Dear ethresear.ch readers,

As part of the Ethereum Foundation's Data Collection Grants Round 2023 which ran between last September and October, an interdisciplinary team involving Nethermind Research and Nethermind core developers received a grant to work on the project "Allowing validators to provide client information privately". Below, we attach our submission in fulfillment of the objectives behind the project. In this deliverable, we have provided the necessary motivation and background for the problem of measuring client diversity, which we then use to propose and analyze three different approaches for validators to privately share their client diversity data—each with their strengths and weaknesses.

Deliverable: Allowing validators to provide client information privately

Executive summary of the proposed approaches.

We briefly summarize the key ideas behind the three approaches before. The reader is referred to the deliverable for a full exposition.

1. Client diversity data on the graffiti field

As the first approach, we have discussed a method for measuring validator client diversity by posting data directly on the graffiti field. We note that this approach has been discussed by the community before. We have outlined necessary changes, such as creating an EngineAPI method for CL clients to retrieve EL client details and agreeing on encoding standards for the data. We have also discussed challenges with this method including dealing with parties that do not participate, multiplexed architectures, and distinguishing between proposer and attester duties.

We have also discussed statistical significance, i.e., how many client data reports are needed to accurately estimate the client distribution from graffiti field data alone. We confirmed that the analyzed method can reach statistical significance quickly (in the order of days) assuming a reasonable participation rate. We discuss these assertions quantitatively in the deliverable.

Finally, we have assessed the feasibility of anonymizing graffiti field reports, concluding that existing methods like encryption or zero-knowledge proofs are impractical to use due to the sequential nature of data collection and the limited space in the graffiti field.

2. Allowing nodes to listen to client diversity data through the gossip network + using nullifiers to hide the identity of validators

As the second approach, we have examined a potential modification to Ethereum's P2P layer to enable crawlers to obtain validator distribution for client diversity. We have explored using a dedicated channel in the GossipSub protocol to share client diversity data efficiently. We have proposed a method that periodically selects validators at random to submit their client diversity data, which is then shared through GossipSub. Each validator forms its client diversity data into a ClientData

object and publishes it via a designated topic. Then, the nodes in this designated topic can receive those objects, verify their authenticity, and aggregate them for the final result. We have also discussed the challenges around this method, particularly concerning network overload.

Furthermore, we have explored anonymizing P2P reports to ensure validators' privacy. We have discussed potential approaches such as encrypting client data or anonymizing the voters' identities using nullifiers and zero-knowledge proofs. We have proposed an approach that uses BLS signatures, nullifiers, and zero-knowledge proofs to hide validators' identities and prevent double submissions. Validators submit encoded client data along with proofs to a P2P network. We have discussed potential deanonymization vectors such as P2P traffic analysis and proposed mitigation strategies like mixnets and approaches based on Dandelion and Dandelion++.

Implementing these strategies may face challenges such as increased latency and complexity. We have stressed our interest in community input regarding the concern level over potential attack vectors and the feasibility of mitigation strategies.

3. Dedicated voting scheme for client data collection

As the third and last approach, we have proposed a voting protocol aimed at collecting data from validators securely and verifiably, avoiding issues like obscurity and centralization found in existing survey methods. We have examined the use of public bulletin boards (PBBs) or blockchains for collecting votes, drawing insights from Vitalik's analysis of blockchains' limitations in elections and the advantages of using blockchains as bulletin boards. Due to its decentralization and cost-efficiency, we have proposed to utilize a blockchain, specifically Ethereum's Holesky Testnet. Regarding how validators submit their votes, we have considered having validators encrypt their client data and share it through a P2P network, and using a trusted committee—called decryption authorities—to receive the encrypted data, submit the received data to a smart contract, and finally, aggregate and decrypt the encrypted client data.

This third method addresses some of the traffic analysis concerns in the second method by leveraging homomorphic encryption of the votes, which requires a trusted committee.

A call for feedback

As the next stage of this research project, we look forward to disseminating and discussing the aforementioned approaches through various channels, including this forum and community calls. Thus, we welcome discussions with the Ethereum community to gauge the impressions on the most suitable approach. For example,

- In the deliverable above, we have provided a rubric that ranks the downsides of each method according to their severity as perceived by the team. From the team's perspective, this analysis positions the second method as the most favorable. Should this rubric be challenged in any way?
- Does the reader see any additional concerns with the proposed methods?
- Are there any variations or suggestions the reader can think of to build upon the methods herein?

We look forward to your impressions and comments!