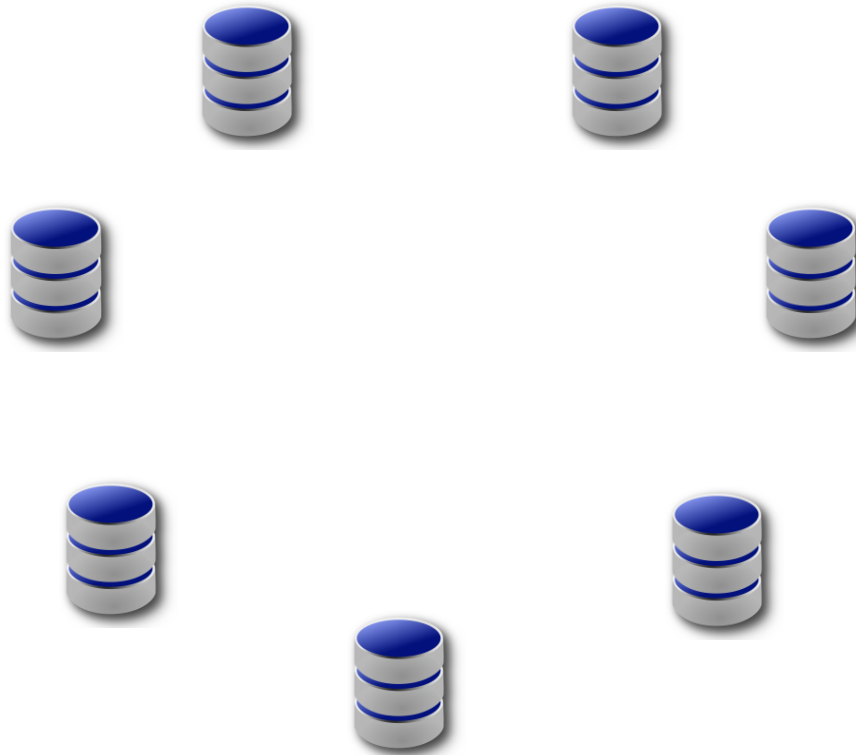
Methodology

What is (de)centralization?

We look at

1. Traffic distribution
2. Reliance on cloud infrastructure

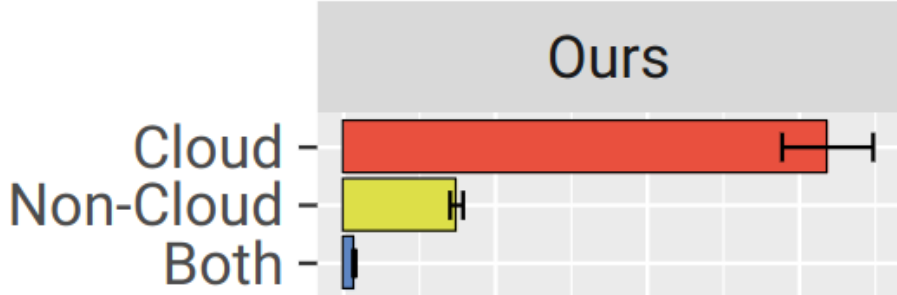
Data Source



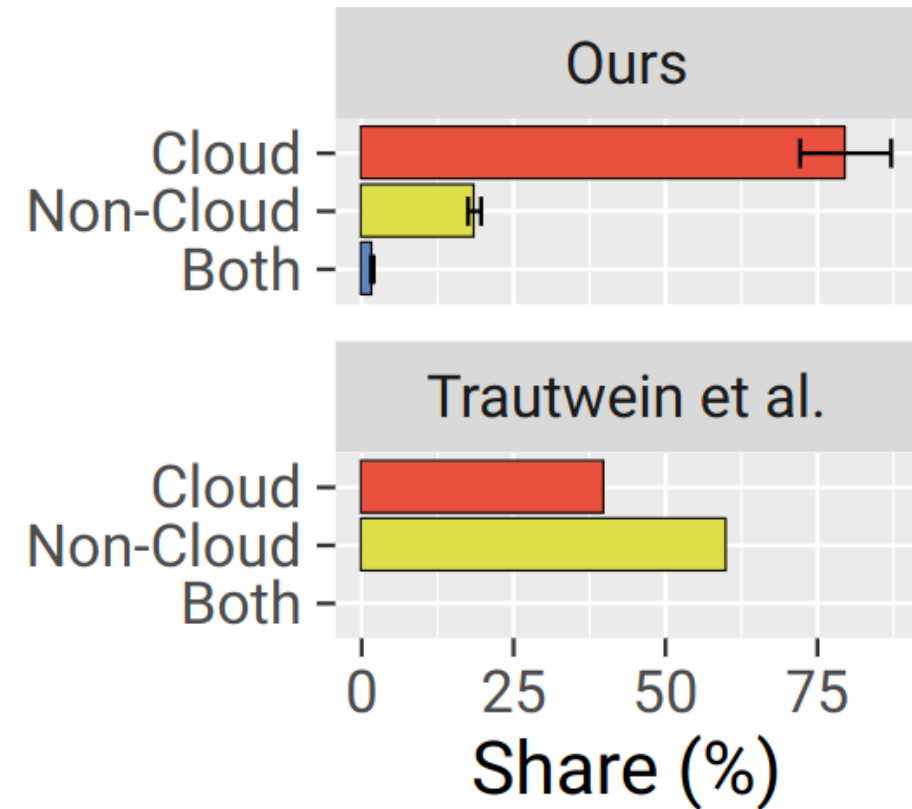
1. Network Crawler
2. DHT Traffic Sniffer
3. Bitswap Traffic Sniffer
4. Gateway Detector
5. DNS Crawler
6. ENS Crawler

Results

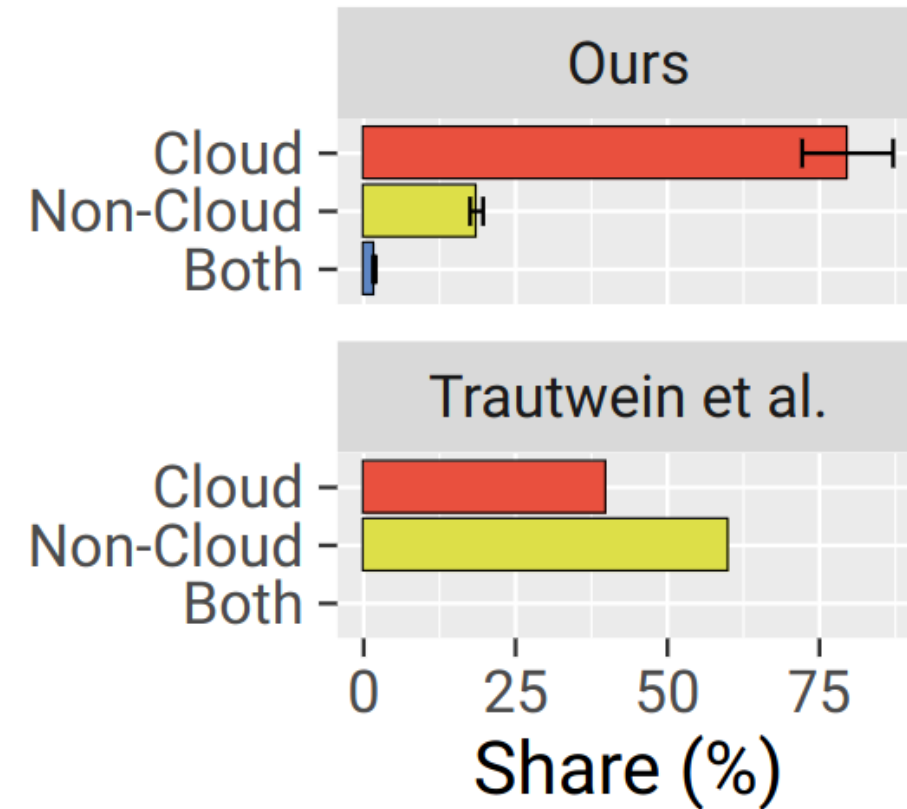
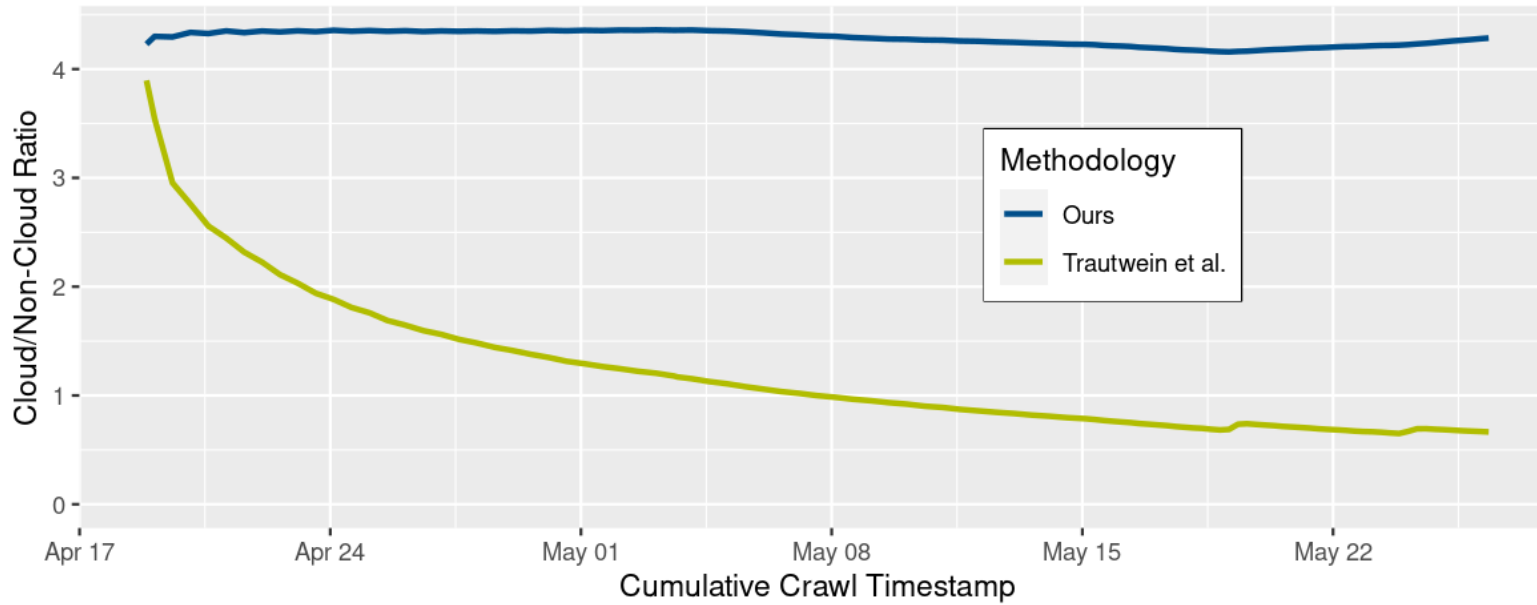
The Network



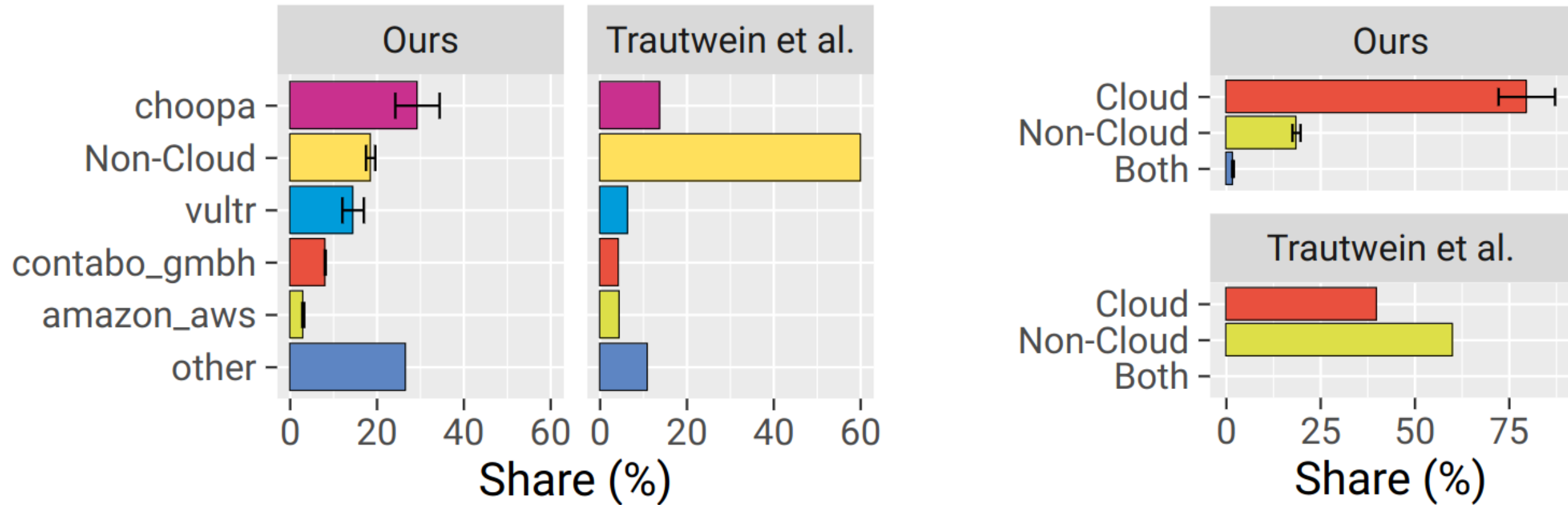
The Network



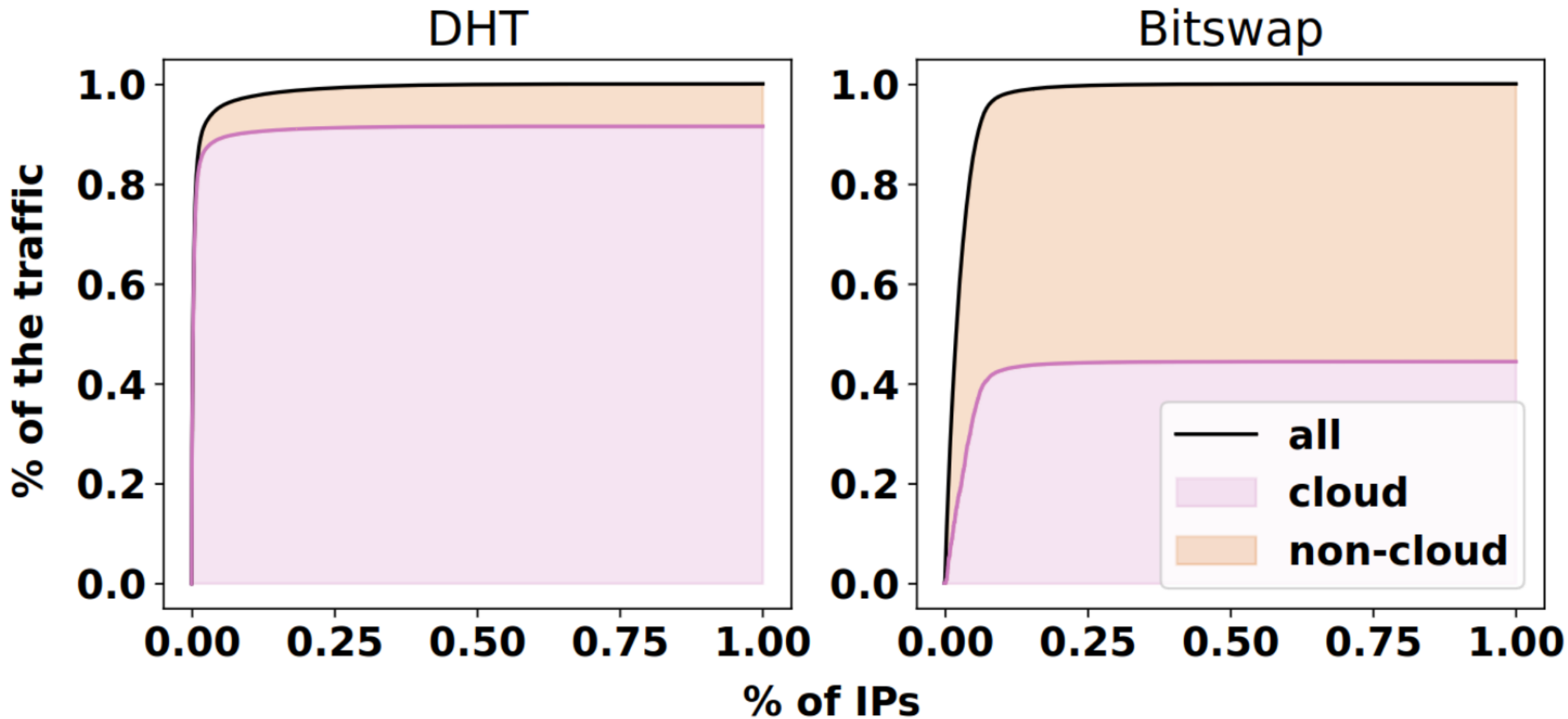
The Network



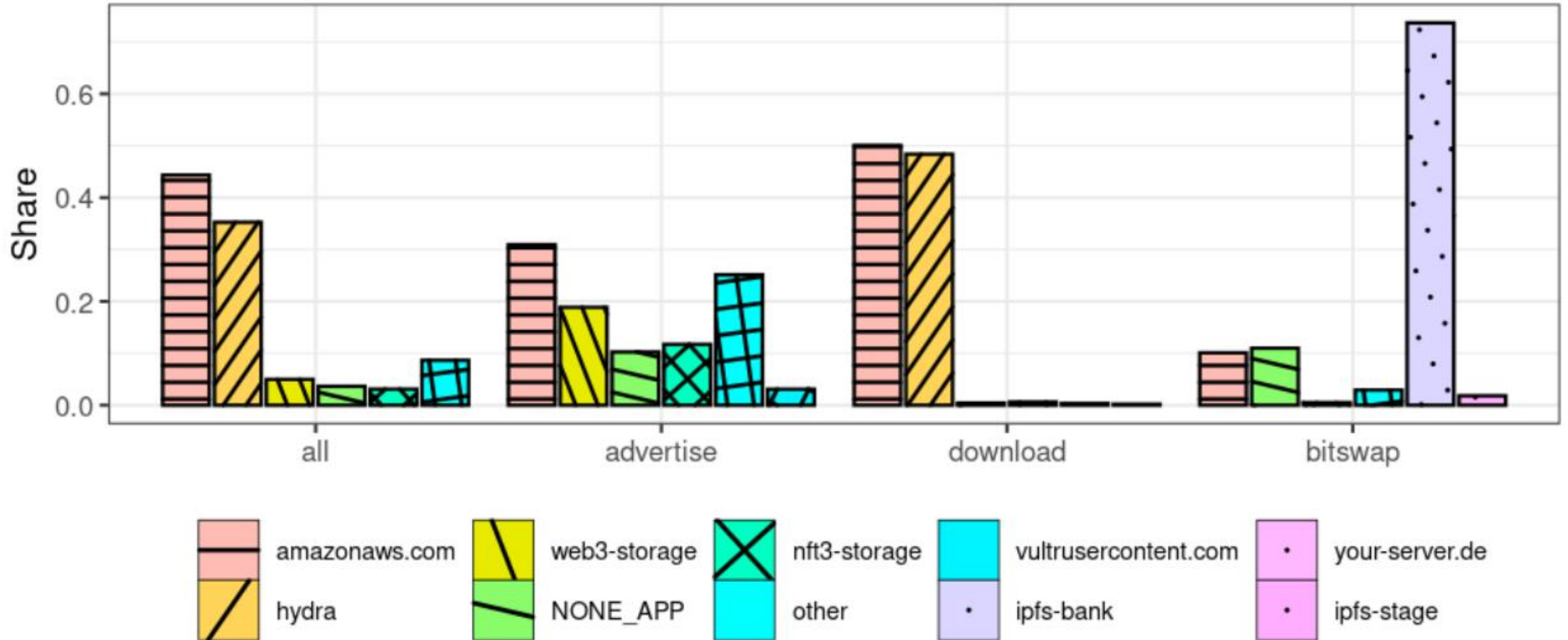
The Network



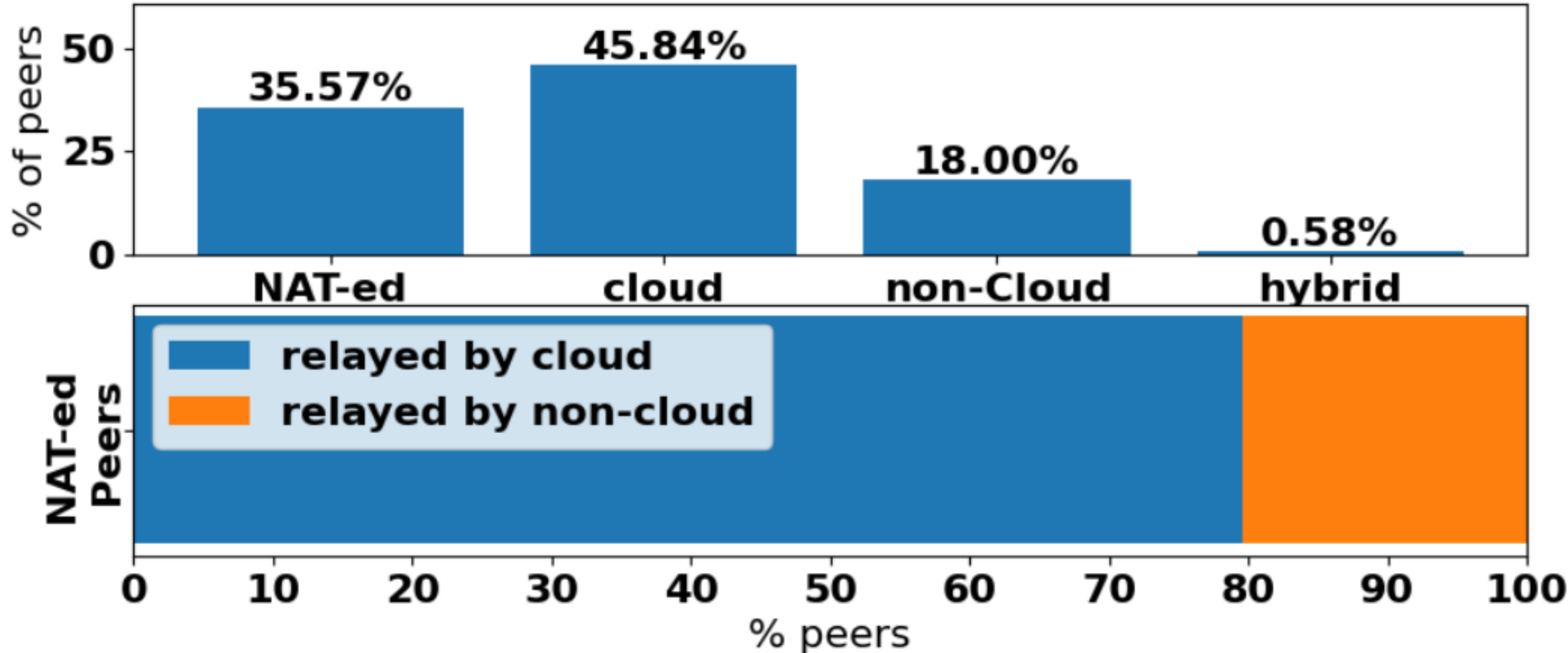
The Traffic



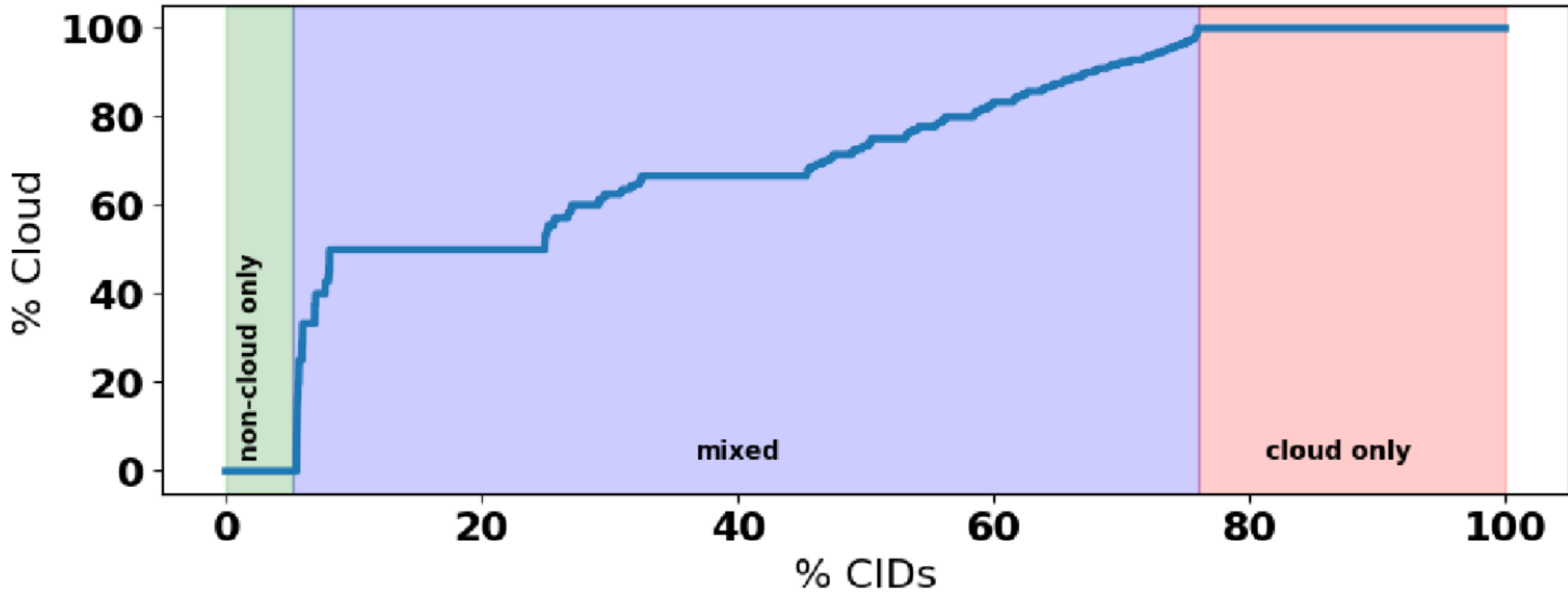
The Traffic



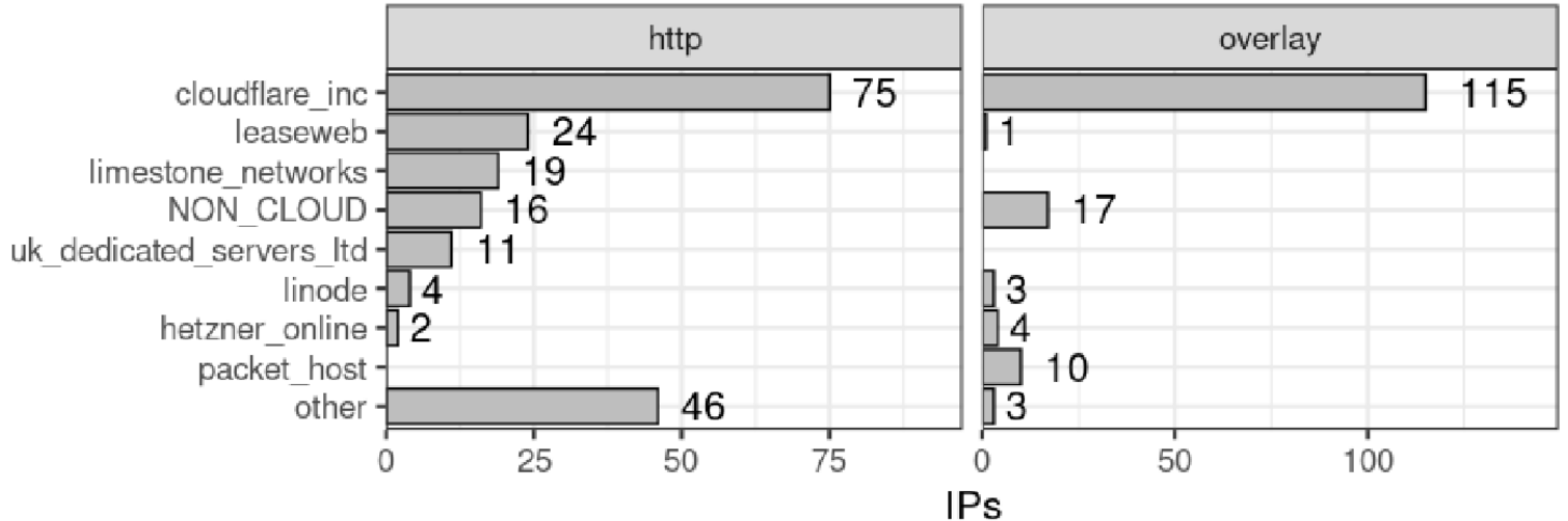
The Content Providers



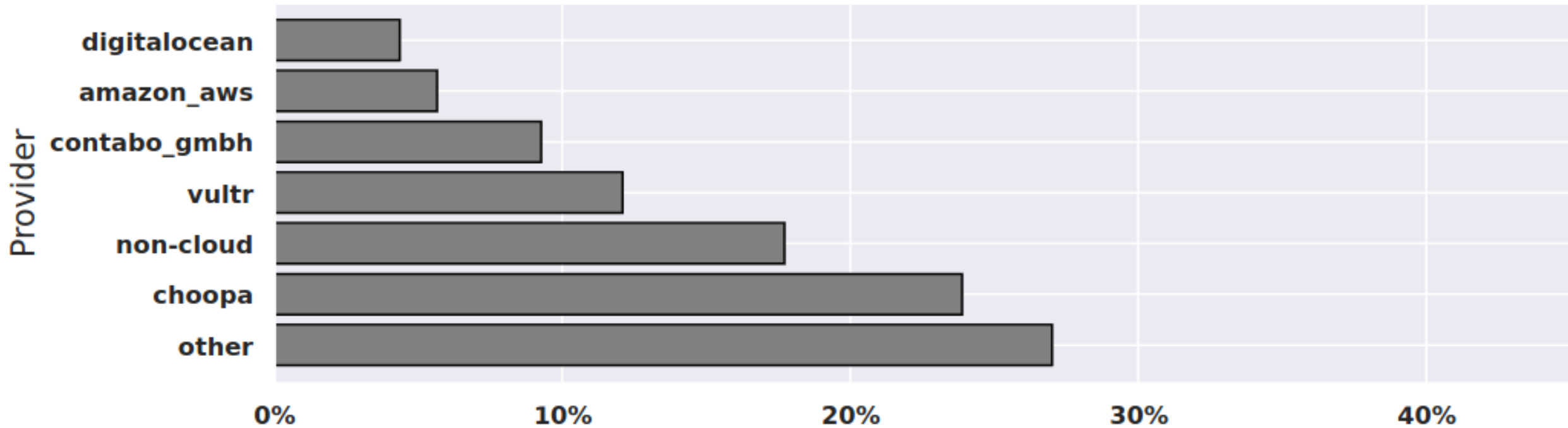
The Content Providers



The Gateways



Ethereum Name System



Result Summary

- Almost 80% of DHT servers are hosted in the cloud.
- The top 3 cloud providers host >51% of the servers.
- The top 5% of the nodes are responsible for up to 95% of the traffic.
 - The largest cloud provider, AWS, generates 96% of all the content resolution requests.
- Nearly 95% of the content is provided by at least one cloud-based node.
 - ≈25% is provided only by cloud nodes.
- Entry points to the network often rely on public gateways, which are dominated by a few large players.

Results

Takeaways

- IPFS is a robust base for the decentralized Web
- It's difficult to achieve decentralization, performance and security [1]
 - Content Resolution
 - Content Transfer
- NAT is still a problem
 - but some solutions are coming
- Non-cloud hosts come and go causing issues with reliability
- Huge progress on practicality and ease of use but still some work is needed

[1] Sridhar, Srivatsan, Onur Ascigil, Navin Keizer, François Genon, Sébastien Pierre, Yiannis Psaras, Etienne Rivière, and Michał Król. "Content Censorship in the InterPlanetary File System." *NDSS'24*