WeRLman: To Tackle Whale (Transactions), Go Deep (RL)

Roi Bar-Zur, Ameer Abu-Hanna, Ittay Eyal and Aviv Tamar

Abstract

Blockchain technology is responsible for the emergence of cryptocurrencies, such as Bitcoin and Ethereum. The security of a blockchain protocol relies on the incentives of its participants. Selfish mining is a form of deviation from the protocol where a participant can gain more than her fair share. Previous analyses of selfish mining make easing, non-realistic assumptions. We introduce a more realistic model with varying block rewards in the form of transaction fees. However, this comes at the cost of an intractable state space. To solve the complex model, we introduce WeRLman, a novel method based on deep Reinforcement Learning. Using WeRLman, we show reward variability can significantly hurt blockchain security.

Motivation

$1,590,000,000,000

Secured in blockchain based cryptocurrencies*

To consider a blockchain secure, it's important to understand when miners are incentivized to follow it.

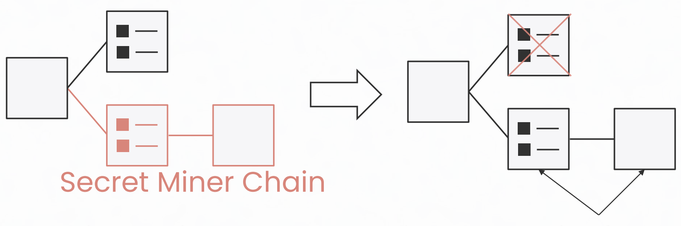
The security threshold is the minimal miner size required to deviate and gain more than the fair share.

Previous work assumes constant block rewards. But transaction fees cause rewards to vary.

* coinmarketcap.com, May 2022

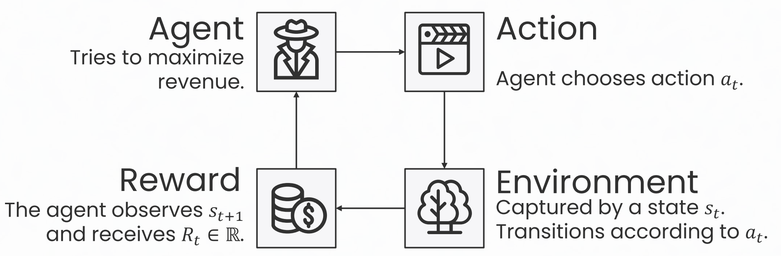
Model

Selfish Mining



Unlike previous work, we consider transaction fees!

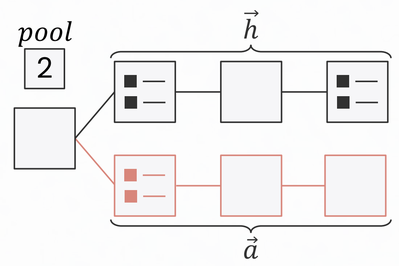
Markov Decision Process (MDP)



Selfish Mining as an MDP

State space:

$$(\vec{a}, \vec{h}, fork, pool)$$



Action space:

- $adopt\ \ell$ – Discard private chain and adopt ℓ blocks.
- $reveal\ \ell$ – Publish ℓ blocks.
- $wait$ – Keep mining on private chain, a new block will be found by either the miner or the rest of the network.

Method

Calculate $Q^*(s,a)$, the optimal revenue after performing action a in state s. Use the 1-step lookahead Bellman equation:

$$Q^*(s,a) = R(s) + \gamma \mathbb{E}\left[\max_{a'} Q^*(s',a') \middle| s' \sim s,a\right]$$

When the state space is too large, train a neural network to learn $Q^*(s,a)$.

Use Monte Carlo tree search to sample simulated trajectories and estimate a deeper lookahead.

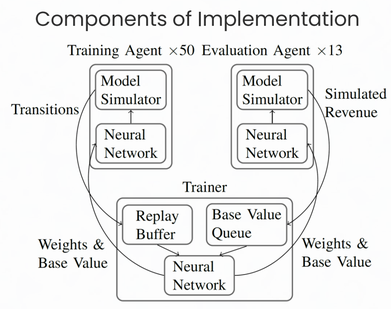
Use two additional techniques:

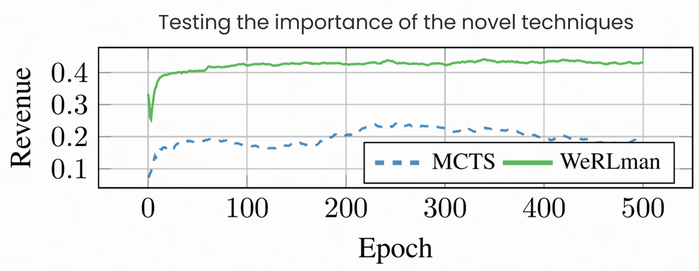
1. Base value

$$Q'(s,a) = \rho + (1-\gamma)Q(s,a)$$

2. Target value normalization

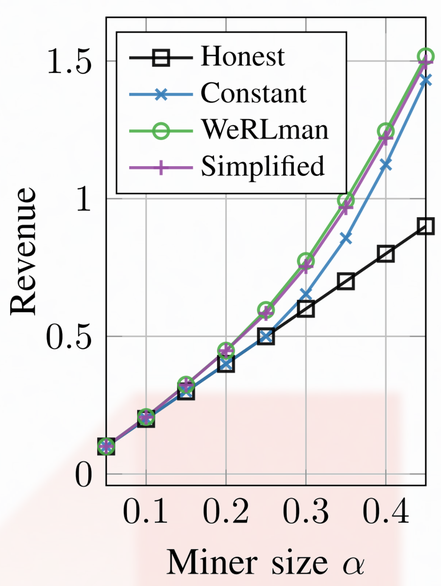
$$t'_n = t_n - \frac{1}{|B|}\sum_{n \in B} t_n$$

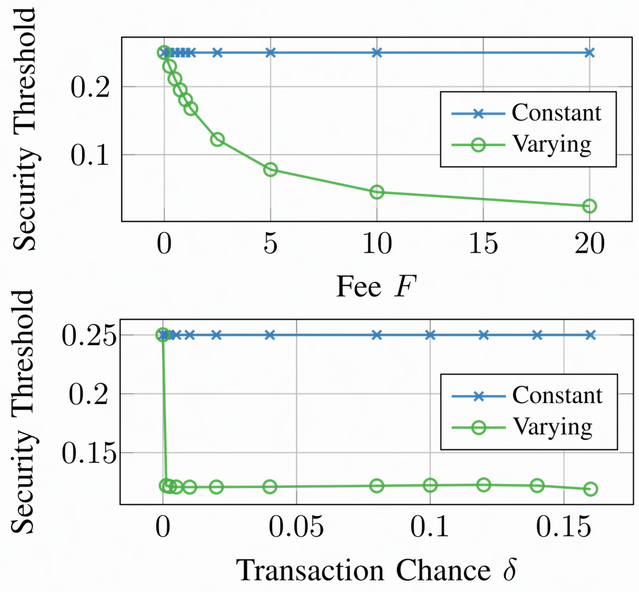




Results

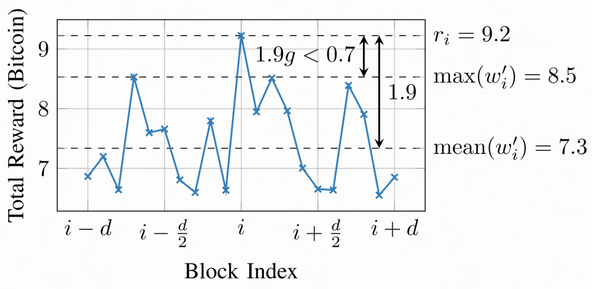
Use WeRLman to maximize revenue for different miner sizes to upper bound the security threshold.

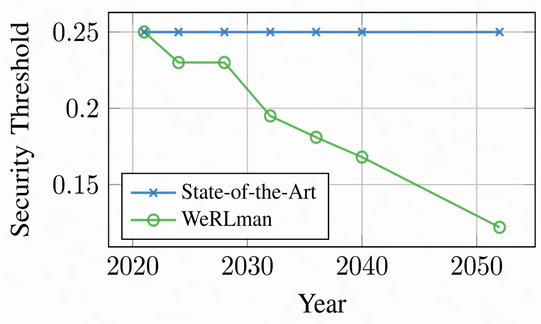




Increasing reward variability lowers the threshold!

Gather data from the Bitcoin network to estimate the security threshold in the future.





We find that Bitcoin is more vulnerable than previously considered. Its security will deteriorate as the block reward gets smaller.

This research was partly supported by the Israel Science Foundation (1641/18, 759/19), an IC3 grant, the Technion Hiroshi Fujiwara Cyber Security Research Center and the Israel National Cyber Directorate.