Expected utility under EIP-7521*

Sandra Johnson[†], Kerrie Mengerson[‡], Patrick O'Callaghan

March 4, 2024

Abstract

Immediate settlement or Single-Slot Finality is a long-term goal for ETH. SSF is understood to be incompatible with the present number (800k) of validators due to the computational burden on the network. Ethereum Improvement Proposal 7251 aims to reduce the number of validators by giving stakers the option to merge existing validators. Key to the success of this proposal therefore is whether stakers choose to merge once EIP-7251 is implemented. It is natural to assume stakers participate only if they anticipate greater expected utility (risk-adjusted returns) as a single large validator. In this paper, we compute the various contributions to an individual staker's pay-off under EIP-7251 vs the status quo. We assume EIP-7251 implies no change to the security of the protocol. We confirm that the probability of a validator being selected as block proposer is equivalent under each regime. This result ensures that the decision of one agent to merge has no impact on the returns of another: in turn ensuring there is no major systemic change to the economics of the protocol. In terms of emergent features, EIP-7251 increases the proportion of solo validators can reinvest and compound returns without requiring a multiple of 32ETH. (To a lesser extent, compounding is beneficial to pooled validators by eliminating queueing times for re-entry.) Compounding also alters the cost of slashing. We conclude with an analysis of the participation distribution under different parametrizations of expected utility and size for the population of stakers.

1 Background

With EIP-7251, validators can choose to auto-compound their stake, instead of having any rewards above the 32 ETH effective balance automatically withdrawn into their account.

Therefore validators do not have to wait until 32 ETH has been accumulated before another 32 ETH deposit can be made, but they can automatically compound their stake with rewards earned. The flow-on effect is that this would increase their rewards since rewards are based on the effective balance of the validator, making it attractive to smaller, typically solo, stakers, who have to wait for a very long time to earn enough ETH for another 32 ETH stake. It should also be attractive to larger stakers who regularly earn far greater rewards, and instead of waiting for a newly created validator to be activated in the long activation queue, that stake will immediately

[*] We would like to acknowledge the initial proposers of EIP-7251: Aditya Asgaonkar and Mike Neuder, respectively. Roberto Saltini contributed to the analysis of the proposal and Barnabé Monnot is reviewing and collaborating with the Ethereum Foundation (EF) grant research team on the variation to the original grant proposal. We thank Ben Edgington and Mikhail Kalinin for their insightful comments and feedback. Finally, we acknowledge that this research would not have been possible without EF Academic Grant ID: FY23-1030 for which we thank the Ethereum Foundation.

[†] ConsenSys Software R&D

[‡] Queensland University of Technology

earn additional rewards. On the other hand consolidation and compounding of validators run a risk of increased slashing penalties.

The two key criteria we need to assess are: risk and benefit.

- In other words what are the risks for a range of stakers - from solo to large stakers, assuming some consolidate a portion or all of their stake, and what are the benefits for the various categories of stakers?

- Are the risks and benefits greater for some categories of stakers, and can we deduce the extent to which these potential differences affect various stakers and importantly the health of the ecosystem.

Having a better understanding of the implications of this proposal on the health of the network, would also better inform the minimum variable deposit proposal.

Most importantly, regardless of the assessment of risk and benefit, the appetite for consolidation of stake needs to be ratified by stakers, large and small.

A security analysis has already been undertaken by the Ethereum Foundation (EF) [**?**] (as mentioned above), but we anticipate that additional in-depth analysis would be beneficial.

2 Removing Unnecessary Stress from Ethereum's P2P Network

The key motivation for the initial proposal by Asgaonkar [**?**] to allow consolidation of validators, was concern over the stress on the network due to the growing large number of validators and consequently peer-to-peer (p2p) messages overloading the network.

Asgaonkar [**?**] highlighted the following that would need further investigation before consolidation of validators can be facilitated:

- If large stakers consolidate some or all of their validators, the weights of the validators in the validator set will vary greatly (weight being the number of 32 ETH stakes, i.e. the number of validators with the maximum 32 ETH effective balance, making up one consolidated validator). This will need close examination since several calculations in the protocol are based on a validator's effective balance.

- He suggested that the following changes will be required to the consensus specifications:
  - Increasing MAX_EFFECTIVE_BALANCE from 32 ETH to a greater maximum (it has since been proposed to have 2,048 ETH as the maximum, a 64-fold increase)
  - Instead of validators exiting and re-entering as consolidated validators, providing in-protocol consolidation (what Asgaonkar termed "one-step" consolidation)
  - Allow stakers to partially withdraw stake
  - "Provide resilient staking infrastructure" by building distributed validator technology (DVT).

3 Increase the MAX_EFFECTIVE_BALANCE – a modest proposal

The initial post by Asgaonkar [**?**] calling attention to the need to reduce stress on the network, was followed up by an *ethresear.ch* proposal to increase the maximum effective balance (EB) for validators from 32 ETH to 2,048 ETH [**?**].

The key strategy of the proposal is “validator set contraction”, with the ultimate aim of advancing, or ideally enabling, the feasibility of single-slot finality (SSF) and enshrined proposer builder separation (PBS), while also reducing “strain on the p2p layer” [?].

Advantages from a protocol perspective

- *eliminating redundancy in the p2p layer*
  For every epoch the entire active validator set is partitioned into 32 slot (or attestation) committees. These committees are then split into a maximum of 64 sub-committees with a minimum of 128 validators each, such that each slot in the epoch has the same number of sub-committees. A subset of validators is randomly chosen from each sub-committee to perform signature aggregation [?]. “Each attestation requires a signature verification and must be aggregated. Many of these checks are redundant as they come from different validators running on the same machine and controlled by the same node operator” [?].

- *reducing network load when processing auto-withdrawals*
  There is a continuous sweep of the validator set that automatically withdraws amounts in excess of the MaxEB. Having a large validator set means a higher network load when processing all the partial withdrawals. However, if validators decide to compound their rewards, it will reduce the withdrawal load.

Advantages from a validator perspective

- *rewards can be auto-compounded, meaning that future rewards are earned on a higher effective balance*
  This will benefit solo-stakers because it currently takes around 11 years to earn 32 ETH, so that they can stake one more validator [?]. Large stakers will also benefit from compounding their stake, since their total daily rewards are sufficient to spin up several more validators. Therefore instead of waiting in the activation queue they can earn additional rewards immediately on the larger stake. Neuder calculated that Coinbase currently earns enough ETH in rewards to spin up an additional 9 validators daily [?].

  However I expect that instead of spinning up “new” validators, each of the solo or consolidated validators will just auto-compound to that validator

- *Reducing the operational overhead of managing many validators*
  This is clearly only relevant and beneficial to large stakers, but judging from comments by Lido, this is not really a huge incentive to consolidate.

Perceived disadvantages of proposal

Two main reasons why increasing the maximum effective balance may be challenging to implement [?].:

- “simplicity of current implementation”

- “considerations around committees

However, after addressing each reason, they are judged to be minor concerns [?].

Mechanisms already in place

The following three mechanisms would not need to be altered if the proposal to increase MaxEB goes ahead because they are based on a validator’s effective balance, meaning that validators with a higher stake carry more weight than those with an effective balance of 32 ETH [?].

1. *attesting validator weight* - in the fork-choice rule validators with a higher stake will have "more influence over the canonical chain (as desired)".

2. *proposer selection probability* - the probability of being selected as the next proposer is influenced by the effective balance of proposers.

   Neuder states "This works as intended even with a higher MaxEB, though it will slightly increase the time it takes to calculate the next proposer index (lower values of EB will result in lower probability of selection, thus more iterations of the loop)."

   Intuitively it is hard to believe that this last statement is true, because with the increased MaxEB from 32 to 2,048 ETH, since I would expect the probability of any solo validator being selected will be much reduced.

   Need to work out the change in probability of being selected to ensure that the statement above is true

   Mikhail explained that despite consolidation into larger validators, the proportion of validators that a large staker controls vs a solo staker is still the same. Hence the probability of being chosen as a proposer should still be the same.

3. *sync committee selection* - A sync committee is selected with replacement from the validator set, and each validator in the set has one vote. Neuder points out that "this logic works as intended even with a large increase to the MaxEB."

   This also seems counter intuitive and should be explored in more detail to ensure that this is indeed the case

   Perhaps the next section on security considerations and spec changes will address these concerns

4 Security Considerations & Spec Changes for MAX_EFFECTIVE_BALANCE Increase

This post by D'Amato and Neuder looks in more depth at the implications of the increase in EB on the Ethereum protocol, and especially any security considerations. Moreover, they suggest spec changes to accommodate the increase in MaxEB [?].

4.1 Committees in Ethereum

- *Slot committees*
  Selection of the 32 committees does not take validator effective balance into account. However, when single slot finality is implemented, the fact that different committees may have different weights would not be relevant since all the active stake will vote in each slot.

  Therefore a consolidated validator with a high stake is equally likely to be selected as a solo staker. If the 'super' validator comprises of 5*32 ETH stake, then it can only be selected to be in one slot committee, where alternatively the staker could have had validators in up to 5 slot committees

  Use an example to check whether not taking weight into account could have adverse effects downstream

  Feedback from Barnabé: Probably the adverse effect here would be by chance having one slot with a lot of weight behind it and one slot with not a lot of weight. But with the move to SSF this disappears anyway (all stake votes every slot)

- *Sub-committees*
  The slot committees are split into a maximum of 64 sub-committees each. The sub-committees "facilitate attestation aggregation at the p2p layer" [?].

- *Sync committee*
  The sync committee has a size of 512 and is more persistent than slot committees, being replaced roughly daily.

4.2 The effect of MaxEB on committees

Slot committees

To ensure that the fork choice rule works as designed, the majority (i.e. $> 50\%$) of a slot committee's weight is controlled by honest validators. However, "if a single attestation is majority adversarial, the most an attacker can do is execute a local reorg" [?].

MaxEB increase and stake consolidation

Increasing the maximum effective balance by a factor of 64 from 32 to 2,048 ETH will result in a less homogeneous validator set with respect to individual validator weights, if some stakers consolidate their validators into 'super' validators having a total stake $> 32\ ETH\ but \leqslant 2,048\ ETH$.

The authors are interested to determine whether the probability of slot committees being honestly controlled will be adversely affected by this this proposal.

They propose that the extreme case would be when all validators consolidate their stake to the new MaxEB. This case is clearly infeasible as small-scale stakers will not have 64 validators to consolidate into one 'super' validator. However, if theoretically this could happen, it follows that the validator set will decrease by a factor of 64. Using current (25 August 2023) values from *Ethereum Validator Queue*, the total number of validators is 750,668 (with another 58,757 in the entry queue). Therefore maximum consolidation would equate to a validator set reduction to only 11,729 active validators. The authors state that this large reduction would "weaken the effect of the Law of Large Numbers on the distribution of weight across committees". Therefore the failure probability for this case would give an upper bound for failure probabilities. Saltini calculated the upper bound and details of the calculations can be found in his post [?].

Why Full Consolidation is the Worst-Case Scenario

We summarise here the arguments made by the authors to demonstrate that full consolidation is the worst-case scenario, when casting committee safety as a game between two parties controlling validators: one being adversarial and the other honest. Having full consolidation:

- increases the probability that one or more committees will be majority dishonest

  If all validators (honest and dishonest) have the same MaxEB, why would this be any different to the current probability of a dishonest committee? Surely this is only the case if the dishonest stakers all consolidate, but not all the honest validators consolidate?

  Feedback from Barnabé: I am quite convinced by Roberto's calculations that the probability increases. The coarser partition of stake means that there is more chances that "one of the bins overflows" (one of the slot has more malicious stake than honest)

- increases the variance of adversarial balances

  If all validators are fully consolidated, why would the variance of adversarial balances increase?

- increases the variance of the distribution of adversarial weight over committees

  Similar to above, why is this the case?

- due to the above, it follows that it is more likely that there will be "a positive deviation from the expected adversarial weight in a committee". Apparently although there may also be "negative deviations" the adversarial party is not concerned because it merely aims to control at least one committee.

  In summary, I can only see these statements holding if there is not also maximum consolidation by honest validators

- On the other hand, full consolidation of honest validators maximises qpthe risk of negative deviations from the expected committee weight.

  Why does it work for one and not the other? I think this statement holds true if MaxEB is 32 ETH and honest validators aim to keep their effective balance at 32 ETH so that they have the maximum weight per committee.

- Honest validators benefit from spreading weight evenly across the committees because it minimises the risk of negative deviations.

  If all honest stakers also consolidated their validators to the new MaxEB, their weight would be evenly distributed across all the committees, so this argument only holds in my opinion if one party does full consolidation and the other not

The authors then conclude that the "worst-case scenario occurs when all validators consolidate their stake". After a detailed analysis under the assumption of equal stake distribution among the two groups of validators, Saltini calculated an upper bound for the failure probability and concluded that this probability is maximised when the validator consisted only of fully consolidated validators [**?**].

Safety margins

Based on the assumption that full consolidation is the worst case, the authors generated a graph using a binomial approximation to demonstrate that safety margins remain high under this proposal. They generated graphs for 2x, 4x and 8x reduction.

Sub-committees

Sub-committees are a legacy from the now defunct sharding design that required $\frac{2}{3}$ of the committee weight to be controlled by honest validators. Currently, their key role is in attestation aggregation, which only requires that at least one honest aggregator exists so that attestations from that subnet make it to the global gossip. The gossiped attestations "influence the fork-choice rule, are included in blocks, earn attestation rewards, and lead to justification of epochs" [**?**].

Aggregators are selected via a verifiable random function (VRF) lottery to yield an expected number of 16 aggregators. According to Edgington [**?**], although the target number of aggregators is 16, it is not infeasible that no aggregators are selected, or that more than 16 validators can be selected. Edgington visualises the distribution of aggregators in a histogram, showing that the expected number of aggregators is 16, with a minimum of 0 and a maximum of 32 for a sub-committee of size 256. The selection of aggregators are done as follows: a random number is generated for each sub-committee member, which is deterministic and verifiable. The assumption is that Boneh–Lynn–Shacham (BLS) signatures have random outputs, i.e. the assumption is that they are uniformly random. In other words, a validator signs the slot number, and hashes their signature giving their 32 bytes of randomness. If this random number is divisible by 23 then the member becomes an aggregator. It is a local computation and can be verified. The target is 16 aggregators per sub-committee, but the outcome could be that no aggregators are selected or that more than 16 are selected. Currently there may be a sub-committee with no aggregators every few weeks, which may mean that around 1.6% of attestations per slot committee ($\frac{1}{64}$) may go missing which is not a great concern [**?**]. Moreover, a block proposer, who notices attestations that have not been aggregated may decide to aggregate those signatures for inclusion in their block. The probability of having zero aggregators is the same as having 32 aggregators and the probability of more than 32 decreases rapidly to be negligible.

TO DO: Ben to check the accuracy of what I summarised for aggregator selection

The selection process does not take effective balance into account, therefore if honest validators decide to consolidate and adversarial validators do not, then the probability for an honest validator being chosen as an aggregator is diminished.

To ameliorate this situation, the authors propose to treat consolidated validators as several 'virtual' validators, so that the validator is chosen as an aggregator if one of its 'virtual validators' is chosen.

- Therefore a consolidated validator, v, having effective balance w will consist of $\frac{w}{32}$ virtual validators
- If there are t validators in a sub-committee, assuming all have 32 ETH effective balance, then the probability of being chosen is $\frac{1}{t}$, and if 16 selections are made, then the probability of being chosen as an aggregator for the committee is $\frac{16}{t}$
- For a validator consisting of $\frac{w}{32}$ virtual validators, the probability of being chosen as an aggregator is shown below

$$
\begin{aligned}
&Let\ \ t = total\ \ number\ \ (actual + virtual)\ \ of\ \ validators\ \ in\ \ the\ \ committee\\
&\ \ i.e.\ \frac{W}{32} = t,\ \ where\ \ W = total\ \ weight\ \ of\ \ committee\\
&Let\ \ p = P(selecting\ \ a\ \ validator\ \ as\ \ an\ \ aggregator) = \frac{1}{t}\\
&Let\ \ \frac{w}{32} = b,\ \ (number\ \ of\ \ virtual\ \ validators\ \ for\ \ consolidated\ \ validator, v,\\
&\ \ of\ \ weight\ \ w)\\
&\therefore P(selecting\ \ validator\ \ v\ \ at\ \ any\ \ one\ \ time) = \frac{b}{t}\\
&\therefore P(selecting\ \ validator\ \ v\ \ as\ \ aggregator) = P(selecting\ \ \geqslant 1\ \ of\ \ b\ \ validators)\\
&\therefore P(selecting\ \ validator\ \ v) = 1 - P(not\ \ selecting\ \ any\ \ of\ \ b\ \ validators)\\
&\therefore P(selecting\ \ validator\ \ v) = 1 - \left(\frac{n!}{x!(n-x)!}\right)\left(\frac{b}{t}\right)^{x}\left(1-\frac{b}{t}\right)^{n-x},\\
&\ \ where\ \ n = number\ \ of\ \ draws,\ \ x = number\ \ of\ \ successes,\ \ and\\
&\ \ success = choosing\ \ validator\ \ v\\
&Assume\ \ there\ \ 16\ \ draws\ \ to\ \ select\ \ aggregators\ \ \implies n = 16, x = 0\\
&\therefore P(selecting\ \ validator\ \ v) = 1 - \left(\frac{16!}{0!16!}\right)\left(\frac{b}{t}\right)^{0}\left(\frac{t-b}{t}\right)^{16}\\
&\therefore P(selecting\ \ validator\ \ v) = 1 - \left(1-\frac{b}{t}\right)^{16}
\end{aligned}
$$

I assume that if more than one of the virtual validators of a consolidated validator is chosen, then it is still just that one validator in the set of aggregators, meaning that it could increase the probability of fewer than 16 distinct aggregators?

Feedback from Barnabé: Isn't it already the case that one can be chosen multiple times for the aggregator role? i.e., isn't the sampling already with replacement?

The authors conclude that the following two beneficial properties hold when the selection of aggregators are performed as suggested.

- The total number of aggregators is unchanged, because one consolidated validator would account for only one aggregator.
- The probability of at least one honest validator is equivalent to the current probability, because the introduction of virtual validators means that the probability of selecting an honest aggregator is the same as currently. The total number of honest validators prior to EIP-7251 is the same as the number of virtual and unconsolidated honest validators.

Sync committee

Selecting a sync committee member is proportional to a validator's effective balance, and the authors maintain that sync committee selection will still work as designed.

Once a validator has been selected from the shuffled index, the check for inclusion in the sync committee is given by:

$$if\ effective\_balance * MAX\_RANDOM\_BYTE \geqslant MAX\_EFFECTIVE\_BALANCE * random\_byte :$$
$$sync\_committtee\_indices.append(candidate\_index)$$
$$MAX\_RANDOM\_BYTE = 2^8 - 1 = 31$$

It appears to me that the acceptance probability will change from $\frac{b}{32}$ to $\frac{b}{2{,}048}$, where $b = validator\ effective\ balance$ because the denominator is MaxEB. Hence when the check is done to confirm or reject inclusion, validators with a small stake will be much more unlikely to be accepted as demonstrated by the logic below. Check flaws in my reasoning.

The only disadvantage the authors suggest, would be that the selection process will take longer because there will be more rejections of candidate validators, but they acknowledge that it will also become more unlikely for solo stakers to be selected as sync committee members.

I presume the increased time to select the committee is due to the increase in the expected number of rejections when doing the inclusion check

If solo stakers do turn out to be disadvantaged, how much of a concern is the under representation of solo stakers in sync committees?

4.3 Churn Invariants

The number of validators entering and exiting the validator set is managed by a churn rate that is dependent on the size of the validator set to ensure that it does not grow or shrink by a large number at any time. However, this 'churn limit' has been steadily increasing with the large number of validators in the activation queue. This exacerbates the concern of the network load increasing even more.

Lion [?] proposed an upper churn limit as a temporarily fix [?], but the overarching objective is to devise a mechanism to cap validator set size permanently.

The `CHURN_LIMIT_QUOTIENT` = 65536, and the churn limit is calculated as the integer, or floor, division of the total number of validators by this constant, i.e.

$$upper\ churn\ limit = \left\lfloor \frac{N}{65536} \right\rfloor,\ where$$
$$N = total\ number\ of\ active\ validators$$

We explore the following scenarios:

- $\frac{2}{3}$ of the current validator set size consolidates validators to the proposed MaxEB. Therefore the validator set size will shrink to 258,048 - far more than the desired maximum church limit.

$$\frac{1}{3} of validator set = 250,229$$
$$\frac{2}{3} of validator set = 500,459$$
$$Max\ consolidation\ of\ \frac{2}{3}\ of\ validator\ set = 7,819\ \ validators$$
$$New\ validator\ set = 250,229 + 7,819 = 258,048$$

- If only one third consolidates to MaxEB, the validator set shrinks to 504,368:

  $$\frac{1}{3}ofvalidatorset = 250,229$$

  $$\frac{2}{3}ofvalidatorset = 500,459$$

  $$Max\ consolidation\ of\ \frac{1}{3}\ of\ validator\ set = 3,909\ \ validators$$

  $$New\ validator\ set = 500,459 + 3,909 = 504,368$$

  Still a lot larger than the maximum churn limit.

The authors propose that the churn limit is weight denominated.

I am not sure whether they are taking the proposal from Lion into account when they propose the changes in the blog post.

It looks like this EIP will make it into Dencun so might be worth mentioning it here, but it doesn't change much the calculations, it's a bit like a parameter freeze.

I need to check the pull request to look at details of the proposed spec changes

The authors then look at validator activations and exits, and again propose that validator weight, which in this context means the validators effective balance, into account. Details are provided for the proposed code changes.

They also analyse top-ups for a validator's effective balance when the balance is below 32 ETH. Validator top-ups are capped at 16 per block, which is a lot higher than the validator activation queue rate, and they propose to cap these top-ups at 32 ETH so that validators do not have a loop hole to bypass the churn limit for activations and exits. Validators are currently able to top-up their effective balance regardless of whether it is below 32 ETH. However with the maximum effective balance presently at 32 ETH, there is no benefit to validators toping up their effective balance over the 32 ETH maximum.

4.4 Withdrawals

Currently there is a sweep of the entire validator set to process withdrawals of all effective balances that are over 32 ETH, the current maximum for effective balances.

The proposal is to introduce a new BLS prefix, `0x02`, so that the automatic withdrawal for these validators only happen when their effective balance exceeds MaxEB (2,048 ETH).

However, for this EIP to be acceptable to large stakers, they want to nominate the amount of a partial withdrawal. For example, perhaps a staker wants to maintain a consolidated validator that consists of five 32 ETH validator, then they may want to withdraw rewards in excess of 160 ETH and not the default value of 2,048 ETH. This requirement was already proposed by Asgaonkar in his initial proposal [**?**].

Neuder has a GitHub pull request (PR) that proposes changes to withdrawal processing. According to the proposed changes for consolidated validators, any balance in excess of 2,048 ETH will be included in the automated partial withdrawal sweep.

5 Upper bound on the probability of one majority dishonest committee in the context of MAX_EFFECTIVE_BALANCE increase

In this post Saltini describes the derivation of an upper bound for the probability of having one dishonest committee if EIP-7251 is implemented.

An assumption made for simplification of the calculations is that the weight of dishonest and honest validators are distributed equally amongst each group, i.e. regardless of the proportion of dishonest validators to honest validators, within each category the weights are evenly distributed.

5.1 Exact Probability Formula

The formula was derived as follows:
Let:

$$\begin{aligned}
&W_t = \textit{total amount of stake, i.e. stake of the entire validator set}\\
&\beta < \frac{1}{2}\ \textit{, the proportion of Byzantine validators in the validator set}\\
&h = \textit{total honest validators}\\
&b = \textit{total Byzantine (dishonest) validators}\\
&K = 32\textit{, number of slot committees per epoch}\\
&n_t = h + b\ \textit{, total validators in validator set}\\
&n_c = \frac{n_t}{K}\ \textit{, total validators in each committee}\\
&w_h = \frac{(1-\beta)W_t}{h}\ \textit{, stake of each honest validator}\\
&w_b = \frac{\beta W_t}{b}\ \textit{, stake of each dishonest validator}\\
&h_c = \ \textit{number of honest validators in committee } c\\
&b_c = \ \textit{number of Byzantine validators in committee } c\\
&H_c = h_c w_h\ \textit{, honest weight of committee } c\\
&B_c = b_c w_b\ \textit{, Byzantine weight of weight of committee } c\\
&W_c = H_c + B_c\ \textit{total weight of committee } c
\end{aligned}$$

Then *committee c* will be majority dishonest if the total weight of the Byzantine validators is greater than that of the honest validators, i.e. $B_c > H_c$, and because the assumption of equal weight distribution has been made, it is equivalent to the requirement that the number of Byzantine validators, b_c, for any committee, c, is greater than the number of honest validators, h_c:

$$
\begin{aligned}
&\therefore B_c > H_c \ \textit{iff} \\
&\therefore b_c > n_c \frac{w_h}{W_c} \\
&\therefore b_c > \frac{n_t}{K}\left(\frac{w_h}{w_h + w_b}\right) \\
&\therefore b_c > \left(\frac{h+b}{K}\right) \frac{\left(\frac{(1-\beta)W_t}{h}\right)}{\left(\frac{(1-\beta)W_t}{h}\right) + \left(\frac{\beta W_t}{b}\right)} \\
&\textit{Multiply top and bottom with } \frac{K * h * b}{W_t} \\
&\therefore b_c > \frac{b(h+b)(1-\beta)}{Kb(1-\beta) + Kh\beta} \\
&\therefore b_c > \frac{b(h+b)(1-\beta)}{K(b(1-\beta) + h\beta)}
\end{aligned}
$$

The hypergeometric probability density function (PDF) for x successes in n trials from N items without replacement is given by:

$$
\begin{aligned}
&f(x;k,N,n) = \frac{\binom{x}{k}\binom{N-k}{n-x}}{\binom{N}{n}} \ \textit{where } x = 0,1,2,...,min(n,k) \\
&\textit{a distribution with three parameters: k, N, and n} \\
&n = \textit{number of trials (or draws)} \\
&k = \textit{no of items of one kind, i.e. Byzantine validators} \\
&\textit{and (n-k) of another kind, i.e. honest validators} \\
&N = \textit{population size, i.e. the entire validator set} \\
&mean: \mu = \frac{nk}{N} = np \\
&variance = \sigma^2 = \left(\frac{nk}{N}\right)\left(1 - \frac{k}{N}\right)(N-n)(N-1) \\
&mode = \left\lfloor \frac{(k+1)(n+1)}{(N+2)} \right\rfloor
\end{aligned}
$$

$F(x; N, K, n) = P(X \geqslant x) = 1 - P(X > x)$ is the cumulative distribution function (CDF) for a hypergeometric distribution, where
$N =$ *population size, i.e. active validator set*
$K =$ *total number of success states in population, i.e. total number of Byzantine validators in the active validator set*
$n =$ *total number drawn from population, i.e. the number of validators in a slot committee*
$x =$ *number of observed successes from n draws, i.e. number of Byzantine validators selected for a committee*

The binomial distribution is a good approximation of the hypergeometric distribution when N much larger than n [**?**].

Moreover, for large n, the binomial distribution can be expressed in terms of an F distribution, and Poisson tail probabilities in terms of a Chi-square distribution.

Saltini provides a detailed outline of determining an upper bound for maximising the probability that the total Byzantine weight of a committee is greater than the total weight of the honest validators in that committee [?]. He concludes with the aid of the Chernoff-Hoefding approach to calculating the upper bound of the tail for that the "tail upper bound is maximised by minimising h" [?]. Therefore this happens when we have maximum consolidation, based on the assumption that the ratio of honest validators to dishonest validators is great than $\frac{1}{2}$.

Minimising both Byzantine and honest validators is through maximum consolidation, i.e. all validators have an effective balance of MaxEB. A graph is generated to of the probability of having a majority dishonest committee, using a MaxEB of 8 times the current MaxEB of 32 ETH and with the total validator set weight equal to the total weight at the time of writing. The graph shows both the hypergeometric distribution and the upper bound.

6 Proposer selection probabilities

The expectation is that "Proposer selection is already weighted by the ratio of their effective balance to MAX_EFFECTIVE_BALANCE. Due to the lower probabilities, this change will slightly increase the time it takes to calculate the next proposer index."

6.1 Proposer selection process

proposer selection is a two-stage process:

1. being *selected as the candidate* from the list of shuffled validator indices.
2. passing the *proposer eligibility* check

The *swap-or-not-shuffle* technique is used to shuffle the validator indices in preparation for the selection of a block proposer. The computation to determine the proposer for the next block is the following:

```
def compute_proposer_index(state: BeaconState, indices: Sequence[ValidatorIndex], seed: Bytes32) -> ValidatorIndex:
    """
    Return from 'indices' a random index sampled by effective balance.
    """
    assert len(indices) > 0
    MAX_RANDOM_BYTE = 2**8 - 1
    i = uint64(0)
    total = uint64(len(indices))
    while True:
        candidate_index = indices[compute_shuffled_index(i % total, total, seed)]
        random_byte = hash(seed + uint_to_bytes(uint64(i // 32)))[i % 32]
        effective_balance = state.validators[candidate_index].effective_balance
        if effective_balance * MAX_RANDOM_BYTE >= MAX_EFFECTIVE_BALANCE * random_byte:
            return candidate_index
        i += 1
```

Therefore, we iterate through the shuffled indices, starting with the first entry and then check whether it passes the selection criteria. If it doesn't, then the next validator index in the array goes through the same check.

As we can see from the code, the validator's effective balance (EB) is multiplied by 255 (i.e. MAX_RANDOM_BYTE $= 2^8 - 1 = 255$) and then compared to the product of the generated random_byte, r and MAX_EFFECTIVE_BALANCE $= 2,048$ETH.

Figure 1 (a) is an exposition of random byte values generated for 716,800 validators from the *random_byte* assignment statement below. Superimposed on the histogram of these random byte integers is a uniform distribution. As expected, the random bytes appear to visually resemble values drawn from a uniform distribution: $r \sim U(0, 255)$. We confirm this in the Q–Q plot in Figure 1 (b). Hence we can assume in subsequent calculations that the random bytes, r, have a uniform distribution, $r \sim U(0, 255)$.

```
random_byte = hash(seed + uint_to_bytes(uint64(i // 32)))[i % 32].
```

7 Slashing penalty analysis

The main motivation for this post is the concern of large stakers regarding slashing penalties for consolidated validators. The proposal consists of two parts:

1. Changes to existing penalties:
    - Changing the *initial penalty* to be either fixed, or scaled sublinearly
    - Changing the *correlation penalty* to scale quadratically rather than linearly.
2. Unchanged penalties:
    - *Attestation penalties*
    - *Inactivity leak penalties*

7.1 Initial slashing penalty

Currently this penalty is proportional to the validator's effective balance, and this would result in a penalty of 1 ETH if the validator had 32 ETH. However, if a fully consolidated validator with 2,048 ETH effective balance would incur an initial penalty of 64 ETH.

One proposal is to make this penalty a constant value of 1 ETH, but "If we decide that a constant penalty is insufficient, ..." it can be adjusted so that it monotonically increases with effective balance, e.g. by choosing from the family of polynomials.

How do the authors propose to check whether the initial penalty is sufficient?

The authors demonstrate the increase in initial penalty such that
$\textit{initial penalty} = \frac{EB^x}{32}, \quad x \leqslant 1$

and generate graphs for $x = 1, \frac{15}{16}, \frac{7}{8}, \frac{3}{4}, \frac{1}{2}$ with respect to the graph of a constant initial penalty of 1 ETH [?].

The authors conclude that visually it appears that $x = \frac{3}{4}\ \frac{7}{8}$ are good choices in terms of balancing the size of the initial slashing penalty and the risk for a consolidated validator.

7.2 Correlation penalty

This penalty is incurred roughly 18 days (4,096 epochs) after the slashing event, half way between its exit epoch and the slashing event. The correlation penalty is important in assessing whether there appears to have been a coordinated attack on the chain.

The penalty for a *slashed validator,* v_j, is then calculated as follows by looking at the previous 36 days from this "half-way" epoch:

$$\begin{aligned}
&\textit{correlation penalty} = \frac{(\sum_{i=1}^{n} v_i) * b * v_j}{T}, \;\; \textit{where} \\
&v_j = \textit{effective balance of validator j slashed 18 days ago} \\
&v_i = \textit{effective balance of } i^{th} \textit{ validator slashed during previous 36 days} \\
&n = \textit{number of validators that were slashed during last 36 days} \\
&b = 3 \; \textit{(multiplier changed to 3 in Bellatrix, from 1 in Phase 0, and 2 in Altair)} \\
&T = \textit{total effective balance of the beacon chain}
\end{aligned}$$

For simplicity this is represented as:

$$\begin{aligned}
&\textit{correlation penalty (penalty)} = \frac{3 * EB * SB}{TB}, \;\; \textit{where} \\
&3 = \textit{Bellatrix multiplier} \\
&EB = \textit{slashed validator's effective balance} \\
&SB = \textit{total slashable balance} \\
&TB = \textit{total effective balance of the beacon chain} \\
&\therefore \textit{if } SB = \frac{1}{3} * TB \implies \textit{penalty} = EB \\
&\textit{Similarly, if } 3 * EB * SB < TB \implies \textit{penalty} = 0 \; \textit{due to integer division}
\end{aligned}$$

As designed, there is no correlation penalty for isolated slashing events. The authors point out that this continues to be the case for a fully consolidated validator with 2,048 ETH effective balance.

They based their calculation on the current staked ETH at the time of writing which is 24 million ETH.

$$\begin{aligned}
&SB = EB \textit{ for an isolated slashing} \\
&\therefore 3 * EB * SB = 3 * EB * EB \\
&\textit{Assuming } EB = \textit{2,048 \& } TB = 2.4 * 10^7, \; \textit{then} \\
&3 * EB * EB = 1.2582912 * 10^7 < 2.4 * 10^7 \\
&\therefore \textit{penalty} = 0
\end{aligned}$$

The authors include two graphs to demonstrate how the correlation penalty increases for solo (32 ETH), partially consolidated (256 ETH) and fully consolidated (2,048 ETH) validators. It is important to ensure that any modifications to the function for calculating the correlation penalty satisfies the requirement that when the total slashed balance is $\frac{1}{3}^{rd}$ of the total balance, the entire balance of the validator is slashed.

The authors propose a function that preserves this requirement if the MaxEB proposal is implemented:

$$\begin{aligned}
penalty' &= \frac{3^2 * EB * SB^2}{TB^2} \\
\therefore \textit{if } SB &= \frac{TB}{3}, \textit{ then} \\
penalty' &= \frac{3^2 * EB * \left(\frac{TB}{3}^2\right)}{TB^2} \\
\therefore penalty' &= EB
\end{aligned}$$

Moreover, the new correlation penalty function scales quadratically as opposed to the current function that scales linearly. Clearly with the proposed new function, the slashing penalties are substantially reduced, not only for fully consolidated validator, but also for partially consolidated and solo validators. The authors demonstrate the comparative correlated slashing penalties for these three types of validators using both the current and the proposed functions [**?**].

7.3 Attestation penalty

Once a validator is slashed, their attestations (source, target and head votes) are deemed to be invalid and hence they incur attestation penalties for the 8,192 epochs until their exit epoch.

Different weights are attached to each vote, but only the source (weight = 14) and target (weight = 26) votes incur penalties. For each of the 8,192 epochs the slashed validator will incur:

$$\begin{aligned}
&Given: \\
&base\ reward = \frac{64}{\left\lfloor \sqrt{TB} \right\rfloor},\ weight\ denominator = 64, \\
&source\ weight = 14\ \&\ target\ weight = 26
\end{aligned}$$

$$\begin{aligned}
&\therefore epoch\ attestation\ penalty = \frac{base\ reward * EB * (14 + 26)}{64} \\
&\therefore\ \ if\ TB \approx 24million\ ETH = 2.4 * 10^6 * 10^9\ \ Gwei \\
&the\ integer\ square\ root\ of\ 2.4 * 10^6\ \ ETH = 4,898,\ and \\
&the\ integer\ square\ root\ of\ 2.4 * 10^6 * 10^9\ \ Gwei = 154,919,333 \\
&base\ reward = \frac{64 * 10^9}{154,919,333} = 413\ \ Gwei
\end{aligned}$$

$$\begin{aligned}
&\therefore for\ a\ \textbf{\textit{solo staker}}\ with\ 32\ ETH: \\
&total\ attestation\ penalty\ for\ 8{,}092\ epochs = 8192 * \frac{413 * 32 * 40}{64}\ \ Gwei \\
&\therefore total\ attestation\ penality \approx 6.767 * 10^7\ \ Gwei \approx 0.06767\ \ ETH
\end{aligned}$$

$$\begin{aligned}
&for\ a\ \textbf{\textit{fully consolidated}}\ validator\ with\ 2{,}048\ ETH: \\
&total\ attestation\ penalty\ for\ 8{,}092\ epochs = 8192 * \frac{413 * 2048 * 40}{64} \approx 4.331\ \ ETH
\end{aligned}$$

This attestation penalty for a large slashed staker seems acceptable, but could potentially be adjusted by changing the number of epochs that the validator is deemed as being “offline”.The size of this penalty needs to be such that the security model is not compromised. In other words it should never be a better option to self-slash to avoid inactivity penalties. Therefore, it needs to be greater than the inactivity penalties for an unslashed validator that is exiting and offline [?].

7.4 Inactivity leak penalty

An inactivity leak is currently defined as the situation when the chain has not been finalising for 4 epochs (this value is set by the protocol). Online validators are not penalised when this happens, i.e. no rewards are earned but the penalty is 0. On the other hand, offline validators, which includes slashed validators waiting to exit, start ‘leaking’ state. The loss of stake means that the relative weight of the online validators will increase, which helps the chain to start finalising again. The inactivity penalty can be quite severe.

Using the current method of calculating inactivity leaks, the authors worked out the penalty for validators with three different effective balances: 32 ETH (solo validator), 256 ETH (partially consolidated validator), and 2,048 ETH (fully consolidated validator).

Table 1: inactivity leak penalties

validator size	**16 epoch leak**	**128 epoch leak**	**1024 epoch leak**
32 ETH	0.000259 ETH	0.0157 ETH	1.00 ETH
256 ETH	0.00208 ETH	0.126 ETH	8.01 ETH
2048 ETH	0.0166 ETH	1.01 ETH	64.1 ETH

8 Challenges & Additional Analysis Requirements

- There are many challenges around exiting and activation of validators for consolidation. Therefore, as Asgaonkar points out, in-protocol consolidation needs to be in place [?].
- Ensure that solo stakers are not unduly disadvantaged with respect to being selected as a proposer.
- If more than one of the virtual validators of a consolidated validator is chosen, then it is still just that one validator in the set of aggregators. Does this mean that the probability of fewer than 16 distinct aggregators will increase?
- Currently the proposer receives both the whistleblower reward and the proposer reward. Moreover, the whistleblower reward is proportional to the slashed validator’s effective balance.
  Analyse the repercussions post-MaxEB for the whistleblower reward.
- What is deemed to be an attack on the chain? With the changes to slashing penalties, where has that line moved to? Based on the infrequency of slashing events in the past, are the figures used in the proposed changes too lenient for large stakers?

- What are the negatives for solo stakers who are considered to be the backbone of Ethereum? Does MaxEB swing it too much in favour of large stakers?
- Explore in more detail, e.g. use an example or two, to ascertain the effect of variations in slot committee balances with the inclusion, or omission of fully consolidated stakers. Could there be adverse downstream effects?
- What is being compromised to enable consolidation? Are there other aspects we have not considered?
- Create some example scenarios that include solo, partial and fully consolidated stakers and how different compositions translate to the selection of validators for specific duties and ideally the probability of dishonest committees.
- Create some example scenarios that include solo, partial and fully consolidated stakers and how different compositions translate to the selection of validators for specific duties and ideally the probability of dishonest committees.
- Create some example scenarios that include solo, partial and fully consolidated stakers and how different compositions translate to the selection of validators for specific duties and ideally the probability of dishonest committees.
- Build a Bayesian network (BN) to capture current factors and interactions, including distributions around factors such as uptake of consolidation, current distribution of different categories of stakers, etc.

9 Stakers and Staking pools

There are several helpful documents explaining in detail the different options available for staking:

- Solo staking [?]
- Staking as a service (Saas) [?]
- Pooled staking - "many of these options include what is known as 'liquid staking' which involves an ERC-20 liquidity token that represents the staked ETH." [?]
- Centralised exchanges [?]

ethereum.org provides a *Comparison of staking options*, outlining the key differences between solo staking, staking as a service and pooled staking. The legal implications of staking are discussed in Consensys blog post *Staking is Data Validation, Not Investment* [?].

Re-staking is deemed to be out of scope for this grant research topic, but there are good resources to explain the concept and discuss the implications of re-staking. A blog post by Buterin *Don't overload Ethereum's consensus* discusses the various approaches to re-staking and explains why some techniques bring "high systemic risks to the ecosystem" [?]. Another interesting perspective is from the Substack post by Tripoli: *Endgame Perils of Restaking*, and the addendum to the post: *Addendum: Endgame Perils of Restaking.*

The Ethereum Foundation and EtherScan conducted a survey of stakers [?], according to the following categories: solo staking, staking pool service, mini node operator (Rocketpool), and those not yet staking. The responses were analysed and the key trends, take aways, and predictions reported. The large majority of the respondents were either solo stakers or mini pool operators.

Justin Drake and Mike Neuder also had conversations with stakers. The biggest stakers they talked to were Coinbase, a few different Lido operators and Kraken. They concluded that there are probably a few reasons to motivate bigger stakers to consolidate their validators if the maximum deposit size is increased from 32 ETH to 2,048 ETH. Arguably this proposal is aimed largely at the bigger stakers rather than the smaller, solo stakers. The reasons they may want to consolidate are: to reduce operational cost, although it is likely not particularly high infrastructure costs, Mike Neuder estimated in the region of a few $100 per week, but key management would be the more challenging side of running many validators that can be consolidated into fewer validators if this proposal is implemented. The questions raised by Lido and the responses provided by the EF are in the documents listed above.

Meetings with Infura, Consensys staking and MetaMask (MM) staking were very insightful and the range of infrastructure configurations was discussed as well as the types of staking. MM staking is mainly targeting integration with fractional liquid staking pools and working closely with a fractional staking provider so that users can easily participate in staking via MM. Estimates of costs to validators are more complex, depending on the staking pool and staking pool costs vary too. In general larger staking pools can keep per validator costs reasonably low due to economies of scale.

In conclusion, the overheads for small stakers are probably slightly higher than staking pools, when we work out the average cost per validator, i.e. spinning up and running a 32 ETH validator.

According to the Staking Launchpad on June 2023:

- Total ETH staked: 19,357,605 ETH
- Total validators: 606,947
- Current annual percentage rate (APR): 4.74%

9.1 Data sources & visualisations

Lido

Lido is a popular liquid staking pool. At the time of writing, 7 June 2023, the total amount of ETH staked with Lido is 7,127,430 ETH.

Rocket Pool

A Rocket Pool node only needs to stake 16 ETH, and this stake is then coupled with 16 ETH from the staking pool to create a validator, known as a “minipool”. Staking in the Rocket Pool can be a stake of a little as 0.01 ETH, which is then issued as an rETH token representing the amount deposited into the pool. A node operator would obtain staked ETH from the pool to form a validator.

There is documentation at the Rocket Pool website that clearly explains Ethereum staking and how it works when staking with Rocket Pool.

Currently, 8 June 2023, Rocket Pool has staked 722,176 ETH and has 2,842 node operators.

If we can identify the number of distinct rETH token holders, we would be able to determine the number of stakers staking ETH in Rocket Pool.

9.2 Staker wealth and distribution

We can at best have an informed guess about current staker wealth distribution. Some staking pools can be identified, but even then we may not be aware of all of the validator nodes and validators that they run. However, the largest staking pool, Lido, is transparent regarding its overall control of validators. Lido operates on a trusted setup, and Consensys staking runs several Lido operators.

10 Validators

10.1 Validator slashing and penalties

Slashing

The situations that lead to a validator being slashed are few, but they are severe violations of protocol rules that may be considered as a potential solo or coordinated attack on the system. Regardless of the reason for the slashing event, they are all handled in the same way. There is an initial slashing penalty, which is currently set at 1 ETH (or $\frac{1}{32}$ of the stake) and this is followed 18 days after the slashing event by another penalty, known as the correlation penalty. The purpose of the latter is to penalise what may be a coordinated attack on the chain. Therefore the correlation penalty takes into account slashings 18 days before and after the slashing event.

A validator gets slashed when they are reported and evidence of the violation of the rules is included in a beacon block. For the valid reporting of a slashing event, the reporting validator receives a reward. The intention is that this will help incentivise the reporting of slashable events. Only one proposer slashing can be included in a report, whereas multiple attestation violations can be included in a report. When the slashing is included in a block, the proposer gets a reward which is a fraction of the effective balance of the validator being slashed (currently $\frac{1}{512}$). Up to 16 proposer slashings can be included in a block and up to 2 attester slashing reports. Therefore, if several slashings have occurred, including these reports in a block can generate a generous reward for the proposer.

For more detailed information on slashing and calculations, please refer to the latest version of Edgington's *Upgrading Ethereum* book [**?**] which incorporates the updates included in the Capella hard fork. In summary the events that lead to slashing are:

1. "making two differing attestations for the same target checkpoint"
2. "making an attestation whose source and target votes 'surround' those in another attestation from the same validator.
3. "proposing more than one distinct block at the same height"
4. "attesting to different head blocks, with the same source and target checkpoints"

The first two relate to Casper Friendly finality gadget (FFG) consensus, and the latter two are related to Latest message driven (LMD) Greedy Heaviest-Observed Sub-Tree (GHOST) consensus.

Edgington points out that 'slashable behaviours relate to "equivocation", which is when a validator contradicts something it previously advertised to the network'. Hence it is important for validators to ensure that they do not 'accidentally' equivocate. This could theoretically happen as a result of bugs in client software, but the vast majority of slashings have been due to node operators running two different nodes using the same validator keys. The reason may have been to improve uptime, but the risk of slashing is too great compared to any potential benefit in uptime [**?**]. There was also an incident where a validator exploited a vulnerability in a relay operator running mev-boost, an open source proposer-builder separation protocol. Flashbots posted a detailed post-mortem of the event.

Apart from the slashing penalties, a slashed validator accrues attestation penalties until such time as they exit, which is not until $2^{13}\ epochs = 8,192\ epochs \approx 36\ days$ after being slashed. Moreover, if there is an inactivity leak at the time, the penalties imposed on slashed validators will be higher. Slashed validators cannot earn any attestation rewards while waiting to exit. It seems rather odd, but a slashed validator can still be elected to be the proposer for the next block. However, their block will be deemed to be invalid. The only duty for which they could receive a small reward is if they are selected to be in the sync committee, but the probability of this happening is very small.

Penalties

Slashing is the most severe penalty a validator is subjected to and as explained they can accrue several extra penalties while they wait to exit. However, there are a number of smaller, less serious 'misdemeanours', or failure to perform their duties that can result in penalties for validators. The validator's stake is reduced by the penalty and the ETH is burnt, thereby reducing net issuance [**?**].

Attestation penalties

- Missed source and target votes (i.e. missed Casper FFG votes), but no penalty for a missed head vote.
- Incorrect source vote, then target vote is missed.
- Incorrect source or target vote, then head vote is missed.

Sync committee penalties

- Non-participation of a member incurs a penalty equivalent to the reward they would have received if it was correct

10.2 Griefing/discouragement attacks

According to Buterin a griefing attack is when a validator acts maliciously inside a consensus mechanism to reduce other validators' revenue even at some cost to themselves to encourage the victims to drop out of the mechanism [**?**].

The two main motivations for reducing the number of participants are most likely because fewer participants:

- mean greater rewards for those remaining in the mechanism
- helps to prepare an attack on the chain by reducing the cost of an attack

Some strategies have already been put in place to avoid discouragement attacks [**?**]:

- inverse square root scaling of validator rewards
- scaling of rewards with participation (viz. for each "source, target, and head vote, the attester's reward is scaled by the proportion of the total stake that made the same vote")
- zeroing attestation rewards during an inactivity leak
- rate limiting of validator exists, which means that an attacker needs to sustain an attack for longer and at greater cost in order to achieve the same outcome.

10.3 Centralisation forces

Centralisation needs to be assessed within the context of this grant and the proposed increase in maximum effective balance, especially with respect to large stakers and staking pools.

10.4 Health of the Ethereum ecosystem

Ether alpha combines various aspects of the ecosystem to give an overall impression of the health of the network. They pull information from various sources for 'Project Sunshine' dashboard [**?**].

Historic and current trends for stake concentration are important to observe as these are warning signs that the chain is becoming more vulnerable to collusion. The consolidation of stake through EIP-7251 (currently in draft form) [**?**] is unlikely to change the dynamics of stake concentration, since it is encouraging consolidation of validators already being operated by stakers. However, with more node operators and validators joining larger staking pools staked ETH will become more one-sided in favour of staking pools.

The Rated network also provides detailed metrics and visualisations to gauge the health of the network [**?**].

11 Bibliography