Sybil Attacks on Auction Based Inclusion Lists (AUCIL)

Abhimanyu Nag • 9 Jun 2026

> Special thanks to Sarisht Wadhwa and collaborators on the AUCIL paper for initial discussions and conversations that led to this post

Inclusion lists (ILs) are the talk of the town as a censorship resistance primitive for Ethereum under proposer builder separation (PBS). Two major designs have been developed as per my understanding:

1. **FOCIL**, which relies on underlying consensus with fixed slot wise participants and no explicit pricing, and
2. **AUCIL**, which is an auction based inclusion list design that relies on economic incentives with strategic participation and price formation.

With FOCIL (EIP-7805) proposed to headline the forthcoming Hegota upgrade, the robustness of IL designs is an immediate concern. It is worth noting that neither design is fully Sybil proof or bribery proof.

FOCIL is inherently sybil resistant at committee selection, but its fee mechanisms can weaken once fake transactions and bribery are allowed. Fortunately, Stouka, Ma and Thiery [1] gave a comprehensive analysis for bribery attacks and safe mechanism design for FOCIL but I do not remember seeing a corresponding analysis for AUCIL.

This blog asks only two questions:

1. *Where can Sybil attacks enter AUCIL and how do they change their censorship guarantees and incentives?* and
2. *What would a Sybil proof AUCIL mechanism for Ethereum look like? How can we improve this?*

Quick Background For The Uninitiated

AUCIL by Wadhwa et al. [2] operates with a mempool $M = m_1, \ldots, m_m$ and a committee of n IL proposers $P_1, \ldots, P_n$. Each proposer may submit an input list $L_i \subseteq M$ containing at most k transactions, with nk bounded by the available block capacity. A transaction m_j offers an inclusion fee f_j, which is also treated

as its utility $u_j = f_j$ (keeping all other utilities constant). The protocol assumes rational parties, a common mempool and a synchronous network (at most Δ delay). I will informally lay out the design (please refer to the original paper for proofs).

Phase I: Input list allocation

Rather than letting every proposer independently choose the highest fee transactions like in FOCIL, the protocol computes a recommended allocation $L = (L_1, \ldots, L_n)$ using a greedy algorithm. Let n_j be the number of proposers assigned transaction m_j, and let γ denote the probability that an assigned input list becomes available. Since the fee is divided among the proposers contributing the same transaction, proposer P_i's expected input-list utility is

$$U_i(L_i) = \sum_{m_j \in L_i} \frac{f_j}{1 + \gamma(n_j - 1)}$$

The greedy allocation repeatedly assigns the transaction with the greatest marginal payoff. At the resulting correlated equilibrium, replacing an assigned transaction $m_a \in L_i$ with an unassigned transaction $m_b \notin L_i$ cannot increase P_i's utility:

$$\frac{f_a}{1 + \gamma(n_a - 1)} \geq \frac{f_b}{1 + \gamma n_b}$$

Thus, transaction coverage and proposer rewards depend directly on the number of proposer identities assigned to each transaction.

Phase II: Aggregation using Auction (Is it really an Auction?)

Each proposer broadcasts its input list together with an availability flag $F_i \in 0, 1$ and obtains a private VRF generated bias

$$b_i \sim \text{Uniform}[0, b_{\max}]$$

Proposers aggregate the available input lists into candidate inclusion lists. If candidate i contains y_i available input lists, its auction score is

$$\ell_i = y_i + b_i$$

(So this is an example of a scoring auction then?)

The block proposer selects the highest scored valid candidate and proves that it beats at least $n - \theta$ other bids, where $\theta - 1$ is the number of tolerated crashes. The winning aggregator receives an aggregation reward u_{agg}, split between the

bid selected immediately and the highest bid observed over later rounds. Proposers also receive their shares of the transaction fees contained in the winning list.

The paper proves that guaranteed censorship through aggregation stage bribery costs at least

$$\frac{(n-\theta)u_{\text{agg}}}{2}$$

Nice right?

Okay So What If I Pull Up With My Friends?

From standard mechanism design literature, Incentive Compatibility $\neq$ Sybil Proofness and I have some claims that make me believe that it is the same for AUCIL too.

More formally, let A be an attacker who controls $q < n$ members of the IL proposers.

Definition 1 *AUCIL will be considered sybil proof if, for every number $q \geq 1$ of controlled proposers and every coordinated strategy of those identities,*

$$U_A^{(q)} \leq U_A^{(1)}$$

Thus, splitting one principal into several identities must not increase its total reward or reduce its cost of censoring a transaction.

We already know for a fact that this is violated in both phases because there exists an AUCIL list allocation that is stable against every individual proposer but unstable against an attacker controlling two proposer identities.

Example. Let

$$n = 3, \qquad k = 1, \qquad \gamma = 1$$

and consider two transactions:

$$f_a = 10, \qquad f_b = 6$$

Algorithm 1 selects transaction a, then b, then a, because the successive marginal rewards are

$$10 > 6, 6 > \frac{10}{2} = 5, \frac{10}{2} = 5 > \frac{6}{2} = 3$$

Suppose the resulting allocation is

$$L_1 = a, \qquad L_2 = b, \qquad L_3 = a.$$

Each identity assigned a receives

$$\frac{10}{2} = 5$$

while the identity assigned b receives 6.

Now we see that no individual attack is profitable (IC property):

$$a \to b : \qquad 5 > \frac{6}{2} = 3,$$

and

$$b \to a : \qquad 6 > \frac{10}{3}.$$

Now suppose A controls identities 1 and 3. Its prescribed utility is

$$U_A = 5 + 5 = 10.$$

If identity 1 remains on a while identity 3 switches to b, A obtains

$$10 + \frac{6}{2}$$

Thus,

$$U'_A > U_A,$$

even though the deviating identity individually loses utility:

$$3 < 5.$$

The attacker A accepts that loss because the attack raises the reward of its other identity from 5 to 10.

Therefore, Algorithm 1 can be IC with greedy allocation and equilibrium while failing to be sybil proof. More interestingly not that the multiplicity of transaction a falls from

$$n_a = 2$$

to

$$n'_a = 1.$$

A Sybil attack can therefore reduce the number of input lists protecting a transaction, even without an external bribe. Let us formalize this a bit more:

Lemma 1 (Phase I Sybil attack) *Suppose h independent proposers and q identities controlled by A are assigned transaction t, which pays utility u_t. The total utility of A from t is*

$$V_q(t) = \frac{qu_t}{1 + \gamma(h + q - 1)}$$

Moreover,

$$V_q(t) - V_1(t) = \frac{u_t(q-1)\big(1 + \gamma(h-1)\big)}{\big(1 + \gamma(h+q-1)\big)(1+\gamma h)} \geq 0,$$

with strict inequality except when $h = 0$ and $\gamma = 1$. Hence AUCIL's fee sharing rule is not sybil proof.
If t is assigned to n_t identities, of which s_t are already controlled by the censor, then a guaranteed Phase I attack costs at most

$$C_{\mathrm{I}}^{\mathrm{Syb}}(t) \leq (n_t - s_t)u_t,$$

since paying each remaining identity u_t dominates its maximum possible reward from t.

Looks hard but the utility expression simply follows by summing AUCIL's fee shares and the difference is obtained by direct subtraction. For censorship, the s_t controlled identities omit t freely and only the remaining $n_t - s_t$ identities require compensation.

AUCIL's original Phase I guarantee is

$$C_{\mathrm{I}}^{0}(t) \geq (n_t - 1)u_t$$

Thus, for $s_t \geq 2$,

$$C_{\mathrm{I}}^{\mathrm{Syb}}(t) \leq (n_t - s_t)u_t < (n_t - 1)u_t,$$

so the original lower bound no longer holds. The exact new cost depends on the replacement transactions available to each proposer.

Lemma 2 (Phase II Sybil attack) *Suppose the censor controls s of the n selected aggregators. Under AUCIL's bribery model, its additional external bribery cost for certain censorship is*

$$C_{\mathrm{II}}^{\mathrm{Syb}}(s) = [n - \theta - s]_+ \frac{u_{\mathrm{agg}}}{2}$$

Therefore,

$$C_{\mathrm{II}}^{0} - C_{\mathrm{II}}^{\mathrm{Syb}}(s) = \min\{s, n - \theta\} \frac{u_{\mathrm{agg}}}{2},$$

where

$$C_{\mathrm{II}}^{0} = (n - \theta) \frac{u_{\mathrm{agg}}}{2}$$

is AUCIL's original aggregation bound.

What Does This Mean and Where Do We Go?

We have barely scratched the surface on an analysis of the ramifications of a sybil attack on AUCIL. There's tons of other ideas such as a whole theory of Sybil Proof Auction Based Inclusion Lists that we can solve for and we also have not grounded this in formal mechanism design theory such as optimal auctions as well as welfare analysis among the different proposers (a huge income inequality between the proposers will definitely incentivise competition and attacks) and maybe we can improve on the scoring auction set up here (equitable auctions is a thread we can see?). One big improvement that I can see that can solve the problem is using stake based partitions of fees instead which will also discourage sybils and improve aggregation complexity as well in a way. More concretely:

1. Can we use stake based partitions for fees and improve upon the correlated equilibrium greedy algorithm to optimize for equity?
2. Can we improve the scoring auction design in Phase II to deal with the Sybil Proofness case as well? How about if instead of the VRF sample, we can design a stake weighted lottery?

I hope AUCIL is implemented sometime in the future and I am excited to see the implementation of FOCIL this time around. Thanks for reading.

I have been Abhimanyu Nag.

References

1. Stouka, A. P., Ma, J., & Thiery, T. (2025). Multiple proposer transaction fee mechanism design: Robust incentives against censorship and bribery. arXiv preprint arXiv:2505.13751.

2. Wadhwa, S., Ma, J., Thiery, T., Monnot, B., Zanolini, L., Zhang, F., & Nayak, K. (2025). AUCIL: An Inclusion List Design for Rational Parties. Cryptology ePrint Archive.