Selfish Mining under General Stochastic Rewards

Maryam Bahrani[⋆], **Mike Neuder**[†→‡], Matt Weinberg[‡]

Advances in Financial Technology

Thursday – October 9th, 2025

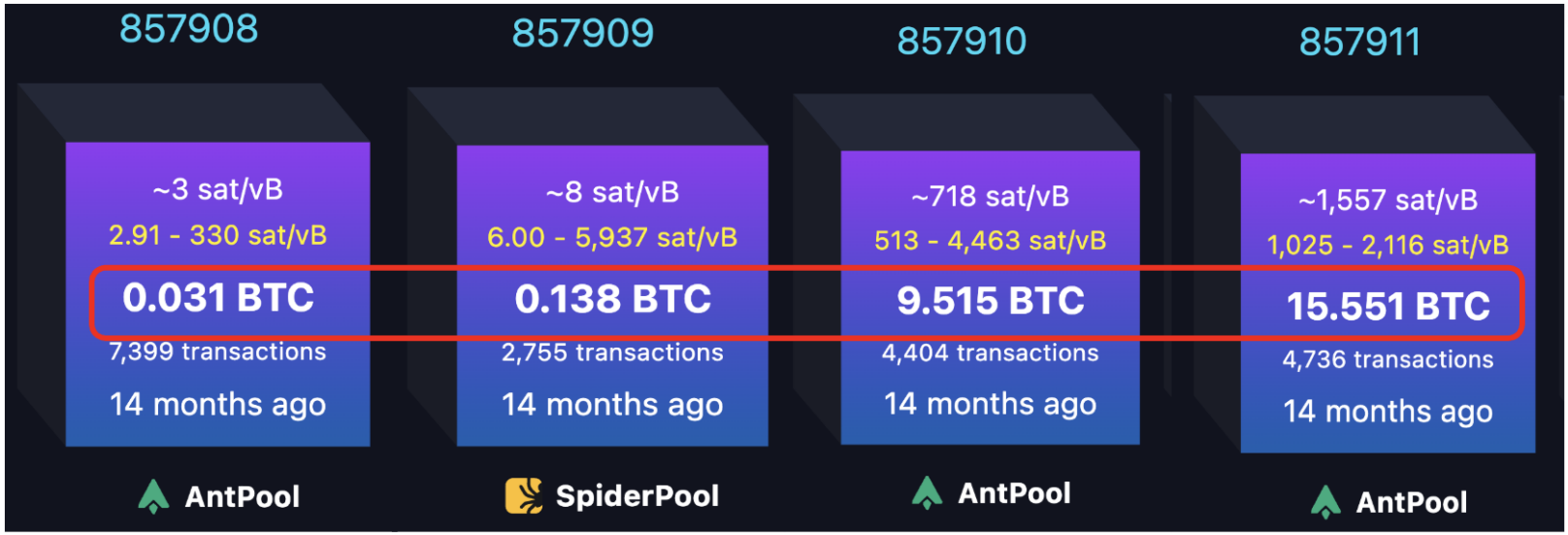
[⋆]Ritual, [†]Ethereum Foundation, [‡]Princeton University

Motivation

- Incentives of consensus mechanisms can depend on exogenous events.
- Selfish mining, where miners seek to earn disproportionately high rewards, should account for that.

Motivating example #1

The launch of Babylon



- In August 2024, bitcoin fees spiked 500× from 0.031 → 15.551 BTC within four blocks.
- The block reward during that period was 3.125 BTC.

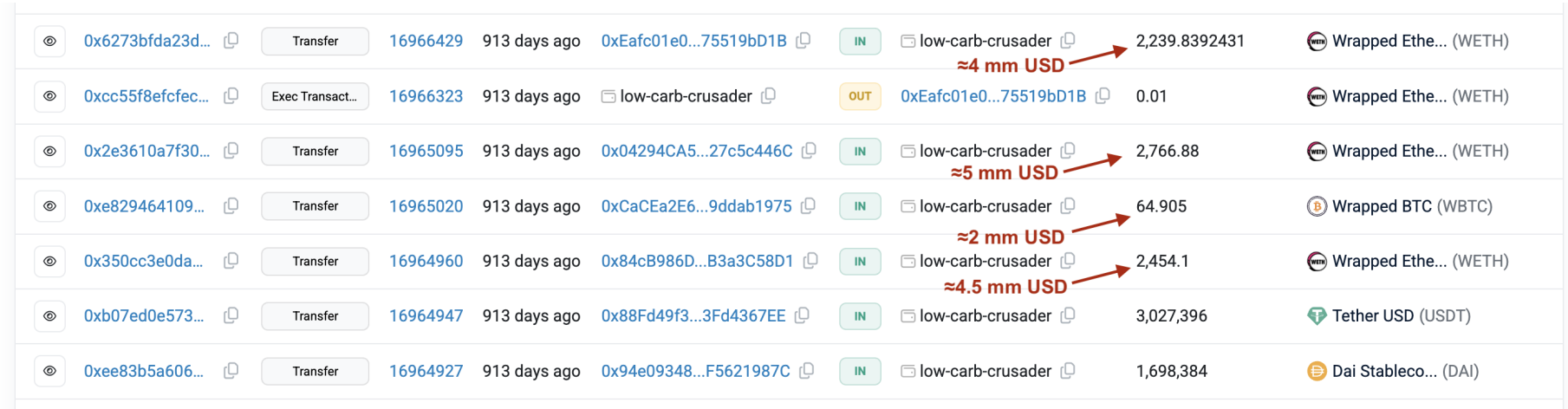
https://mempool.space/block/00000000000000000000151faeaa14ca333a9a5edc3fa7da906413d27a1fe2532

Motivating example #1

The launch of Babylon

Motivating example #2

The Low-Carb Crusador



- In April 2023, an Ethereum validator exploited MEV software to extract 20mm USD worth of assets from other trading bots.
- The attack was enabled by participating in Ethereum consensus. The validator signed conflicting blocks, but the slashing penalty was 1 ETH.

https://ethresear.ch/t/equivocation-attacks-in-mev-boost-and-epbs/15338

Motivation

Related work

- The incentives of consensus mechanisms can depend **heavily** on exogenous events.
- Yet, the selfish mining literature has largely focused on the *block reward regime.*
- [ES'13] show that selfish miners earn disproportionately high block rewards.
- [SSZ'16] find optimal selfish mining strategies.
- Many papers generalize to other consensus protocols, but still focus on maximizing endogenous (to the consensus mechanism) rewards (e.g., Tezos [NMRP'19], Algorand [FHWY'22], Ethereum [SNMATT'22]).

[ES'13/18] Ittay Eyal and Emin Gün Sirer. Majority is not enough: Bitcoin mining is vulnerable. Communications of the ACM, 61(7):95–102, 2018. Original preprint from 2013.

[SSZ'16] Ayelet Sapirshtein, Yonatan Sompolinsky, and Aviv Zohar. Optimal selfish mining strategies in bitcoin. In Financial Cryptography and Data Security: 20th International Conference, FC 2016, Christ Church, Barbados, February 22–26, 2016, Revised Selected Papers 20, pages 515–532. Springer, 2016.

Motivation

Related work cont'd.

- There are a few notable exceptions!
- [CKWN'16] show that selfish mining is still profitable (and actually even more profitable) when only considering transaction fees.
- [ZHET'23] show that "whale transactions" (those which pay extremely high transaction fees) can also lower profitability thresholds.
- Quoting [FNGA'25] "we find that only 3 works include transaction fees in their modeling; 2 consider both block rewards and transaction fees."

[CKWN'16] Miles Carlsten, Harry Kalodner, S Matthew Weinberg, and Arvind Narayanan. On the instability of bitcoin without the block reward. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, pages 154–167, 2016.

[ZHET'23] Roi Bar Zur, Ameer Abu-Hanna, Ittay Eyal, and Aviv Tamar. Werlman: To tackle whale (transactions), go deep (RL). In 44th IEEE Symposium on Security and Privacy, SP 2023, San Francisco, CA, USA, May 21-25, 2023, pages 93–110. IEEE, 2023. doi:10.1109/SP46215. 2023.10179444.

[FNGA'25] Colin Finkbeiner, Mohamed E Najd, Julia Guskind, and Ghada Almashaqbeh. SoK: Time to be selfless?! demystifying the landscape of selfish mining strategies and models. Cryptology ePrint Archive, 2025.

Our contributions

- **Practical:** We analyze selfish mining under a reward function encompassing block rewards, transaction fees, and MEV spikes.
  - Answers the pragmatic question: "how vulnerable is Proof-of-Work to selfish mining under combined rewards?"
  - Painting a more unified picture between thresholds of $\alpha = 0.329$ from [SSZ'16] and $\alpha = 0.07$ from [CKWN'16].
- to do $\uparrow$, we also develop a more general framework $\downarrow$
- **Modeling**: We characterize *properties* of reward functions and perform case studies demonstrating how the properties arise under different assumptions.
- **Methodology**: We present a new technique for calculating expected attacker profits under general stochastic reward functions.
- This talk focuses mainly on the modeling contributions, while showing a flavor of the methodological and practical results.

Model

Nakamoto Consensus Game

- We study the *Nakamoto Consensus Game*.
- A set of miners M, where at time t, each $m \in M$ has...
  - the *public view*, V_t, and
  - a *private view*, $V_t^m \supseteq V_t$.
- In each round, every miner picks a block to mine on.
- The wining miner can publish their block (update the public view) or keep it private; **(new)** this decision now factors in exogenous rewards.
- **(new)** Because rewards are time-dependent, we have to explicitly model difficulty adjustment.

Model

Reward function examples

- Miners make decisions about how to mine and publish based on the *rewards* of various blocks.
  - *Block rewards* are constant at 3.125 BTC.
  - *Transaction fees* are modeled as linear in time since parent block $t - \texttt{timestamp(B)}$ in [CKWN'16].
  - *MEV/Whale transactions* are modeled as the result of a Bernouli trial

$$R = \begin{cases} 10 & \text{if } X = 1 \\ 0 & \text{otherwise} \end{cases}, \quad \text{where } X \sim \text{Bernoulli}(p),$$

    in [ZHET'23].
- Our model is general enough to capture all of these.

Model

Static rewards

- Static rewards depend on the *time since the parent block*.

Definition (Static Rewards)

A reward function R is *static* if:

- for all $\Delta > 0$,
- all times t_1, t_2, and parent blocks B_1, B_2 such that $\texttt{timestamp}(B_1) = t_1 - \Delta$ and $\texttt{timestamp}(B_2) = t_2 - \Delta$, we have
- for all valid blocks $B' \in \mathcal{B}(t_1, V_1, B_1, r)$, we have

$$\Pr_r[R(t_1, B_1, r, B') = x] = \Pr_r[R(t_2, B_2, r, B') = x]$$

for all x.

- We can rewrite static reward functions as a function of Δ.
- Note that static rewards can still be *random* and/or *non-linear* in Δ.

Transaction fee example

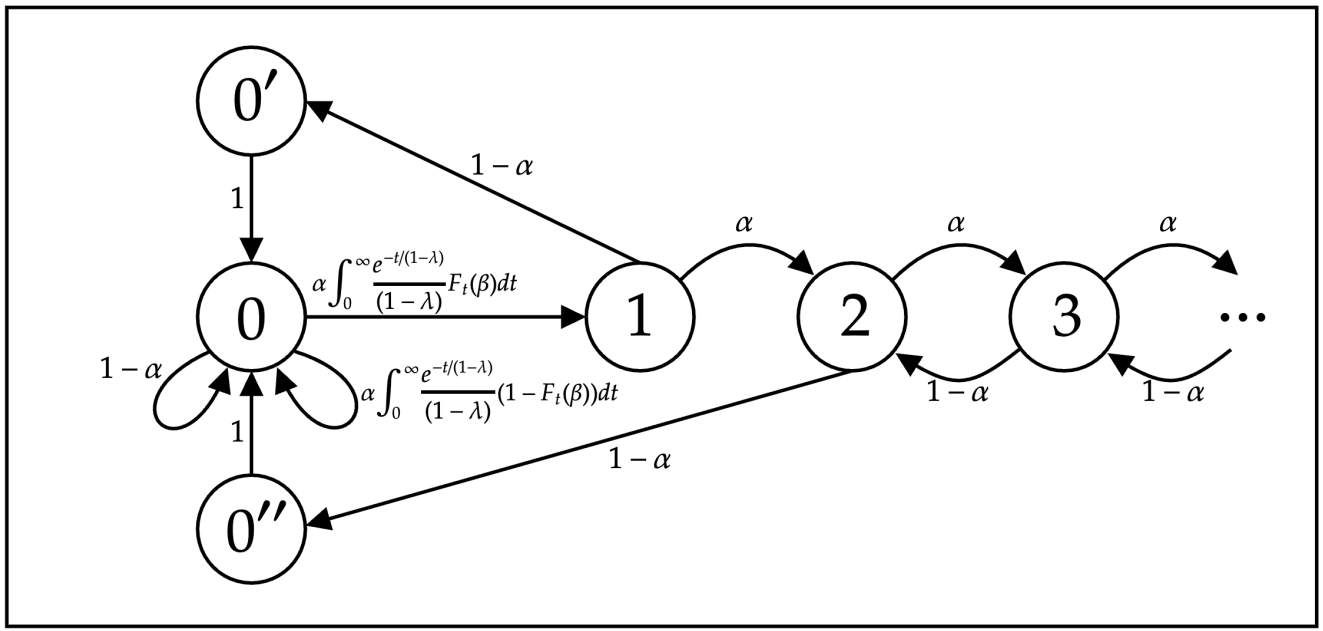
Static rewards cont'd.

- If transaction fees arrive at a **constant rate** (i.e., no contention),
- and blocks are **fully claiming** (i.e., no congestion),
- then transaction fees are static.
- Conversely, if transactions arrive at different rates during different times of the day (e.g., higher trading volume during daytime hours in Asia), then they are not static.
- Also, if blocks are finite capacity and transactions are patient, then transaction fees are not static.

Methodology

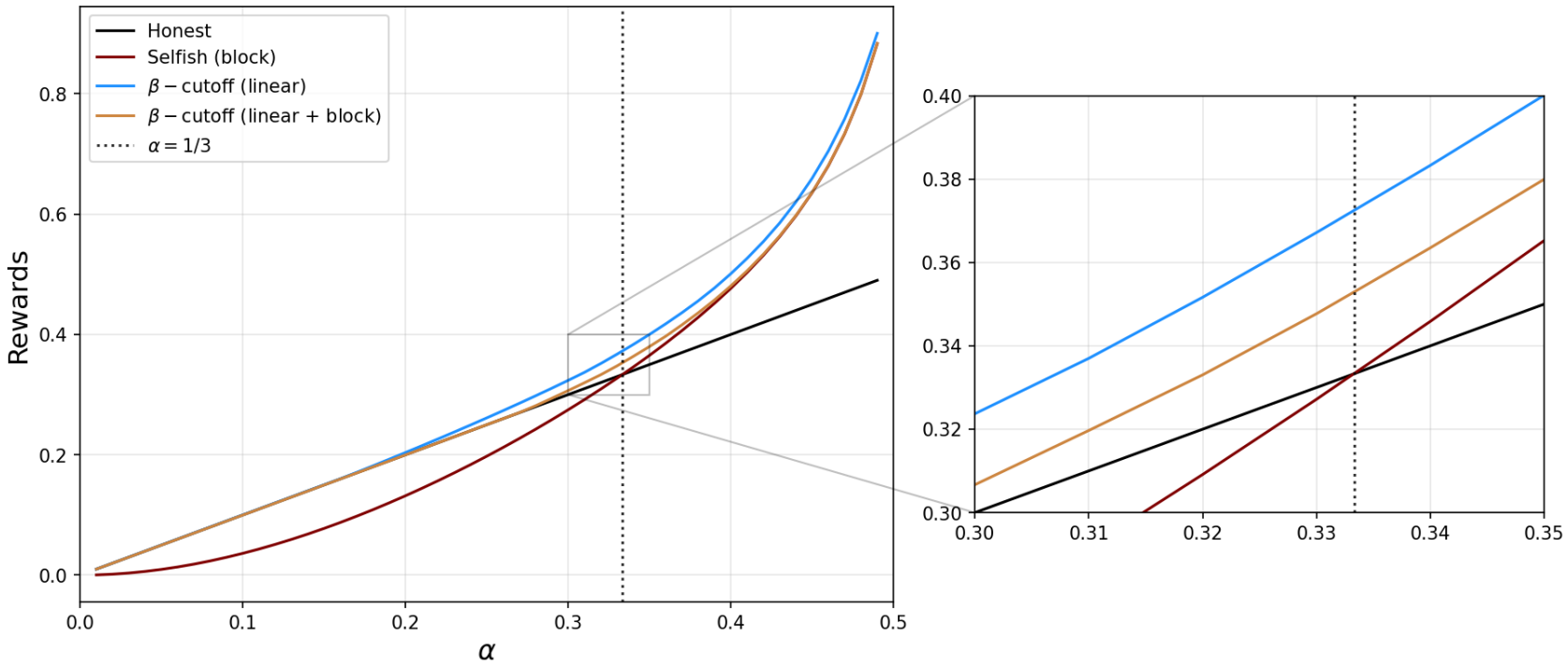
Markov Chain

- We use a similar Markov Chain to [ES'13] and incorporate the "β-cutoff parameter" from [CKWN'16].
- Each state indicates how much longer the attacker chain is than the canonical chain.



Results

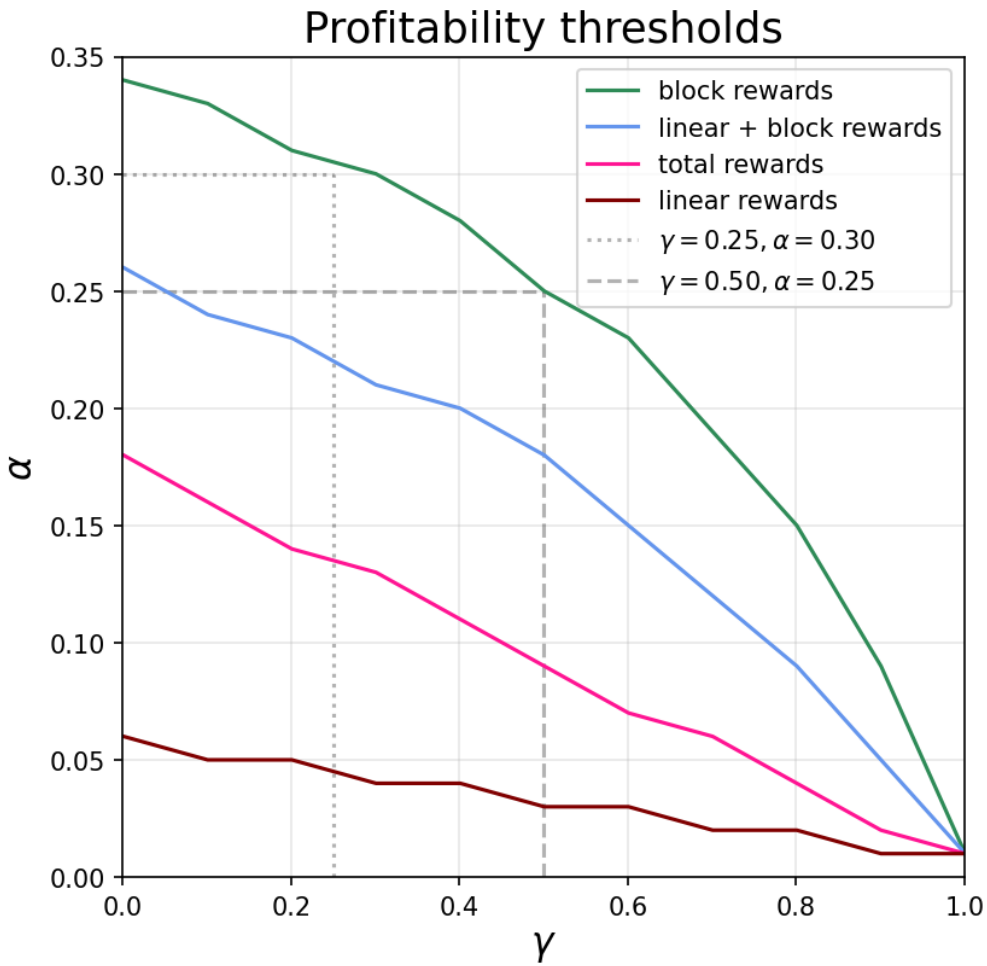
Block rewards and linear transaction fees



- We combine the two most basic reward functions from [ES'13] (block rewards) and [CKWN'16] (linear transaction fees).
- Selfish mining is not profitable until $\alpha > 1/3$, but both other reward functions become profitable much earlier.

Results cont'd

Profitability thresholds



- **Key point**: Different reward functions yield *extremely* different assessments of protocol vulnerability.
- Green, just block rewards (i.e., [ES'13]), looks pretty safe.
- Red, just linear rewards (i.e., [CKWN'16]), looks pretty unsafe.
- Pink, the combined reward function (i.e., this paper) looks moderately unsafe.

thanks :)