

Reading Ethereum's Tea Leaves with Xatu Data

Andrew Davis &
Sam Calder-Mason
DevOps Engineers @ ethPandaOps

Toni Wahrstätter
EF Researcher

Leo Bautista-Gomez
MigaLabs

Workshop Overview

- Introduction
- Xatu Genesis
- Datasets
- Using the data
- **Leo Bautista-Gomez** from **MigaLabs**
- **Toni Wahrstätter** from ARG at **Ethereum Foundation**

Introduction



**Andrew
Davis**



Sam Calder-Mason

DevOps Engineers @ ethPandaOps

Combined height > African bush elephant

Combined IQ of 2 adult orangutan's

Combined weight of 9 moo deng's



ethPandaOps

Embedded within the Ethereum Foundation

DevOps for the Ethereum Protocol itself



barnabasbusa
DevOps Engineer



mattevans
DevOps Engineer



parithosh
DevOps Engineer



pk910
DevOps Engineer



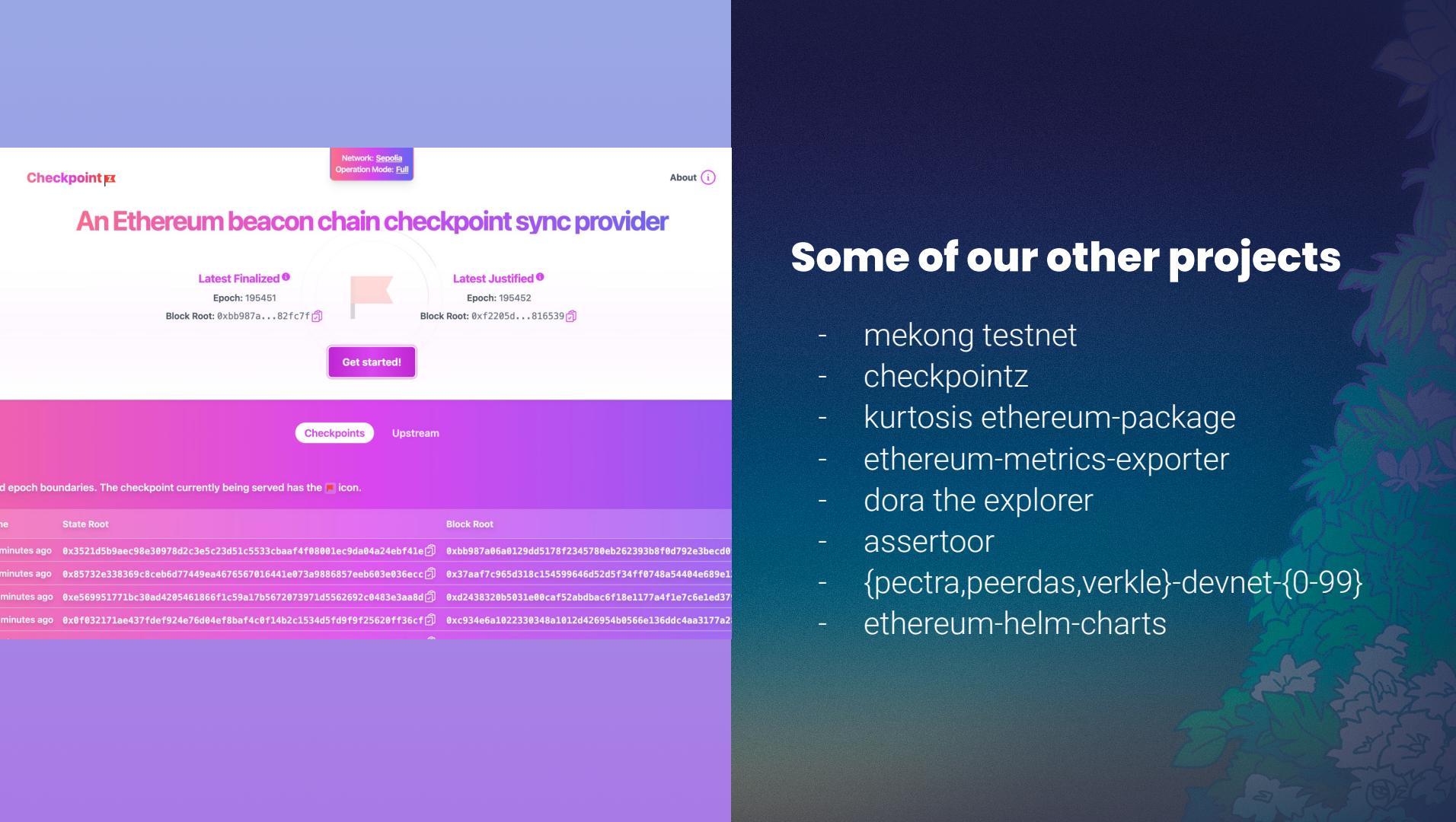
samcm
DevOps Engineer



savid
DevOps Engineer



skylenet
DevOps Engineer



Some of our other projects

- mekong testnet
- checkpointz
- kurtosis ethereum-package
- ethereum-metrics-exporter
- dora the explorer
- assertoor
- {pectra,peerdas,verkle}-devnet-{0-99}
- ethereum-helm-charts

Xatu Genesis

Setting the scene

- Late 2022. Merge complete! 🎉
- **Much more important to know WHEN something happened**
- Lack of data at the network level
- Researchers start capturing this data by themselves at varying scale



- Needs to integrate with existing beacon node implementations
- Existing observability tools aren't suitable
- Beacon API has an /eth/v1/events endpoint
 - Has support for when the beacon node saw:
 - A block
 - An attestation
 - More!
 - All beacon nodes support this endpoint!

Xatu Genesis

Designed to collect data from different sources in a distributed fashion. Built with Go and GRPC.

2 initial modules:

- **Xatu Server**

- Aggregates events, annotates with metadata
- Forwards events somewhere else

- **Xatu Sentry**

- Runs as a sidecar to a beacon node
- Connects to Beacon API



Event Structure

- All events follow the same structure
 - Easy to add new events
- Clear separation of where the data was derived

```
"event": {  
    // When the event occurred on the client  
    "date_time": "2024-11-09T00:00:00.817951361Z",  
    // Event Name  
    "name": "BEACON_API_ETH_V1_EVENTS_BLOCK_V2"  
},  
"data": {  
    "block": "0x35f87 ... ",  
    "slot": "10357198"  
},  
"meta": {  
    // Meta derived on the client (e.g. Xatu Sentry)  
    "client": {},  
    // Meta derived on the server (e.g. Xatu Server)  
    "server": {}  
}
```

?



Clickhouse



Join us at KubeCon, Nov 12-15 at booth A7! →

ClickHouse Products Use cases Docs Resources Pricing Contact us 37.6k Sign in Start free trial

The **real-time** data warehouse for analytics

Unlock faster queries without skyrocketing costs.

[Start free cloud trial](#) [Download open-source](#)

Integrations Court with support All systems operational

Delivery periods

key_period	String	count	floats percentage	floats count
morning	morning	37.11	37.11	14000
evening	evening	37.11	37.11	14000
night	night	37.11	37.11	14000
afternoon	afternoon	37.11	37.11	14000

DELIVERY BREAKDOWN

1861600

MORNING EVENING NIGHT AFTERNOON



ethPandaOps
@ethPandaOps

1,006,660,619,135 🎉

We've hit 4 commas of Ethereum attestations in our [@ClickHouseDB](#) cluster!

Some stats:

- 3 networks
- 1 trillion rows
- 14TB compressed
- 500TB uncompressed 😱

Want to dig in to the data? We publish it all here:
ethpandaops.io/data/xatu/



Datasets

Multiple Datasets

Consensus
Layer

Beacon API
Events

Canonical
Beacon

libp2p

MEV Relays

Execution
Layer

Transaction
Mempool

Node
Discovery

Canonical
Execution

Hermes by ProbeLab



Cryo by Paradigm



Scalable and sovereign EVM data: modern data engineering best practices

- ⌚ Nov 14th — 3:30 PM - 3:55 PM
- 📍 Talk - Stage 3
- 👤 Storm Slivkoff

Open sourcing Xatu Data

The screenshot shows a blog post titled "Open Sourcing Xatu Data" published on March 25, 2024. The post has 853 words and a reading time of 5 minutes. It features two authors: samcm (DevOps Engineer) and savid (DevOps Engineer). The post discusses the open-sourcing of the Xatu dataset, which contains detailed information about the Ethereum network, including beacon chain events, mempool activity, and canonical chain events. The dataset is available on GitHub under the CC BY 4.0 license. A QR code at the bottom provides a direct link to the dataset.

Open Sourcing Xatu Data

25 March 2024 · 853 words · 5 mins ·

Xatu Data Mainnet Sepolia Holesky

samcm DevOps Engineer

savid DevOps Engineer

Introduction

We're thrilled to share that the EthPandaOps Xatu dataset is now open source!

The dataset contains a wealth of information about the Ethereum network, including detailed data on beacon chain events, mempool activity, and canonical chain events.

ethpandaops/xatu-data

Shell ⭐ 13 ⚡ 3

Summary:

- The data is dedicated to the public domain under the CC BY 4.0 license
- The entire schema is available [here](#)
- Data is partitioned by hour or day in [Apache Parquet](#) files

https://data.ethpandaops.io/xatu/mainnet/databases/default/beacon_api_eth_v1_events_block/2024/3/20.parquet



Contribute to Xatu data

The screenshot shows a dark-themed website for "Xatu". At the top left is a logo featuring a panda head and the text "ethPandaOps". The top navigation bar includes links for Posts, Projects, Links, Data, Team, and a search icon. Below the header, a large image of a panda's face is overlaid on a background of abstract, glowing energy fields and geometric shapes.

Contribute to Xatu: Join the Community Data Collection Effort

4 October 2024 · 1216 words · 6 mins · [🔗](#)

Xatu Consensus Layer

samcm
DevOps Engineer
[Twitter](#) [GitHub](#)

savid
DevOps Engineer
[Twitter](#) [GitHub](#)

We're excited to announce that we are opening up the Xatu data collection pipeline to the Ethereum community! This initiative enables community members to contribute valuable data to the Xatu dataset.

As discussions regarding the potential increase in maximum blob count continue **we hope to shed light on the perspective of Ethereum's most crucial participants - home stakers.**

Summary:

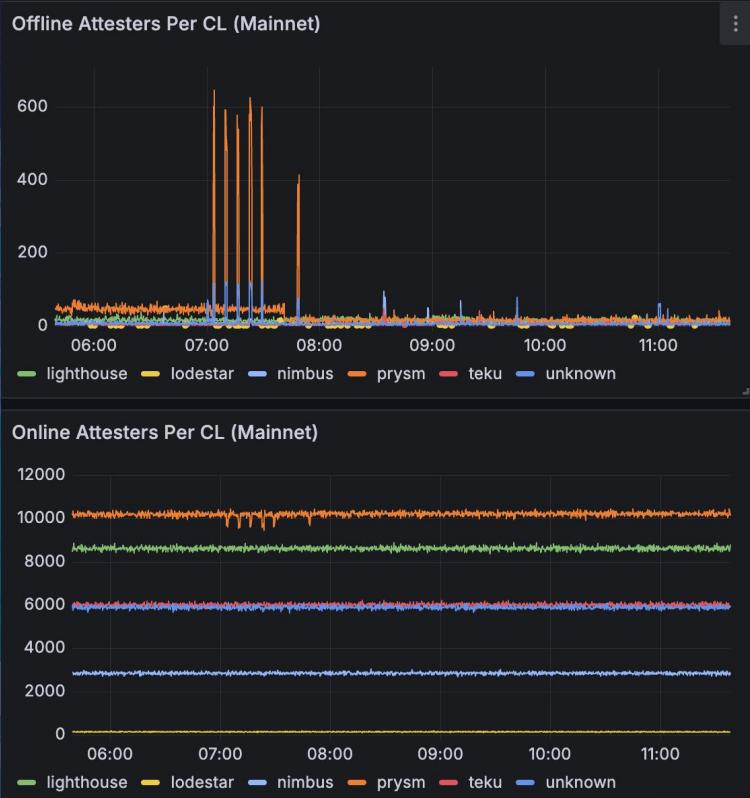
- Privacy focused. Multiple privacy levels that allow contributors to only disclose data they're comfortable with
- Initially restricted to known community members
- Data is published daily by the EthPandaOps team**

On the right side of the page, there is a sidebar with links: Data Collection, Privacy groups, Get Started, and Wrapping Up.



Using the data

EthPandaOps + Sigma Prime



ProbeLab

 ProbeLab

Search

IPFS

Ethereum

- Block Arrival Times
- Methodology
- [Week 2024-45](#)
- Week 2024-44
- Week 2024-43
- Week 2024-42
- Week 2024-41
- Week 2024-40
- Week 2024-39
- Week 2024-38
- Week 2024-29

discv5 weekly reports

- Avail
- Celestia
- Filecoin
- Polkadot
- Tools & Data
- About & Contact

Visit probelab.network

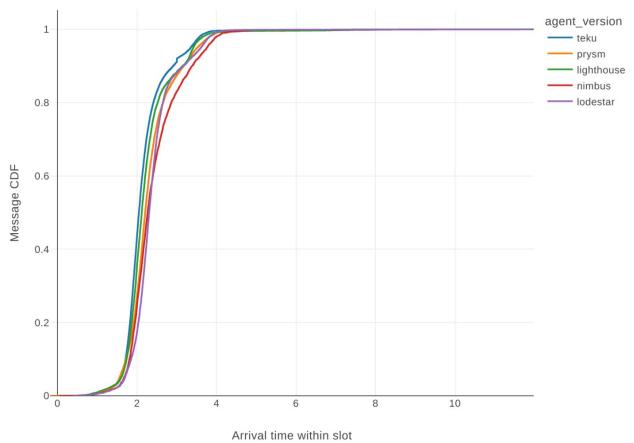
Ethereum block broadcasting latency report 2024-45

The following results show measurement data that were collected in **calendar week 45 of 2024** from 2024-11-04 to 2024-11-11.

This report provides charts and metrics for the beacon block broadcasting latency on the Ethereum network. The methodology we used is available [here](#).

Block arrival time within the slot

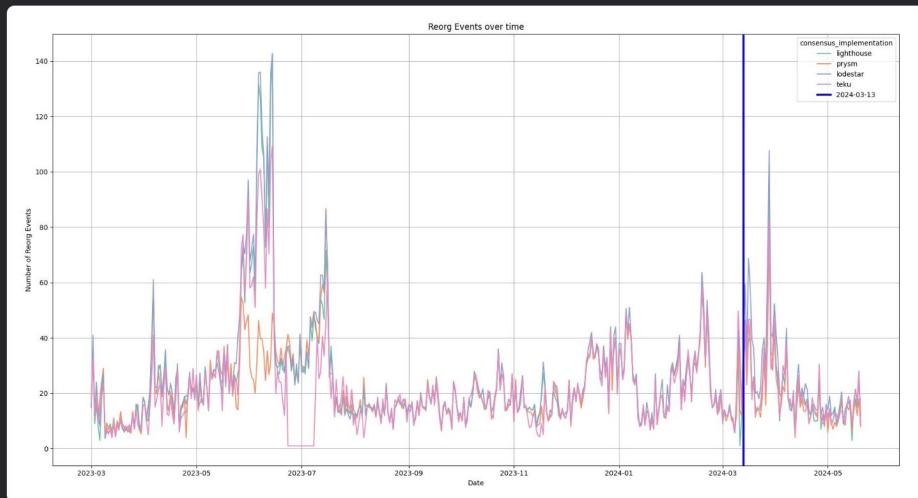
The line plot displays the Cumulative Distribution Function (CDF) of the block arrival times within the slot for the different Ethereum clients (listed [here](#)). The line plot aggregates all the locations we collect data from for each client.



4844 Data Challenge

Blobs, Network Propagation, and Chain Stability

The simplest measurement we could do to discuss the impact of 4844 on chain stability was to investigate how 4844 the hard fork impacted re-orgs:



Though there was an initial uptick in re-orgs, these were due to implementation bugs that were patched shortly after the fork. Since these patches have been released, re-orgs appear to be on par with their pre-4844 levels.



Primev



Analyzing Blob Inclusion Rates and Market Strategies



Primev 0xfa0B



OxEvan April 26th, 2024

Mint

Special thanks to @sui414 for review, @ethpandaops for mempool, beacon chain data, and @envio_indexer for execution layer data.



ethresear.ch

Supporting decentralized staking through more anti-correlation incentives

■ Proof-of-Stake



vbuterin

4 🖊 Mar 27



On Attestations, Block Propagation, and Timing Games

■ Proof-of-Stake ■ mev



Nero_eth ⚡

3 🖊 Aug 14



Big Block Diffusion and Organic Big Blocks on Ethereum

■ Sharding



leobago

Nov 2023



Thank you!

@ethPandaOps
@savid
@samcmAU

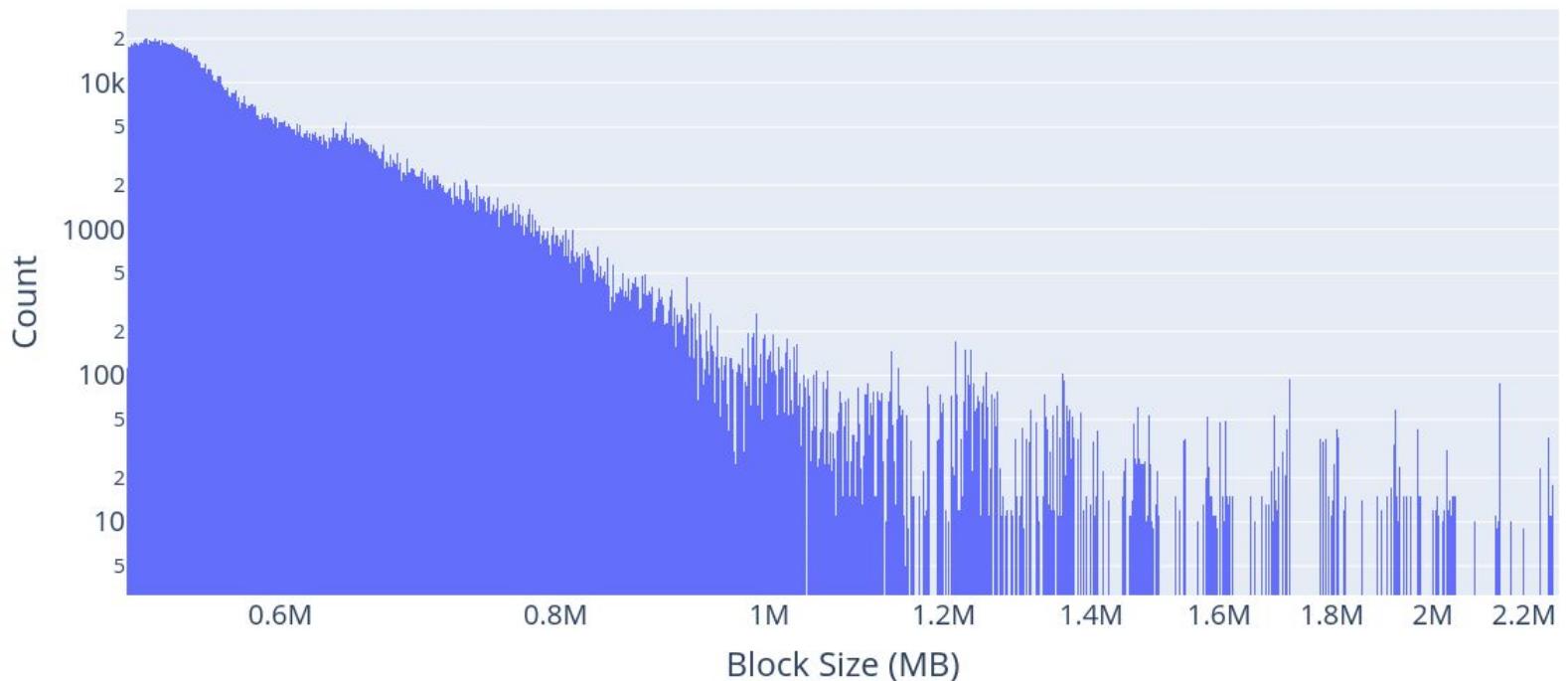
Big Organic Blocks on Mainnet

Leo – MigaLabs



Block Size Distribution

Block Size Distribution



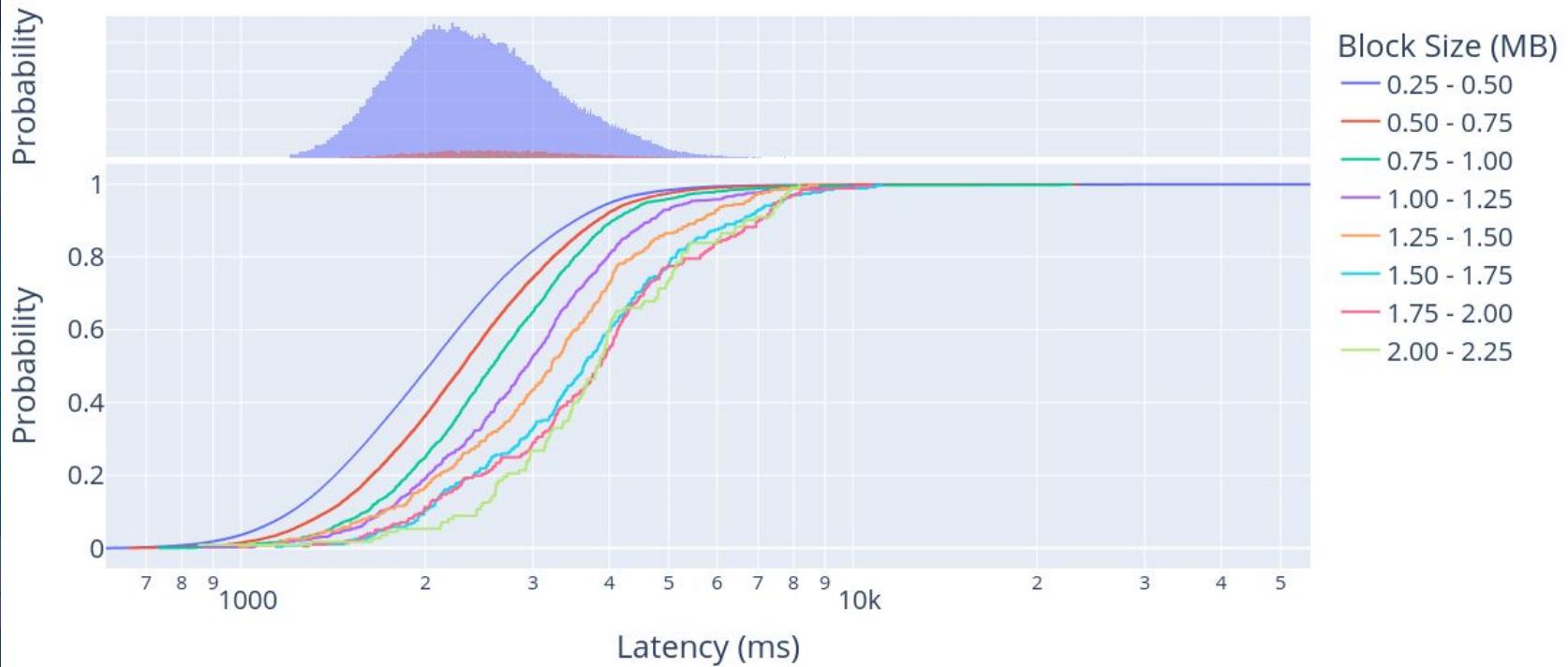
Latency Distribution



Latency Distribution

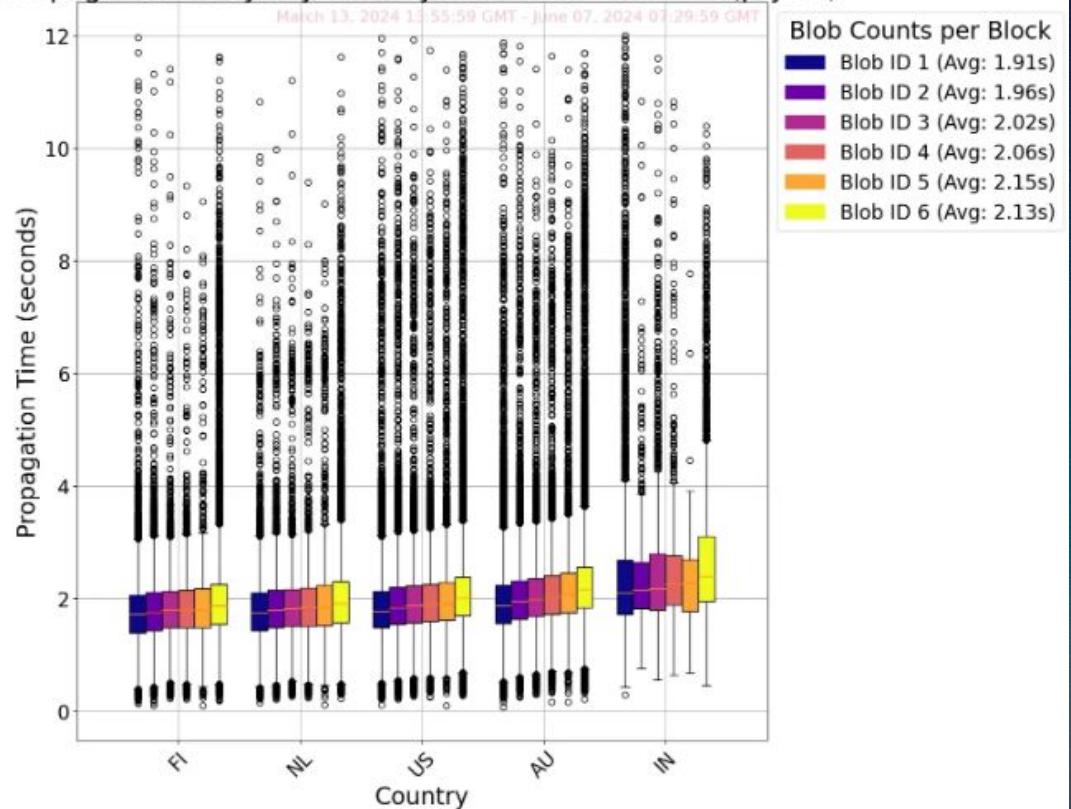
Latency CDF prysm

Latency (ms)



Blob Latency Distribution

Propagation Delays by Country and Block - mainnet (prysm)



Ethereum Blobs

Ethereum Blobs Public Dashboard

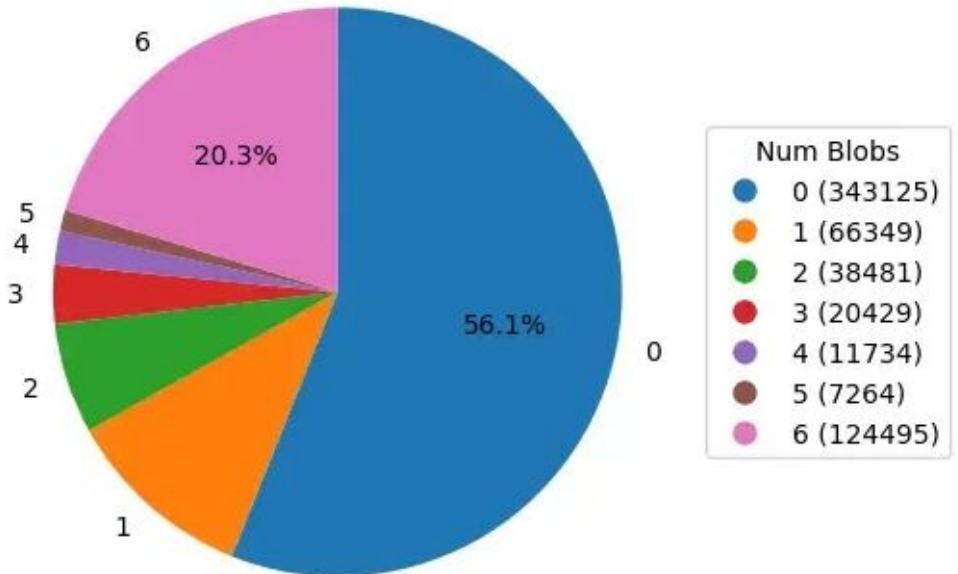
Data retrieved from [Goteth](#) by MigaLabs.

Disclaimer: The information provided in this dashboard only represents a partial view of the data

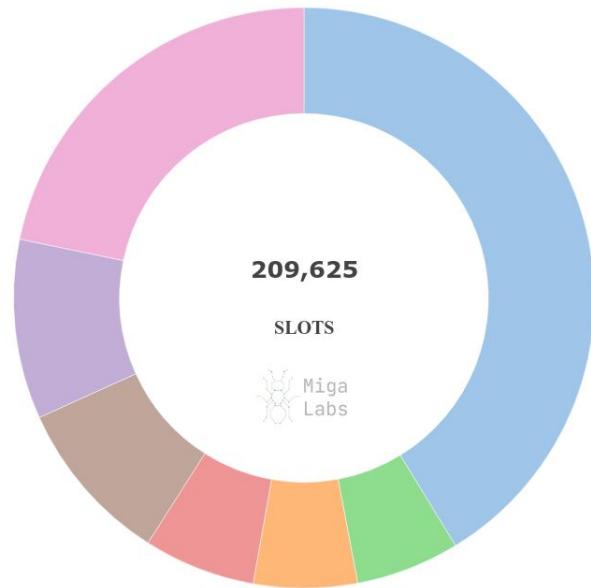


Blob Count and Propagation

Distribution of number of blobs count per slot



Slots by blob count - Latest 6750 epochs (1 month)

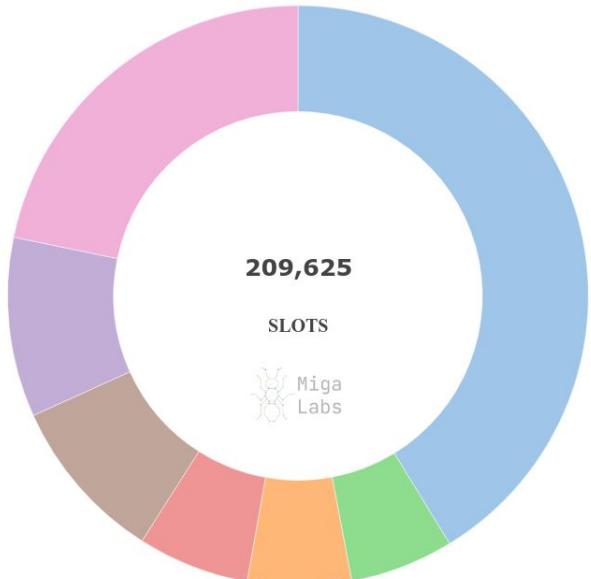


March 13, 2024 13:55:59 GMT - June 07, 2024 07:29:59 GMT

0 (41.262%) 6 (21.763%) 5 (9.994%) 1 (9.224%) 2 (6.216%)
4 (5.777%) 3 (5.764%)

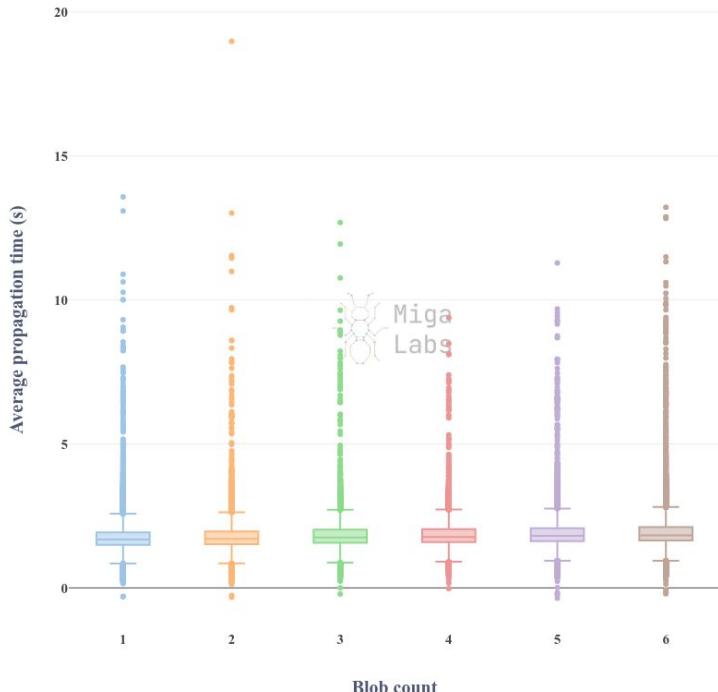
Blob Count and Propagation

Slots by blob count - Latest 6750 epochs (1 month)



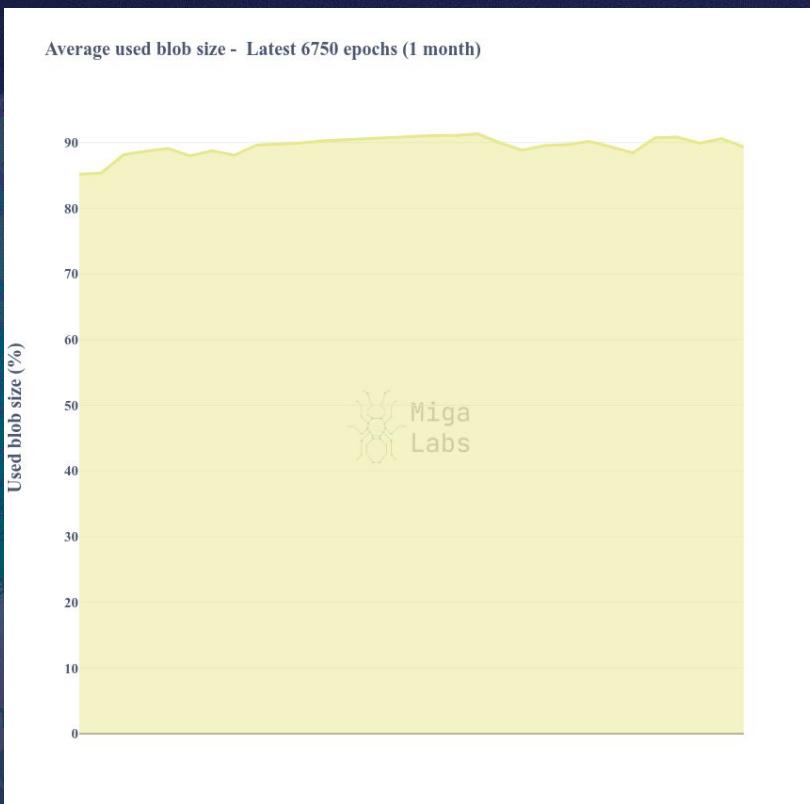
■ 0 (41.262%) ■ 6 (21.763%) ■ 5 (9.994%) ■ 1 (9.224%) ■ 2 (6.216%)
■ 4 (5.777%) ■ 3 (5.764%)

Blob propagation by blob count - Latest 6750 epochs (1 month)



Blob Space Used and Missed Blocks

Average used blob size - Latest 6750 epochs (1 month)



Blob count before missed block - Latest 6750 epochs (1 month)

