TLDR

- Currently, 90% of Ethereum POS blocks use MEV-Boost for block building, resulting in an increase of proposer profits by 55%. However,

the design is not perfect.

- The default relay implementation adds significant latency to the block auction, have high operating costs, and benefit from colocation.

This incentivizes vertical integration which in turn reduces proposer fees and Ethereum's security budget.

- The

ultra sound relay

published and deployed an "optimistic relay", removing latency from the builder submission flow and increasing proposer rewards. bloXroute has been running partially optimistic since earlier this year. The next iterations of optimistic relaying, referred to as v2 and v2.5, seek to reduce latency by another 100ms and minimize the reliance on the relay itself.

- The performance improvements of

optimistic relays make them an

inevitable future

since they increase proposer profits, are cheaper and easier to maintain, and are a path towards ePBS.

Frontier Research has deployed an optimistic relay

to demonstrate the improvement in maintainability, and support the diversity of the ecosystem - we are looking for someone interested in helping test and maintain it.

Introduction

Relays are an important piece in the MEV-Boost architecture. They act as a trusted auction house between block-builders and block proposers (validators). Mechanisms to achieve this trust add latency and costs for the relay, reduce proposer fees, and increase protocols reliance on the relay. In this article, we discuss improvements to MEV-Boost relays categorized as

optimistic relaying

which address these shortcomings. Although these improvements come at the cost of increased "optimism" in block-builders we discuss mechanisms to disincentivize dishonest block-builders. We prove that these improvements result in:

- more profitable blockspace,
- reduction in colocation incentives,
- reduction in relay costs, and
- reduction in reliance on relays.

This article is divided into 3 sections, we first discuss the role of relays. Second, we discuss the shortcomings of the current relay design with respect to latency, costs, censorship and block propagation. And finally, we demonstrate 3 improvements to the current relay design, along with their inherent benefits and challenges.

What are relays and how do they work?

MEV acts as a centralisation force across the protocol stack. Due to the sophistication needed to

c

ollect MEV, block proposers are incentivized to specialise and centralize. To counter this incentive, Proposer-Builder Separation (PBS) was accepted as the path forward by the Ethereum community. PBS creates a new

builder

role which is responsible for extracting as much MEV as possible and delivering it to the proposers as

proposer fees

. The

MEV-Boost

architecture was implemented as an out-of-protocol solution to enable PBS in Proof-of-Stake (PoS) Ethereum.

The MEV-Boost architecture aligns the priorities of block-proposers and MEV actors (searchers). Currently, 90% (

in the last 2 weeks

) of Ethereum PoS blocks use MEV-Boost for block building, resulting in an increase of proposer profits by 55% (due to MEV). However, the design is not perfect.

Relays add latency and costs in the system, reduce proposer fees, and increase the reliance on the relay. In the remaining section, we discuss how relays work and in the next section we look at shortcomings in the current relay design, and how they reduce the security budget and increase centralization.

The MEV-Boost architecture enables permissionless participation of multiple block-builders and relays, ensuring a competitive market for maximizing proposer fees. Relays hold an auction for every block where block-builders participate with their bids. The block-builder which proposes the block with the highest proposer fees wins the auction.

Due to access to block contents, the relays can grief block-builders by censoring and stealing transactions, also they can grief the proposers by delivering blocks with suboptimal proposer fees. This is why relays need to be trusted by both block-builders and proposers.

What does the current relay landscape look like?

In the first few months of MEV-Boost, Flashbots was the dominant relay (source

mevboost pics

) covering almost 90% of MEV-Boost block production. As time progressed more relays were launched each having its own unique trait, reducing Flashbots relay dominance to ~25%. Currently, 11 different relays are responsible for producing blocks in the market.

Let us look into the inner workings of a relay with respect to its inherent latency, its costs, censorship resistance and block propagation.

Impact of latency

Relays introduce 2 types of latency (specific data collected from the ultra sound relay) in the system (figure above):

These latencies reduce the effective time of the block auction. In the figure above we can see that if a bid (red bid, Block B) is sent before the end of the auction but cannot account for the simulation latency, it will remain inactive for the block auction.

The above figures display the estimates of latencies due to block delivery and simulation. On average, a builder based in North America will observe a cumulative latency of 270ms, whereas a builder based in Europe will observe a cumulative latency of 160ms (the relay is based in Europe). These latencies reduce the effective time of the block auction. Moreover,

the block delivery latency acts as a forcing function for builders to colocate with relays.

The above figures represent the cost introduced by latencies on block-builders. On average the winning blocks are submitted in the last 0.011s after the end of the auction. The time difference can be positive because on average proposers request the auction results

400ms late

from relays. In the right figure, we show the cost of latency in the system. In 1% of cases, block-builders see an improvement of at least 961

g

w

e

i

/

μ

s

gwei/\mu s

g

w

e

i

/

μ

s

in their bids. For these 1% of cases, a 270ms latency improvement can mean a proposer fee improvement of 0.025 ETH.

These latencies reduce the proposer fees, which is the security budget of Ethereum. Note, for a block-builder even a minor improvement in proposer fees will mean consistently beating the competition and winning more blocks.

Costs of a relay

Every time a relay receives a block bid they need to spend compute resources and simulate the validity of the block content. The need for simulations increases significantly by the end of the auction which causes a significant surge in pricing on the relay. On average, there is a

400% increase in blocks submitted per unit time

in the last 2s of an auction vs the earlier 10s. A conservative estimate

suggests

~$100k/year as compute costs for running a relay.

A bad actor can DoS a relay by submitting multiple block bids consuming unnecessary compute resources. To bypass this several relays use high-priority queues and low-priority queues. Access to the priority queue can be either achieved by being a trusted builder or having a high inclusion rate (blocks accepted over time). The priority queues also have limits on the

number of bid submissions per auction, resulting in an

implicit latency

in the auction.

Censorship by relays

Since relays can access all the block content (last look) and are responsible for making bids active in the auction, they can censor transactions. Relays can both censor transactions due to regulatory requirements (OFAC compliance) or censor transactions of competing block-builders.

Although it's harder to measure if relays are censoring blocks of competing block-builders (using public data), we can measure OFAC compliance of relays. According to

mevwatch.info

, 26% of blocks were OFAC compliant in the last 24hrs, this is down from ~70% censored blocks in Nov 2022. In practice, censorship means longer transaction confirmation times a

recent study

found a median 1 block delay experienced by OFAC transactions.

Relays as block propagators

On April 3rd 2023, a malicious proposer exploited a bug in MEV-Boost relay implementation to steal ~$20M from sandwich bots. The relay code did not verify the signed block header properly before sending the block contents to the proposer. The proposer got access to sandwich bundles and was able to unpack them and steal funds.

Although the

bug was fixed

, a malicious attacker can still pick up block contents while they are being attested, propose a malicious block and race to get the malicious block attested faster (aka

equivocation

). To counter this race relays will now propagate the block to the attesters first and send the block contents to the proposers with a one-second delay (resulting in

missed slots

). This hack has

increased the trust

block-builders and proposers put in the

relays

.

In this section, we saw the current shortcomings in MEV-Boost due to the mechanisms to ensure trust in relays. In the next section let's see how these can be overcome due to optimistic relays.

What are optimistic relays and how do they work?

Optimistic relays reduce the latencies in the system by optimistically assuming honest behaviour by the block-builder. If a block-builder deviates from honest behaviour and causes a loss of funds to the proposer, they are asked to refund the

proposer and their collateral is used as a last resort. Optimistic relays also progressively remove relay responsibilities and

converge to

a system closer to enshrined PBS.

In this section, we discuss 3 proposals to improve relay design and their trade-offs.

Optimistic relaying v1: asynchronous block validation

The first iteration of optimistic relaying moves the simulation of the blocks to an asynchronous task and immediately marks the bid as active in the block auction. This increases the average time of the auction by ~150ms, results in more proposer fees, and reduces the burst simulation costs incurred by relays. A solution which benefits block-builders, relays, and proposers altogether.

Current status:

ultra sound relay has already

implemented

and deployed

optimistic relay v1

. While b

loXroute offers

skipping simulations to their VIP tier customers since last 2 months. The above figure shows

an 8% improvement

in the Ultra sound relay's block contribution after they switched to optimistic relaying. Since other relays are connected to different block-builders and different proposers, they might still win the block auction due to access to valuable private order flow or access to the next block proposer. The figure on right shows a 17% improvement in block submission due to optimistic relaying for the same builder connected to 3 different relays. The improvement in the right figure is more pronounced since it eliminates the impact of private order flow.

This improvement in winning blocks relayed comes due to 2 factors:

At what cost?

This approach comes at the cost of increased optimism in the behaviour of block builders. The builders can construct an invalid block or not pay the fees to grief the proposer. To counter this, the participating builders need to post collateral with the relays, which can be used to refund proposers if they misbehave. This adds

2 types of costs to the architecture:

Optimistic relaying v2: Asynchronous payload delivery

The second iteration of optimistic relaying removes the block delivery latency. V2 makes the bid active for auction as soon as it receives the block header, while asynchronously downloading the remained of the block content. This increases the effective time of auction by ~10ms-100ms depending on the location of the block-builder. This latency introduced by block delivery will further increase after EIP-4844 since it results in larger blocks. Note, v2 will reduce the incentive to colocate but not eliminate it completely.

Historically, Ethereum PoW introduced uncle rewards to compensate for block delivery latency in the p2p network. 1M extra gas per block would have

resulted in a 1.82% increased

chance of a mined block becoming stale. To compensate for this loss and the centralization effect of block delivery, uncle blocks (stale blocks) received 2-5% of the

total mining reward

.

At what cost?

This approach does not change the trust assumptions around the builders compared to the v1.

Optimistic relaying v2.5: Unconditional payments

Until now relays are responsible for ensuring payments are made to the proposer either by simulations or by escrowing collateral. Optimistic relaying v2.5 skips the download of the block body altogether and accepts

unconditional payments

from builders. This reduces the reliance on relays to act as a data-availability layer and a block propagator.

The steps involved with v2.5 are:

Since multiple optimistic relays can rely on the same public escrow contract for funds it will reduce the barriers to entry for builders and reduce the operational costs of the relays.

At what cost?

This approach changes the data availability assumptions for proposers and increases operational costs for builders.

Conclusion

This article explores the role of relays in the Ethereum MEV-Boost ecosystem and proposes several design improvements to optimize the system's performance. We delve deeper into the current relay landscape and discuss the latency in relays, highlighting their costs, the need for priority queues, censorship capabilities and the demand for block propagation.

The article then discussed three designs for improving relay performance, each with its trade-offs.

- Optimistic Relaying v1 removes simulation latency by assuming "honest" behaviour by the block-builder. This approach benefits relays by reducing the simulation costs and benefits builders and proposers by improving the proposer fees. These benefits come at the cost of increased operating costs for relays and increased barriers to entry for builders.
- Optimistic Relaying v2 removes the block delivery latency, putting the block up for auction as soon as the relay receives the block header. This approach benefits the protocol by reducing incentives for colocation and benefits block-builders and proposers by improving the proposer fees.
- Optimistic Relay v2.5 skips the download of blocks altogether and accepts unconditional payments from block-builders. This approach benefits the protocol by reducing dependency on relays. These benefits come at the cost of reduced data availability and increased barriers to entry for block-builders.

In our opinion, optimistic relays are the

inevitable future

since they increase proposer profits are cheaper and easier to maintain and are a path towards ePBS.

References