

Measuring Eth1 network metrics with increased block sizes and topology changes

Project requirements & design May 2020

Background

- EIP-2028 reduced calldata gas cost from 64 to 16 gas/byte, which is likely to result in large block sizes, especially as more rollups are deployed.
- EIP-1559 will double the block gas limit.
- Planned changes such as PoS and stateless Ethereum will likely lead to topology changes, mostly away from large, highly connected mining pools towards a more decentralized, homogeneous network.
- Starkware did some testing and concluded “we see no adverse effect of increasing today the average block size by a factor of 7.5x, nor of increasing the individual blocksize by a factor of 22x”
<https://ethereum-magicians.org/t/eip-2028-transaction-data-gas-cost-reduction/3280/24>.
 - According to James Hancock, “there is concern that the result from Starkware is “too good to be true” and highly dependent on our current network topology. As we move forward into Stateless Ethereum/Eth2 and the topology changes, we want to make sure we aren't prematurely shooting ourselves in the foot.”
- There is concern from the Geth team that the network will not be able to handle a series of large blocks

High level goals - can be addressed in separate stages

- Study the impact of increasing the block gas limit on uncle rate, and implications for network security.
2. Study the impact of changing network topology (away from highly connected block producers) on uncle rate and transaction propagation.
 3. Consider: Put in place a permanent system to collect data on Mainnet performance so that unusual events can be analyzed.
 - Exactly what data to capture? TBD.
 - How valuable/important/urgent is this? TBD.
 4. Consider (Sandra): Create a probabilistic graphical model, a Bayesian network model, to represent, for example Stateless Ethereum, including all the factors that are critical to making the network function, including any interactions or dependencies between the factors. The insights gained from the previous steps would contribute to this model.

5. Consider (Tas): Would it be useful to study the effect of increased block propagation latency due to the possible future addition of a privacy layer in the Eth2 network for at least some nodes? How much latency can be tolerated?
6. Address the test cases specified in <https://eips.ethereum.org/EIPS/eip-2028>

More precise goals and requirements

- Piper thoughts:
 - Gather data suitable for answering current questions related to EIP-1559 and EIP 2028:
 - Relationship between block size, uncle rates, gas limits
 - Gather data suitable for answering future questions for active research topics
 - Network capacity for things like witnesses
 - Topology of the network
 - Understanding transaction and block propagation.
 - Ability to measure how changes impact the network
 - If we had this infrastructure we could have measured the broader effects of [EIP-2464](#) on the network.
 - Future changes that are projected to affect the network can actually be measured.
- Stage 1
 - TBD
- Stage 2
 - TBD

Testing and data analysis methodology

- Piper thoughts:
 - Modify go-ethereum (and maybe other clients) to collect network level information and metrics
 - I need help specifying the exact metrics we should collect. Here is my best guess at a starting point.
 - Transaction propagation
 - For each peer record each time they broadcast a transaction.
 - Block Propagation
 - For each peer record each time they broadcast a block.
 - Transaction and block sizes

- Peer latency (ping/pong)
 - Block execution time
 - Deploy *many* of these nodes into data centers that are globally geographically distributed.
 - Aggregate the collected data into a single database.
 - Use the data to gain insights about the network or to build probabilistic models of the network. Use these to answer the desired questions.
- Horacio's thoughts:

(some Alethio info moved to a comment...)

Some metrics that I'd like to see, as reported by a "probe node" regarding each peer it sees:

 - Bandwidth / transmission speed / retries
 - Protocol support
 - Nodes are incentivized to know about the latest block, but might not care about transactions. Does this happen? Does this mean there are different network topologies to deal with?

Regarding modifying clients to gather extra data: it can be useful, but I'd take that as a strong hint that better measurement capabilities should be standardized. Though that probably is out of scope here.

Alethio's data gathering client and server are open source and MIT-licensed.

1. Tomasz - Stage 1
 - prepare a set of basic formulas
 - i. maximum potential size of a block as a function of x === gas limit and y === isEip1559Enabled
 - ii. average max block size utilization ratio
 - iii. annual max potential block bodies database growth
 - iv. maximum number of blocks synchronized at once at maximum size
 - v. expected block bodies initial sync times one year from now at max blocks
2. Stage 2
 - TBD

Possibly useful tools

- <https://github.com/testground/testground>
- <https://github.com/ConsenSys/wittgenstein>
- <https://github.com/ethereum/hive>
- <https://whiteblock.io/genesis/>
- <https://github.com/Alethio/ethstats-network-server>

Sandra's Bayesian Network Idea

EXAMPLE OOBN - Very incomplete & WIP - put here to show what BN models look like

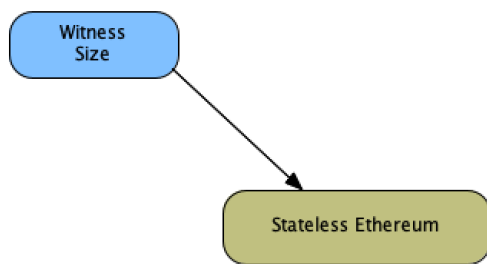


Fig1: High level view of OOBN model

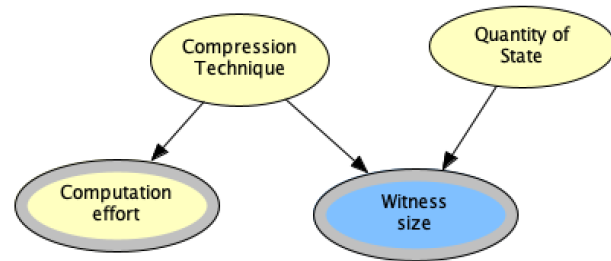


Fig 2: Witness size OOBN submodel

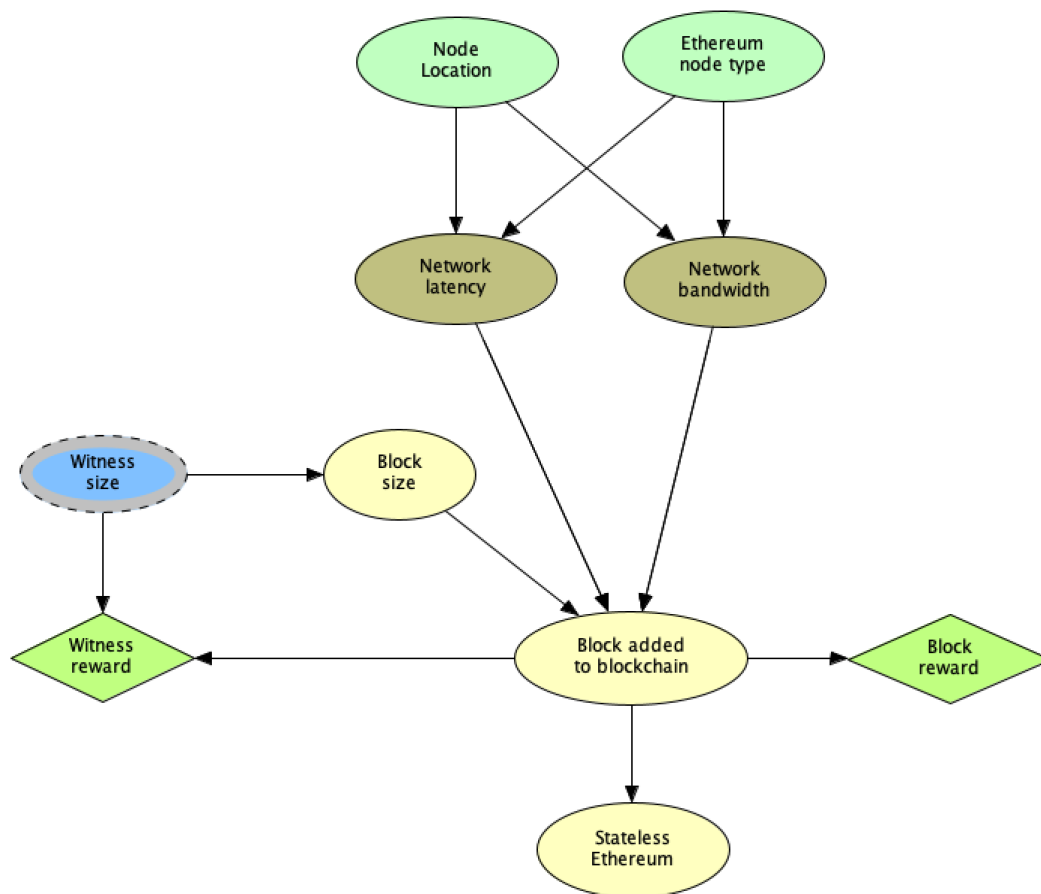


Fig 3: Stateless Ethereum OOBN submodel

So we need to include all the aspects of Stateless Ethereum in the model. I started putting a few thoughts in there, but it is only 10% complete imo, Please add here any additional factors to be added and I (Sandra) will include them

Please also have a think if this approach is of interest. Each of the factors (nodes) have states. For an initial model we would typically have discretised states, e.g. for Stateless Ethereum I have 'Possible' and 'Impossible'. When we run scenarios through the model, we get probabilities. For example, if this is the situation what is the probability that Stateless Ethereum is possible. So you don't get a definite yes or no answer, but rather percentages.