

# Staking Pools on Blockchains\*

Hans Gersbach	Akaki Mamageishvili	Manvir Schneider
CER-ETH and CEPR	CER-ETH	CER-ETH
Zürichbergstrasse 18	Zürichbergstrasse 18	Zürichbergstrasse 18
8092 Zurich, Switzerland	8092 Zurich, Switzerland	8092 Zurich, Switzerland
hgersbach@ethz.ch	amamageishvili@ethz.ch	manvirschneider@ethz.ch

This Version: April 2022

## Abstract

Several proof-of-stake blockchains allow for “staking pools”, i.e. agents interested in validating transactions can open a pool to which others can delegate their stake. We develop a game-theoretic model of staking pool formation in the presence of malicious agents who want to disrupt the blockchain. We establish existence and uniqueness of equilibria. Moreover, we identify the potential and risk of staking pools. First, staking pools can never increase current blockchain security over a system in which such pools are not allowed. Yet, by optimally selecting the distribution of the validation returns, honest stake holders obtain higher returns, which may be beneficial for future blockchain operations. Second, by choosing welfare optimal distribution rewards, staking pools prevent from allocating large rewards to malicious agents. Third, when pool owners can freely distribute the returns from validation to delegators, staking pools disrupt blockchain operations, since malicious agents attract all delegators by distributing most of the returns to them.

**Keywords:** Delegation, Staking Pool, Blockchain, Governance

**JEL Classification:** C72, D02, D60, G23

---

\*We would like to thank Philémon Bordereau for helpful comments.

# 1 Introduction

From both the financial side and the security side, there are reasons why a proof-of-stake blockchain may want to allow the formation of staking pools. With staking pools, agents interested in validating transactions are allowed to open a pool, such that others can delegate their stake for some time to it. For delegating agents who are not interested in validating transactions, this can provide an additional income on their token holdings. In turn, their stakes are blocked and cannot be used for other purposes during the time of commitment. By agents we mean all participants of the decentralized system who own at least some stake, which usually takes the form of a native token. Agents interested in running a staking pool could earn a higher income from their transaction validation activities.<sup>1</sup>

Ideally, such a staking system makes it more attractive to hold tokens, provides incentives for a sufficient number of agents to run staking pools and act as transaction validators, and increases the share of honest agents, weighted by the stakes they hold, involved in transaction validation. As every staking pool is a validator, we occasionally use the word “validator” for a staking pool.

However, malicious agents also run staking pools and may thus enlarge the share of the stake they control in transaction validation.<sup>2</sup> This may undermine the security of the blockchain and lead to a collapse of this blockchain, as the malicious agents take over.

We explore how such a system can be modeled and designed so that it operates beneficially for the decentralized consensus mechanism—i.e. by lowering the share of malicious agents who want to disrupt the validation of transactions—and for the ecosystem as a whole. The model invokes a measure of honest agents who are interested in the returns from holding a stake (of tokens) in a proof-of-stake blockchain and thus are also interested in the proper functioning of the blockchain. An agent is honest if he is prepared to run the software for validation as required by the system. Otherwise, honest agents choose actions to maximize their expected returns. Honest agents have different costs in participating in transaction validation, as availability of appropriate computer software, the knowledge to run a secure validation node and opportunity costs to engage in validation activities may differ across agents. There is also a measure of malicious agents who are only interested in disrupting the blockchain.

A *Blockchain Designer* aims at maximizing the chance that the blockchain is working (maximizing blockchain security), which will be captured by maximizing the number of honest validators. We will also consider an alternative objective where the Blockchain

---

<sup>1</sup>An agent that decides to run a pool and validate transactions is usually referred as a node of a network.

<sup>2</sup>Our assumption that malicious agents always run staking pools is justified, since if they do not, they do not affect blockchain functioning and thus cannot be called malicious. If they delegate their stakes, they act like honest delegators and thus again are not malicious in any way.

Designer trades off the probability that the blockchain is running correctly with the aggregate costs for all honest agents of validating transactions. This is a standard economic welfare criterion. Two further aspects can be important for a Blockchain Designer: Reducing the rewards for malicious agents, as this decreases their future influence and distributing the rewards to validators as equally as possible.

The Blockchain Designer has two basic options when designing the market for staking pools. First, he can fix the return distribution between the pool owner and the pool delegators. We call this “return fixing”. Second, he can allow competition of pool runners regarding how the returns from transaction validation are shared between the pool owner and the pool delegators. This is called “return competition”.

We model the ensuing interaction as a three-stage game. In the first stage, agents decide whether (i) to open a staking pool, (ii) to delegate their stake to some pool or (iii) to abstain from validation activities. Setting up pools for validating transactions is costly, and these costs may differ between agents. In the second stage, either the Blockchain Designer determines the shares uniformly for all running pools (return fixing) or pool owners determine how returns should be shared between pool owners and delegators (return competition). In the third stage, transactions are validated and, depending on the share of stakes controlled by malicious agents, validation either works properly or the blockchain is disrupted.

Our main insights start from the observation that honest agents with high costs to set-up a node as a validator may want to delegate their stake to other pool owners, while honest agents with low costs may want to open their own pool. Malicious agents always open a pool, as this increases their chances to disrupt the blockchain.

We establish existence and uniqueness of equilibria of the stake pool formation game with fixed return distribution between pool owners and delegators and show that they are of the threshold type. We show next that there exists a unique sharing rule of the returns from validation between delegators and pool owners that maximizes the probability that the blockchain operates correctly and we do the same for maximizing welfare of honest agents when costs of running staking pools are taken into account in addition to blockchain security. Then, we provide numerical illustrations of the equilibria for uniform distribution of costs.

Subsequently, we identify the potential and risk of staking pools. First, staking pools can never increase current blockchain security over a system in which no such pools are allowed. The reason is as follows. Without staking pools, a share of honest agents participates in validating transactions, as the additional reward is higher than the costs. With staking pools and if the returns are shared with the validators, the return to pool owners declines, as the total rewards from validating transactions is given. Hence, less honest agents are willing to open staking pools, so that malicious agents will control a larger share of stakes in validating transactions.

Yet, by optimally choosing the distribution of the validation returns to delegators, the allocation of rewards to honest stakes involved in validation increases, which may be beneficial for subsequent blockchain operations. We show how return splitting between pool owners and delegators has to be determined in order to minimize the rewards to malicious agents.

Second, by choosing welfare optimal distribution rewards, staking pools may decrease blockchain security, but it prevents allocating large rewards to only a fraction of agents. Third, when pool owners can freely distribute the returns from validation to delegators, staking pools decrease blockchain security, since malicious agents attract delegators by distributing most of the returns to them.

Finally, we show how our results can be extended to situations in which not only running a pool but also the act of delegation is costly. Moreover, as we use a continuum model for tractability reasons, we show how our analysis can be recasted—albeit with a more complex formal apparatus—in a discrete setting.

The paper is organized as follows. In the next section, we discuss related literature. In particular, literature motivating the formation of staking pools in proof-of-stake blockchains is reviewed. In Section 3, we introduce the model and preliminaries. In Section 4, we analyze the equilibria of the fixed return game. In Section 5, we discuss designs for staking rewards that either maximize security or maximize welfare. Section 6 addresses a uniform cost distribution and provides numerical examples. In Section 7, we analyze a return competition game, where pool owners individually decide on the reward sharing scheme. In Section 8, we study two extensions of our basic model. Section 9 concludes.

## 2 Related Literature

*Staking Pools:* Many blockchains have already implemented staking or will implement it in the near future. Such examples include, but are not limited to Cardano<sup>3</sup>, Solana (Yakovenko (2017)), Polkadot (Wood (2016)), Tezos (Goodman (2014)) and Concordium (Damgård et al. (2020)). All these allow staking pools in which agents who do not run their own staking pool will be able to delegate their stake to an existing pool and benefit from rewards. By delegation, agents are indirectly involved in block proposal and validation, via their stake.

Brünjes et al. (2020) study staking pools only among honest agents from a mechanism design perspective. Their reward schemes ensure that pool running agents have their costs covered and reporting these costs is the dominant strategy.

*Blockchain and Consensus Protocols:* There is a rapidly growing literature on blockchains

---

<sup>3</sup><https://www.cardano.org> (retrieved January 20, 2022).

and consensus protocols. The most common reference is Nakamoto (2008), a Bitcoin whitepaper, in which the consensus protocol proof-of-work is introduced. Many papers focusing on consensus protocols, that have been published recently, include John et al. (2020) and Benhaim et al. (2021). A widely cited paper on the game-theoretic analysis of blockchain is Biais et al. (2019). Benhaim et al. (2021) and Amoussou-Guenou et al. (2020) consider games in the presence of malicious agents. This aspect is also part of our model. In particular, we argue that the probability that the blockchain is well-functioning is increasing with the share of honest validators and achieves complete security (probability one) when the share of malicious agents falls below a certain threshold. More specific work on blockchain mining rewards from a game-theoretic angle are studied in Chen et al. (2019) and Kiayias et al. (2016).

*Model Assumptions:* Herrera et al. (2014) study a voting game with two parties where voters vote for one party or abstain. In their model, voters have individual costs that are drawn according to some distribution function. Similarly to our paper, Herrera et al. (2014) characterize equilibria of their game which take the form of so called “cut-off thresholds”. They obtain a pair of thresholds, one for each party. Then, a citizen whose cost is below the corresponding threshold will turn out and vote for his party. If a citizen’s cost is above the threshold, he will abstain. Our model works in a similar manner. We characterize threshold equilibria such that if the cost for an individual is below that threshold, he will run a staking pool, and delegate or abstain otherwise.

Similar to Herrera et al. (2014), Castanheira (2003) studies a voting game where citizens’ costs come from a specific distribution, namely, costs are distributed uniformly random. In our paper, we also consider a uniform distribution for costs. One key difference, however, is that the costs in Herrera et al. (2014) and Castanheira (2003) are costs of voting<sup>4</sup>, whereas in our paper, the costs are associated with running a pool and the costs connected to it.

In our paper, we use continuum approach to model the measure of agents, and thus follow the approach in Gersbach (2009) and Halaburda et al. (2021), for instance. More concretely, agents are modeled as infinitely small. The continuum approximates large communities and it proves to be a tractable approach for the staking pool formation game. The continuum model is thus a limiting case where the number of agents becomes large and delegation to staking pools is done uniformly at random. In our model, besides a pool ID (or address), no further information is provided to delegators. From addresses or pool IDs, no information about pool owners can be inferred. Hence, every pool has equal chances to be chosen by agents.<sup>5</sup> In the basic version of the model, delegation

---

<sup>4</sup>See the rational and costly voting literature (Palfrey and Rosenthal (1983), Ledyard (1984), Börgers (2004) and Gersbach et al. (2021b)).

<sup>5</sup>In practice, delegators may have more information about pool owners, but they remain anonymous. On the Cardano blockchain, for example, agents can find all staking pools on `pool.pm`, which visualizes all staking pools with the pool IDs, the current stake of the pools, the number of delegators and which

entails no cost.

*Crypto-Democracy and Vote Delegation:* There is an extensive research on democracy and blockchains, such as voting on blockchain (see Leonardos et al. (2020), Osgood (2016) and Allen et al. (2017), for instance). Very recent work on vote delegation in the presence of malicious agents is Gersbach et al. (2021a). The delegation of votes can be seen as a delegation of stakes in the blockchain environment. Further literature in the field of vote delegation is known under “liquid democracy”, where besides voting and abstaining, agents have the additional option to delegate their vote to other agents (see Bloembergen et al. (2019), for example).

*Automated Market Maker (AMM):* Staking is also used in other contexts such as yield aggregation<sup>6</sup> and automated market maker. The analogy to staking pools is the following. There are liquidity providers who add funds to a liquidity pool that is managed by an AMM. The liquidity providers correspond to agents who delegate their stake in our model. Hence, the AMMs collect funds in a liquidity pool that can be traded later.<sup>7</sup> In return, liquidity providers receive tokens proportional to their staked amount, which correspond to the rewards in our model. Besides these analogies, AMMs have other tasks as well, such as determining prices for traders. A well-known AMM is Uniswap, for example (see Hayden Adams (2020)).

## 3 Model

### 3.1 The General Set-up

There is a continuum of measure  $H$  of *honest* agents and there is a continuum of measure  $M$  of *malicious* agents. The continua can be represented by intervals on the real line, with length  $H$  and  $M$ , respectively. We assume  $H > M$ , so that, honest agents are in the majority. Each agent (malicious or honest) has one unit of the stake. Hence, the aggregate amount of stakes is  $H + M$ .

Each honest agent is identified by his cost level for validation of transactions, respectively, for running a staking pool on the blockchain.<sup>8</sup> Costs are denoted by  $c$  and are heterogeneous across honest agents, as they depend on the availability of appropriate computer capital and human capital. Let the random variable  $X$  correspond to the costs for honest agents. Specifically, the costs for honest agents are distributed according to the atomless density function  $f(c)$  defined on  $[0, T)$ . Note that the support interval can

---

blocks were produced by which staking pool.

<sup>6</sup>Under yield aggregation, investors can passively earn rewards by sending tokens to reward generating smart contracts (see Cousaert et al. (2021)).

<sup>7</sup>Note that the liquidity pools hold at least two different assets, which is a major difference to staking pools.

<sup>8</sup>Costs include, for example, the costs for registering, running the software and forwarding messages on transactions.

be  $\mathbb{R}_+$ , that is,  $T$  can be equal to  $\infty$ . The corresponding cumulative distribution function is denoted by  $F(c)$ . Malicious agents are of the Byzantine type and do not care about the costs and returns of running a pool. Hence, we set their costs to zero.

There is also a reward  $R \in \mathbb{R}_+$ , paid for creating the next block<sup>9</sup>. The agents' types are private information. As assumed above, all honest and malicious agents have the same amount of stakes, equal to one unit. There is also a *Blockchain Designer*. The blockchain is assumed to be functioning better if more of validators are honest.

## 3.2 Objectives

The Blockchain Designer and the two types of agents have the following general objectives:

- Maximize the chance that the blockchain is working (maximizing blockchain security), which will be captured by a maximization of the number of honest validators (**Blockchain Designer**).
- Maximize expected reward minus cost (**Honest Agents**).
- Maximize the measure of stakes delegated to them (**Malicious Agents**).

We will formally specify the manifestation of these objectives later. As to the Blockchain Designer, we will consider an alternative objective where he trades-off the probability that the blockchain is running correctly with the aggregate costs for all honest agents to validate transactions. While we consider maximizing blockchain security as the most important objective, arguably one could also consider a standard economic welfare criterion as a guiding principle for the Blockchain Designer.

To further aspects can be important for a Blockchain Designer. First, he may aim at minimizing the rewards received by malicious agents, as this decreases their future influence. Second, the Blockchain Designer may want to distribute rewards to validators as equally as possible, which is an original motivation of staking. We will discuss to which extent these aspects materialize when we present our results.

## 3.3 Staking Pool Formation Game

We consider the following game, which consists of three stages:

Stage 1: Agents decide either to form a staking pool or not (both honest and malicious). Agents who decide to become a pool owner obtain an identification number, denoted by  $i$ .

---

<sup>9</sup>In some contexts it is called mining reward.

Stage 2: Agents who did not register for a staking pool decide whether to delegate their stake to some staking pool or to remain idle.

Stage 3: The blockchain runs, validation takes place (or not), and rewards are distributed.

If an agent  $i$  forms a staking pool, we denote by  $s_i$  the amount of stakes he is receiving. We also denote by  $P$  the measure of honest agents who form a staking pool.  $D$  and  $I$  denote the measure of honest agents who delegate their stakes or stay idle, respectively. We have  $H = P + D + I$ .

Our main assumption for this game is that delegators distribute themselves evenly across all possible pools. The rationale is that the type of a pool owner is private information, and for delegators, pool owners are all alike. Hence, invoking measure consistency, this assumption implies

$$s_i = \frac{D}{P + M}, \forall i.$$

We note that all pools obtain the same amount of delegated stakes and thus we write  $s$  for  $s_i$  in the following. The total size of the pool—stakes of the pool owner and the number of delegated stakes—is then  $s + 1$ . We note that the aggregate amount of stakes is  $(s + 1)(P + M) + I = D + P + M + I = M + H$ .

The payoffs are determined as follows: The next validator<sup>10</sup> is chosen among the available pools proportionally to the pool size. The total reward for the next block is given by an amount  $R$ . Since all pools have the same size, the return distribution is a uniform distribution with density  $\frac{1}{P+M}$ . Hence, the individual reward a pool expects to receive is  $r = \frac{R}{P+M}$ . We note that both the individual return for an individual and the cost of running a pool have zero weight in the aggregate amount of return  $R$  and the aggregate amount of costs, respectively, since we are working with a continuum to simplify the description. However, we show in 8.3, how the analysis—albeit in a much more complicated way—can be performed in a discrete framework and how we can recover results in that framework.

The blockchain designer sets a parameter, denoted by  $\lambda$ ,  $0 < \lambda \leq 1$ , which determines how rewards have to split between pool owners and delegators. To sum up, the individual expected rewards are as follows:

- A pool receives  $r = \frac{R}{P+M}$ .
- The pool owner obtains  $\lambda \cdot r$ .
- An individual delegator obtains  $\frac{(1-\lambda) \cdot r}{s}$ . The total amount given to the delegators in a pool is  $(1 - \lambda) \cdot r$ .

---

<sup>10</sup>In blockchains, transaction validation is done by creating a new block.

- An idle agent obtains 0.

We measure the probability that the blockchain operates correctly by a function

$$P_c(\cdot) : [0, T] \rightarrow [0, 1],$$

that depends on the share of honest agents running staking pools, that is (weakly) increasing as a function of this share, and which may reach probability one if a sufficient share of honest agents is participating in validating transactions. Later in this paper, we will study two different versions of this probability function and reward schemes that depend on it. In the basic version of the staking pool formation game, we assume that the returns are paid, no matter whether the blockchain operates correctly or not. The motivation for this assumption is as follows: Whether the blockchain operates correctly or not may not be immediately detected or agents maybe able to sell their rewards immediately after writing the next block. Hence, agents involved in validating transactions aim at maximizing the immediate returns from these activities in such cases. The Blockchain Designer is, of course, interested in how well the blockchain is functioning. In Section 8.2, however, we make rewards dependent on the operation of blockchains, that is, rewards are only distributed if the blockchain operates correctly.

The main design parameter  $\lambda \in [0, 1]$  is a non-negative real number set by the Blockchain Designer. A game with a payoff structure as above, together with  $(H, M, R, F, \lambda)$ , is called a “staking pool formation game” and is denoted by  $\mathcal{G}$ .

## 4 Equilibrium Analysis

In this section, we analyze the equilibria of a staking pool formation game.

### 4.1 Equilibrium Concept

In a staking game with non-zero reward  $R$ , an honest agent will never stay idle. The reason is that since delegation is free of cost and delegators earn some reward, the expected payoff is positive for a delegator, whereas the payoff for staying idle is zero. Only in the case  $\lambda = 1$ , agents would be indifferent between delegation and staying idle. Hence, delegation weakly dominates staying idle and so  $H = P + D$ .

In Section 8.1, we will consider the case where delegation comes at a small fixed cost and so, staying idle will not be dominated in general. For tractability we assume a tie-breaking rule. First, if an honest agent is indifferent between delegating and staying idle, he will choose to delegate. This implies that when there are no delegation costs, no agent remains idle.

We now proceed by focusing on equilibria of the threshold type. In particular, we solve for the threshold equilibrium by looking at the agent with a specific cost level  $c^*$  at which this agent is indifferent between running a pool and delegating, i.e., the expected utility from delegating is equal to the expected utility from running a pool.

An important remark is in order. If a threshold equilibrium exists, it is unique. If a threshold equilibrium does not exist, we end up in a corner solution—either no honest agent will run a pool or all honest agents will run a pool, and thus there will be no delegation.

We introduce the following definition:

**Definition 1 (Threshold Equilibrium)**

*A cost level  $c^* > 0$  is called a “threshold equilibrium” if an agent with cost  $c^*$  is indifferent between running a staking pool and delegating. Furthermore, all agents incurring a cost that is lower than  $c^*$  will run a staking pool, and all agents with a cost greater than  $c^*$  will delegate.*

In the threshold equilibrium we have  $P = F(c^*)H$  and  $D = (1 - F(c^*))H$ .

## 4.2 Equilibrium Characterization

Note that a strategy profile in which no honest agent runs his/her own pool is always an equilibrium of the game  $\mathcal{G}$ . In the following we are looking for the other equilibria, in which at least some fraction of honest agents decide to run own pools. We characterize the equilibria of the staking game:

**Theorem 1**

*There exists a unique threshold equilibrium to the game  $\mathcal{G}$  if and only if*

$$\lambda > \frac{M}{H + M} \text{ and } T > \frac{\lambda R}{H + M}. \tag{1}$$

**Proof.** First, we have to set up the indifference condition for an honest agent. That is, we have to equate the expected utility from being a delegator with the expected utility from running a pool. More precisely, the expected utility from being a delegator is

$$\frac{(1 - \lambda)r}{s},$$

which is the share  $(1 - \lambda)$  of the reward  $r$ , divided by the number of delegators, for the particular pool. Similarly, the expected utility from running a pool is

$$\lambda r - c,$$

which is the share  $\lambda$  of the reward  $r$ , minus the cost  $c$  of the particular pool owner. In the equilibrium point  $c^*$ , an honest agent is indifferent between delegating and running

a pool, that is,  $c^*$  solves the indifference equation:

$$\frac{(1-\lambda)r}{s} = \lambda r - c^*. \quad (2)$$

From the equilibrium definition, we know that  $P = F(c^*)H$  and  $D = (1 - F(c^*))H$ . Plugging in these values in (2) and simplifying, we obtain:

$$\frac{(1-\lambda)R}{(1-F(c^*))H} = \lambda \frac{R}{F(c^*)H + M} - c^*. \quad (3)$$

We reorder equation (3) as follows:

$$c^* = \frac{\lambda R}{F(c^*)H + M} - \frac{(1-\lambda)R}{(1-F(c^*))H}. \quad (4)$$

We note that the left hand side (LHS) of equation (4) is obviously increasing in  $c^*$ , while the right hand side (RHS) is decreasing in  $c^*$ . Indeed, the derivative of the RHS with respect to  $c^*$  is

$$-\frac{\lambda R F'(c^*)H}{(F(c^*)H + M)^2} - \frac{(1-\lambda)R F'(c^*)}{(1-F(c^*))^2 H} < 0.$$

Since the LHS of (4) is increasing and is equal 0 for  $c^* = 0$ , and the RHS is decreasing in  $c^*$ , the necessary condition to have a solution to the equation is that the RHS is positive for  $c^* = 0$ . That is, we have the condition

$$\frac{\lambda R}{F(c^*)H + M} - \frac{(1-\lambda)R}{(1-F(c^*))H} > 0,$$

which is, for  $c^* = 0$ , equivalent to the condition in the theorem,

$$\lambda > \frac{M}{H + M}.$$

The indifference condition of the equilibrium of equation (4) to have an internal solution is obtained by taking  $c^* = T$ . In this case, the LHS has to be larger than the RHS, which is equivalent to:

$$T > \frac{\lambda R}{H + M}.$$

Note that here we assume that the expected gains from delegation,  $\frac{(1-\lambda)R}{(1-F(c^*))H}$ , are equal to 0 if all honest agents are running their own pools. This is consistent with the continuum model, where each agent is assumed to have a size zero.

To establish uniqueness, suppose that  $\lambda > \frac{M}{H+M}$ . As shown above, if we focus on threshold equilibria, there exists a unique equilibrium characterized with the cost level  $c^*$ . Suppose that an equilibrium exists which is not of the threshold type. Without loss

of generality, assume two cost levels  $c_1$  and  $c_2$  with  $c_2 > c_1$ , with the following property. A agent with cost  $c_2$  will run a pool, while a agent with  $c_1$  will delegate. Hence, for the first agent, it must hold that

$$\frac{\lambda R}{F(c^*)H + M} - c_2 > \frac{(1 - \lambda)R}{(1 - F(c^*))H},$$

while for the second agent, we must have the opposite inequality, that is,

$$\frac{\lambda R}{F(c^*)H + M} - c_1 < \frac{(1 - \lambda)R}{(1 - F(c^*))H}.$$

Together, this implies that

$$c_2 < \frac{\lambda R}{F(c^*)H + M} - \frac{(1 - \lambda)R}{(1 - F(c^*))H} < c_1,$$

which contradicts the assumption  $c_2 > c_1$ . Hence, any equilibrium will be of threshold type. ■

Since all equilibria are of the threshold-type, we simply refer to threshold-type equilibria as “equilibria”.

The interpretation of the lower bound condition on  $\lambda$  in the theorem is straightforward. To have a positive measure of pools owned by honest agents, the share of the reward for the pool owners should be higher than the share of malicious agents in the whole system. As long as  $\lambda$  satisfies this condition, we have a unique equilibrium of the game  $\mathcal{G}$ . From the proof of Theorem 1, it is straightforward to see that if  $\lambda = \frac{M}{H+M}$ , then the only solution to the indifference condition is  $c^* = 0$  and hence, all honest agents will delegate and malicious agents control all stakes. If  $\lambda < \frac{M}{H+M}$ , then there exists no equilibrium solution, where a positive measure of honest agents run pools.

The second condition of the theorem is automatically satisfied if the support of the random variable  $X$  is infinite. On the other hand, if it is finite, then for the right end point of the support,  $T$ , the payoff coming from everyone running their own pool should be lower than the cost.

If the first condition of the theorem holds, but the second condition is violated, then in the equilibrium, everyone runs a staking pool. However, it is useful to look at polar cases. For instance, if the Blockchain Designer wants to encourage some level of delegation, i.e., not all honest agents should run their own pools in the equilibrium, the second condition of the theorem has to hold.

## 5 Optimal Reward Design

In this section, we use the framework developed in the previous sections to design blockchains that maximize security or, alternatively, maximize welfare.

### 5.1 Maximal Blockchain Security

We obtain the equilibrium solution of (3) for a given  $\lambda$  by simply solving for  $c^*$ . We denote it by  $c^*(\lambda)$ . The inverse function is denoted by  $\lambda(c^*)$ , and can be trivially found from (3). Namely,

$$\lambda(c^*) = \frac{c^* + \frac{R}{(1-F(c^*))H}}{\frac{R}{F(c^*)H+M} + \frac{R}{(1-F(c^*))H}}. \quad (5)$$

$\lambda$  is a designer's variable.

We assume in this section that the probability that the blockchain operates correctly is given by:

$$P_c(c^*) := \frac{P(c^*)}{P(c^*) + M}.$$

That is, the probability that the next block consists of correct transactions is equal to a share of honest pool runners. This is a simple formulation, reflecting that the next block writer is chosen uniformly at random. However, our analysis holds qualitatively for any probability function that is increasing in the share of honest agents and may reach 1 if a sufficient share of honest agents is achieved.

The first goal of the designer is to maximize the share of honest agents running pools, that is, to maximize  $P(c^*)$ . The probability that the blockchain is run correctly depends on  $P$ , which is increasing with increasing  $c^*$ . Increasing  $\lambda$  has two effects on the honest agents' decision. First, it motivates an agent to run a pool, as a greater share of the rewards is allocated to the owner of the pool. On the other hand, since a higher  $\lambda$  motivates many agents to run pools, there are many pools, and therefore, lower chances for each of them to win the reward. However, we obtain the following result:

**Proposition 1**

*The fraction of honest agents running a pool is maximized for  $\lambda = 1$ .*

**Proof.** Note that the RHS of (4) is increasing in  $\lambda$ . By increasing  $\lambda$ , we have to increase  $c^*$  to have equality, as the RHS is decreasing in  $c^*$ . That is, if  $\lambda_1 \leq \lambda_2$ , then  $c^*(\lambda_1) \leq c^*(\lambda_2)$ . Taking the maximum value  $\lambda = 1$  transforms equation (4) into

$$c^* = \frac{R}{F(c^*)H + M}. \quad (6)$$

The solution to this equation maximizes the share of honest validators. ■

We note that by setting  $\lambda = 1$ , in the threshold equilibrium of game  $\mathcal{G}$ , we do replicate the levels of honest and malicious stakes involved in transaction validation which would arise in the simple game without delegation and no staking pools. In this game, instead, honest agents are allowed to either validate transactions or abstain. In such a game, the indifference condition of the threshold equilibrium corresponds to  $c^* = r$ , equivalent to  $\lambda = 1$  in the pool formation game. Yet, with  $\lambda = 1$  and pool formation, all returns from validation are channelled to staking pool owners while delegators receive nothing. This is a concern for the future evolution of the blockchain since stake holding may be more and more concentrated on pool owners.

We will see next that the solution  $\lambda = 1$  does not maximize social welfare, and does not distribute the rewards on honest agents that have high costs of running a pool either.

## 5.2 Welfare Optimal Reward Schemes

In this section, we consider the alternative objective the Blockchain Designer may pursue, namely taking into account that achieving maximal security may involve large costs, as too many honest agents with high costs participate in the validation process. For the alternative objective, we normalize the returns from a successful operation of the blockchain per honest agent to one and express the costs relative to these returns. To quantify the gains and losses, we introduce social welfare of the game  $\mathcal{G}$ .

### Definition 2 (Social Welfare)

*Social welfare of the game  $\mathcal{G}$  is defined as*

$$W = \frac{P(c^*)}{P(c^*) + M} H - P(c^*) \mathbb{E}[X | X < c^*]. \quad (7)$$

The social welfare is calculated as the probability that the blockchain runs correctly, times the total number of honest agents, minus pool running costs incurred by honest agents.

Increasing  $\lambda$  has two competing effects on social welfare. First, it increases the number of honest agents who create their own pools. Therefore, the likelihood that malicious agents will write the next block is decreasing. Second, increasing the number of honest agents who create their own pools wastes a lot of costs of running pools. These two effects work against each other. In the following, we show that for a wide class of distribution functions, the social welfare optimum value is not polar. We obtain the following result:

### Theorem 2

*Let the cost distribution function satisfy  $F''(c) \leq 0$  for any  $c \in [0, T)$ . If  $HM > (H + M)^2 T$ , then the optimum value maximizing welfare is  $c^* = T$ . On the other hand, if  $HM < (H + M)^2 T$ , then there is a unique optimum value of  $c^*$  that maximizes the social welfare.*

**Proof.** The welfare (7) can be rewritten as

$$W = \frac{F(c^*)H^2}{F(c^*)H + M} - F(c^*)H\mathbb{E}[X|X < c^*] = \frac{F(c^*)H^2}{F(c^*)H + M} - F(c^*)H \frac{\int_0^{c^*} xf(x)d(x)}{F(c^*)}.$$

The derivative of  $W$  with respect to  $c$  is equal to:

$$W' = \frac{f(c)H^2M}{(F(c)H + M)^2} - H \cdot cf(c).$$

The second derivative is equal to:

$$W'' = \frac{f'(c)H^2M(F(c)H + M) - 2f(c)^2H^3M}{(F(c)H + M)^3} - Hf(c) - cHf'(c).$$

We note that given  $F''(c) < 0$ , the second derivative is always negative, that is, the welfare function is concave. It is easy to verify that  $W'(0) > 0$ . That is, if  $W'(T) < 0$ , which is equivalent to  $HM < (H + M)^2T$ , we have a unique optimum solution. If, on the other hand,  $W'(T) > 0$ , equivalent to  $HM > (H + M)^2T$ , then the optimum is achieved in the point  $c^* = T$ . ■

Note that the uniform distribution function satisfies the condition  $F''(c) \leq 0$ . In fact, all distribution functions of the type  $F(c) = c^\alpha$ , where  $\alpha \leq 1$  satisfy the condition of the theorem.

## 6 Uniform Cost Distribution

In this section, we analyze the equilibria that maximize blockchain security and calculate the welfare optimal values for  $c^*$  and  $\lambda$  for the case when costs are distributed uniformly on  $[0, 1]$ . Furthermore, we study for the uniform distribution how rewards for malicious agents can be minimized.

### 6.1 Security Maximization

To maximize the share of honest staking pools, we have to solve equation (6). Hence, for uniform distribution on the interval  $[0, 1]$ , that is,  $F(c^*) = c^*$ , we have to solve the following quadratic equation:

$$c^2H + cM - R = 0. \tag{8}$$

The positive solution is given by

$$c^* = \frac{-M + \sqrt{M^2 + 4HR}}{2H}. \quad (9)$$

If  $H = 1$  and  $M = 0.4$  and  $R = 1$ , then the solution to (8) is  $c^* = \frac{-1 + \sqrt{26}}{5} \approx 0.82$ . This solution maximizes the share of honest staking pools. That is, approximately 82% of honest agents will run pools in the equilibrium. More numerical values are provided in Section 6.3.

## 6.2 Welfare Maximization

For uniform distribution  $F$ , the welfare measure (7) simplifies to

$$W = H \left( \frac{c^*H}{c^*H + M} - \frac{1}{2}c^{*2} \right),$$

where we used that  $P(c^*) = F(c^*)H = c^*H$  and

$$E[X|X < c^*] = \int_0^{c^*} xf(x|x < c^*)dx = \frac{\int_0^{c^*} xf(x)dx}{F(c^*)} = \frac{\frac{1}{2}c^{*2}}{c^*} = \frac{c^*}{2}.$$

The derivative of  $W$  with respect to  $c^*$  is

$$W'(c^*) = \frac{H^2M}{(c^*H + M)^2} - c^*H.$$

Note that  $W'(c^*) = 0$  has three solutions. Furthermore,  $W''(c^*) = -\frac{2H^3M}{(c^*H + M)^3} - H < 0$  for any non-negative  $c^*$ . This means that the real-valued extremum (which is in the interval  $[0, 1]$ ) of  $W$  is a maximum. By simply solving  $W'(c^*) = 0$ , we obtain the following real-valued solution which maximizes  $W$ :

$$c^* = \frac{(1 + i\sqrt{3})M^2}{2^{2/3}3Z} + \frac{(1 - i\sqrt{3})Z}{6\sqrt[3]{2}H^2} - \frac{2M}{3H}, \quad (10)$$

where  $Z = \sqrt[3]{-27H^5M - 2H^3M^3 + \sqrt[3]{3}\sqrt{27H^{10}M^2 + 4H^8M^4}}$ . The welfare function is maximized at  $c^*$ . Numerical examples follow in the subsequent section.

## 6.3 Numerical Illustrations

In this subsection, we provide detailed security maximizing (see Section 6.1) and welfare maximizing (see Section 6.2) equilibrium values for three sets of parameters in Table 1 and Table 2. In Table 1 the values for  $c^*$  are given by Equation (9). In Table 2 the values for  $\lambda$  and  $c^*$  are given by Equations (5) and (10), respectively.

$H$	$M$	$R$	$\lambda$	$c^*$	$P(=c^*H)$	$W$
1	0.5	1	1	0.78	0.78	0.305
1	0.4	1	1	0.82	0.82	0.341
1	$\frac{1}{3}$	1	1	0.85	0.85	0.357

Table 1: Maximizing the measure of honest staking pools:  $P$ .

$H$	$M$	$R$	$\lambda$	$c^*$	$P(=c^*H)$	$W$
1	0.5	1	0.83	0.5	0.5	0.375
1	0.4	1	0.8	0.497	0.497	0.431
1	$\frac{1}{3}$	1	0.77	0.491	0.491	0.475

Table 2: Maximizing social welfare:  $W$ .

For example, consider the case where  $M = 0.4$ . Social welfare is maximized for cost  $c^* = 0.497$ . That is, it is quite different from  $c^* = 0.82$ , which is the value for maximizing the honest agents running a pool. On the other hand,  $\lambda = 0.8$ , the value optimizing welfare is also quite different from  $\lambda = 1$ , the value maximizing the share of honest agents running their own pools.

## 6.4 Malicious Reward Minimization

The share of rewards received by malicious agents as a function of  $\lambda$  is calculated in the following way:

$$\mu(\lambda) := \frac{M\lambda}{F(c^*)H + M}. \quad (11)$$

To show the main result of this section, we first show a lemma that holds for any cost distribution function:

### Lemma 1

*Rewards for malicious agents when  $\lambda = \frac{M}{H+M}$  are lower than rewards for  $\lambda = 1$ , for any cost distribution function  $F$ .*

**Proof.** In the first case, where  $\lambda = \frac{M}{H+M}$ , we have  $c^* = 0$ , and therefore, using (11), rewards are equal to

$$\mu\left(\frac{M}{H+M}\right) = \frac{M}{0+M}\lambda = \frac{M}{H+M}.$$

In the second case, rewards are equal to

$$\mu(1) = \frac{M}{F(c^*(1))H + M}\lambda = \frac{M}{F(c^*(1))H + M}.$$

Since  $F(c^*(1)) \leq 1$ , the lemma is proved. ■

Next, we show the main result of this section:

## Proposition 2

Rewards to malicious agents  $\mu(\lambda)$  are minimized for  $\lambda = \frac{M}{H+M}$  for a uniform cost distribution function on the interval  $[0, 1]$ . Moreover,  $\mu(\lambda)$  is increasing or first increasing and then decreasing on the interval  $[\frac{M}{M+H}, 1]$ .

**Proof.** We verify the derivative of the share of rewards received by malicious agents with respect to  $c^*$  for a uniform cost distribution. It is given by,

$$\frac{d \frac{M\lambda}{F(c^*)H+M}}{dc^*} = \frac{M\lambda'(c^*)(c^*H + M) - HM\lambda}{(c^*H + M)^2}. \quad (12)$$

Note that the denominator is always positive. In Equation (5),  $\lambda$  is given as a function of  $c^*$ . The derivative is

$$\lambda'(c^*) = \frac{1}{R(H + M)}(-3c^{*2}H^2 + 2c^*(H^2 - HM) + HR + MH).$$

After plugging this into Equation (12), the numerator of (12),  $M\lambda'(c^*)(c^*H + M) - HM\lambda$ , is given by

$$\frac{H^2c^{*3}(HM - M) + H^2c^{*2}(HM - 4M^2) + 2Hc^*(HM^2 - M^3) + HM^3}{R(H + M)}.$$

Again, note that the denominator is always positive and hence, we only consider the numerator,  $H^2c^{*3}(HM - M) + H^2c^{*2}(HM - 4M^2) + 2Hc^*(HM^2 - M^3) + HM^3$ . For  $c^* = 0$  (equivalent to  $\lambda = M/(H + M)$ ), it is positive, and for  $c^* = \frac{-M + \sqrt{M^2 + 4HR}}{2H}$  (equivalent to  $\lambda = 1$ ), it is, depending on  $H, M$  and  $R$ , either positive or negative.

As increasing (decreasing)  $\lambda$  yields increasing (decreasing)  $c^*$  and vice versa, (12) translates easily. Therefore, given  $\mu(\frac{M}{H+M}) < \mu(1)$  from Lemma 1, the minimum value is achieved in the point  $\lambda = \frac{M}{H+M}$ . From the observation  $\mu(\frac{M}{H+M}) < \mu(1)$ , we see that the derivative cannot always be negative. Therefore, it is either always positive, implying that  $\mu(\lambda)$  is increasing on the whole interval  $[\frac{M}{H+M}, 1]$ , or the derivative is first positive and then negative, implying that  $\mu(\lambda)$  is first increasing and then decreasing. ■

This result suggests that rewards to malicious agents are minimized in the corner case, where only malicious parties run pools. In that case, however, the probability that the blockchain functions correctly is equal to 0. Therefore, from the blockchain security perspective, designer needs a higher  $\lambda$ . Our result suggests that once  $\lambda$  is large enough ( $\lambda \geq t_\lambda$ ) for the blockchain security to cross the required threshold specified by a system, the designer only needs to verify endpoints of the interval  $[t_\lambda, 1]$ , for minimizing rewards to malicious agents.

We next provide numerical examples for the share of rewards that malicious agents receive for a uniform cost distribution when  $\lambda = 1$  and  $\lambda < 1$ .

First, for  $\lambda = 1$ , the malicious agents' reward share is  $\frac{M}{P+M}$ , with  $P = F(c^*)H$ , where

$c^*$  is given by (9). For  $\lambda = 1, H = 1, R = 1$  the share of rewards that malicious agents receive is given by

$$\frac{2M}{M + \sqrt{M^2 + 4}}.$$

We summarize some values in Table 3.

$H$	$M$	$R$	Share of Reward
1	0.5	1	0.390388
1	0.4	1	0.327922
1	$\frac{1}{3}$	1	0.282376

Table 3: Share of rewards for malicious agents when  $\lambda = 1$ .

Second, for  $\lambda < 1$ , the malicious reward share is  $\frac{M}{P_\lambda + M}\lambda$ , with  $P_\lambda = F(c^*)H$ , where  $c^*$  is the real-valued non-negative solution to (3). We first solve the indifference equation as follows: With uniform distribution, we obtain from (3),

$$0 = H^2c^{*3} + (MH - H^2)c^{*2} - (RH + MH)c^* + \lambda R(H + M) - RM. \quad (13)$$

To find the roots of this cubic equation, we can write (13) as  $0 = ac^{*3} + bc^{*2} + cc^* + d$ , where

$$\begin{aligned} a &= -H^2, \\ b &= H(H - M), \\ c &= H(M + R), \\ d &= -\lambda R(H + M) + MR. \end{aligned}$$

Let

$$\begin{aligned} \Delta_0 &= b^2 - 3ac \\ &= H^2((H - M)^2 + 3H(M + R)), \\ \Delta_1 &= 2b^3 - 9abc + 27a^2d \\ &= H^3(2(H - M)^3 + 9H(H - M)(M + R) + 27HMR) - \lambda \cdot 27H^4R(H + M). \end{aligned}$$

Then, we define

$$C = \sqrt[3]{\frac{\Delta_1 + \sqrt{\Delta_1^2 - 4\Delta_0^3}}{2}}, \quad (14)$$

and the three roots of (13) are given by

$$c_k^* = -\frac{1}{3a} \left( b + z^k C + \frac{\Delta_0}{z^k C} \right), \quad (15)$$

where  $z = \frac{-1+i\sqrt{3}}{2}$  and  $k = 0, 1, 2$ .

We summarize some values in Table 4.

$H$	$M$	$R$	$\lambda$	Share of Reward
1	0.5	1	0.99	0.395647
1	0.5	1	0.9	0.414172
1	0.5	1	0.8	0.416164
1	0.5	1	0.7	0.409608
1	0.5	1	0.6	0.396822
1	0.5	1	0.5	0.378318
1	0.4	1	0.9	0.35305
1	0.4	1	0.8	0.357138
1	0.4	1	0.6	0.345172
1	0.4	1	0.5	0.331877
1	0.4	1	0.4	0.313697
1	$\frac{1}{3}$	1	0.9	0.307422
1	$\frac{1}{3}$	1	0.5	0.294599
1	$\frac{1}{3}$	1	0.4	0.28047
1	$\frac{1}{3}$	1	0.3	0.261626

Table 4: Share of reward that malicious agents receive.

## 7 Return Competition

In this section, we reconsider the staking pool formation game. Instead of the Blockchain Designer, we allow that pool runners choose both their own levels of rewards and the rewards they want to distribute to delegators. A pool owner  $i$  sets his own  $\lambda_i$ . This game is a variant of the game  $\mathcal{G}$  studied so far in the paper.

In the first part of this section, we allow free return competition, that is, pool owner  $i$  can choose any  $\lambda_i \in [0, 1]$ . We denote this game by  $\mathcal{G}_0$ . Hence, the game unfolds as follows:

Stage 1: Agents (both honest and malicious) either decide to form a staking pool or not. Agents who decide to become a pool owner obtain an identification number, denoted by  $i$ , and set  $\lambda_i$ .

Stage 2: Agents who did not register for a staking pool decide whether to delegate their stake to some staking pool or to remain idle.

Stage 3: The blockchain runs, validation takes place (or not), and rewards are distributed.

We obtain the following result:

**Proposition 3**

*In any equilibrium of the game  $\mathcal{G}_0$ , malicious agents control all stakes involved in transaction validation and the blockchain is disrupted.*

**Proof.** Suppose that there exists an equilibrium in which honest agents run staking pools. Since running such staking pools is costly and there is zero measure of honest agents with zero costs, such an equilibrium necessarily must involve that the minimal value of all offered values  $\lambda_i$  by staking pool owners, denoted by  $\hat{\lambda}$ , must be positive. Otherwise, honest agents are better off by delegating their stakes. Note that the last statement holds, since an individual honest agent has no influence on the probability that the blockchain operates correctly by his/her decision whether to run a staking pool or to delegate,

However, every malicious agent has an incentive to deviate and to set a lower value of  $\hat{\lambda}$  for his/her own staking pool in order to attract more delegators, thereby making staking pools for honest agents unattractive. Hence, all honest agents delegate. This is a contradiction that honest agents run staking pools. Hence, in any equilibrium, malicious agents control all stakes involved in transaction validation and the blockchain is disrupted. ■

In the second part of this section, agents are only allowed to choose their corresponding  $\lambda_i$  from the interval  $[\bar{\lambda}, 1]$ , where  $\bar{\lambda} > \frac{M}{H+M}$ . We denote this game by  $\mathcal{G}_{\bar{\lambda}}$ . Invoking standard Bertrand competition logic, we obtain the following result in this case:

**Proposition 4**

*The equilibrium of the extended game  $\mathcal{G}_{\bar{\lambda}}$  is the same as the equilibrium of game  $\mathcal{G}$ .*

**Proof.** First, we note that in the equilibrium, all malicious agents choose the lowest possible level  $\bar{\lambda}$ . If honest agents choose any  $\lambda$  that is strictly larger than  $\bar{\lambda}$ , then nothing is delegated to them. Therefore, they also choose the same  $\bar{\lambda}$ . ■

That is, imposing a lower bound on  $\lambda$  also guarantees the same upper bound. This adds to the robustness of the result obtained in Theorem 1 and offers a way to implement the equilibrium solution. On a practical side, the blockchain system does not need to force the agents to have the same level of rewards. Rather, they reach it through rational play. Note that setting any lower bound  $\bar{\lambda} \leq \frac{M}{H+M}$  would result in the same (bad) equilibrium obtained in Proposition 3.

## 8 Extensions

In this section, we study two extensions of our basic model. In first extension, we introduce a cost of delegation and in the second extension, we study the case when rewards are realized endogenously. Finally, we outline how the analysis has to be performed in a

discrete framework and how we can recover results as the one in Theorem 1 in a discrete framework.

## 8.1 Costly Delegation

In this subsection, we include costs of delegation. Arguably such costs are non-zero, since agents have to obtain the knowledge how to delegate safely their stakes to staking pools, which lock-up conditions are attached to such operations and how to observe the returns from delegation in comparison to other alternatives. This takes time and involves opportunity costs. We denote by  $c_d > 0$  the costs of delegation.

Costs for delegation introduce new trade-offs, as large values of  $\lambda$  may motivate honest agents to run pools but simultaneously discourage honest agents to delegate to pools, as they may simply stay idle. We will see that staking pools and suitably chosen sharing parameters for  $\lambda$  may increase blockchain security in such cases.

To prepare the analysis, let us first assume that the strategy set only consists of running a pool and delegating. We denote this game by  $\mathcal{G}_d$ . The indifference condition in this game is the same as (4), with only one difference—the RHS has the additional summand  $c_d$ .

$$c^* = \frac{\lambda R}{F(c^*)H + M} - \frac{(1 - \lambda)R}{(1 - F(c^*))H} + c_d. \quad (16)$$

We obtain the following auxiliary result:

### Proposition 5

*For any value of  $\lambda > \frac{MR - c_d MH}{HR + MR}$ , the indifference cost level of the equilibrium solution of the extended game  $\mathcal{G}_d$  is higher than the cost level of the equilibrium solution of  $\mathcal{G}$ .*

**Proof.** The proof is analogous to the proof of Theorem 1. The additional term  $c_d$  increases the RHS of (16), and therefore, the intersection of the LHS and RHS curves corresponds to higher  $c^*$ . By plugging in  $c^* = 0$ , and requiring LHS is lower than RHS, and simplifying we get:

$$\lambda > \frac{MR - c_d MH}{HR + MR}.$$

■

Note that the lower bound on  $\lambda$  for a positive equilibrium in the game  $\mathcal{G}_d$  is lower than the lower bound obtained in Theorem 1. However, in this game, it might be that the expected utility of running a pool minus the expected utility of delegation is negative, but adding  $c_d$  makes the RHS of (16) positive. In that case, individual rationality of pool running agents is violated. Therefore, if we consider an extended game  $\mathcal{G}_d^i$ , in which honest agents are allowed to be idle, then those with high costs would choose to stay idle instead of delegating. This also changes behavior of agents that run pools, in particular,

they stay idle. That is, the strategy sets of agents are extended, and also the definition of the threshold equilibrium takes a new strategy into account. Namely, in the indifference condition only positive values are compared, as the strategy to stay idle gives zero utility. Formally, we obtain the following result:

**Proposition 6**

For  $\lambda \in [\frac{MR - c_d MH}{HR + MR}, \frac{M}{M + H}]$ , in the equilibrium solution of the game  $\mathcal{G}_d^i$ , no honest agent runs a pool.

**Proof.** When  $\lambda \in [\frac{MR - c_d MH}{HR + MR}, \frac{M}{M + H}]$ , the expected gain from running a pool minus expected gain from delegation is negative:  $\frac{\lambda R}{F(c^*)H + M} - \frac{(1 - \lambda)R}{(1 - F(c^*))H} < 0$ , which violates individual rationality constraint for agents of the game  $\mathcal{G}_d$  that run pools. That is, a weakly dominant strategy for them is to stay idle, in the game  $\mathcal{G}_d^i$ . ■

Next, we ask if costly delegation can help to increase the level of honest agents running a pool. Note that  $\lambda = 1$  maximizes the level without costly delegation. From equation (6), we find the level of  $c^*$  such that agents with cost lower than this threshold will run a staking pool and those agents with costs higher than the threshold will stay idle, since delegation is costly. However, if we decrease  $\lambda$  in such a way that the expected return from delegation compensates the delegation cost  $c_d$ , then all honest agents that were staying idle will delegate or run a staking pool.

We obtain the following result:

**Proposition 7**

Costly delegation does not increase the highest blockchain security level of the benchmark game  $\mathcal{G}$ , for any distribution function  $F(c)$ .

**Proof.** We first look at the decision of agents whether to delegate or to stay idle. For values of  $\lambda$  that satisfy the following inequality:

$$\frac{(1 - \lambda(c^*))R}{(1 - F(c^*))H} \geq c_d, \tag{17}$$

honest agents delegate instead of staying idle.  $\lambda_d(c^*)$  is calculated from (16), that is, from the indifference condition of the game  $\mathcal{G}_d$ . That is,  $c^*$  should satisfy both conditions (16) and (17).

Recall that in the benchmark case  $\lambda = 1$  and  $c_d = 0$ , the equilibrium cost level  $c^*$  is a solution to the following equation:

$$c = \frac{R}{F(c)H + M}.$$

In the costly delegation game, however,  $c^*$  solves the equation (16), which together with (17) gives:

$$c^* = \frac{\lambda R}{F(c^*)H + M} - \frac{(1 - \lambda)R}{(1 - F(c^*))H} + c_d \leq \frac{\lambda R}{F(c^*)H + M} \leq \frac{R}{F(c^*)H + M}.$$

Given monotonic decreasing property of  $\frac{\lambda R}{F(c^*)H + M}$ , we get that the equilibrium  $c^*$  of the costly delegation game is always lower than the equilibrium level of game  $\mathcal{G}$ . ■

## 8.2 Endogenous Rewards

Throughout the paper, we assumed that rewards for writing the next block are exogenously given and are equal to a constant number  $R$ , no matter what fraction of honest agents runs pools and participates in the validation. In this section, we assume that rewards are realized only if the blockchain functions correctly, with probability one. This probability is calculated by the following formula:

$$P_c(c^*) := \min \left[ 1, \left( \frac{P}{P + M} + \theta \right) \right].$$

Here  $\theta$  is a real number in  $[0, \frac{1}{2}]$  that describes the tolerance of a system regarding the share of malicious agents it can handle without compromising network security.

We say that full network security is achieved when

$$\frac{P}{P + M} \geq 1 - \theta.$$

Typically, for example in Byzantine-fault-tolerant protocols,  $\theta$  is about  $\frac{1}{3}$  in many consensus protocols (see Lamport et al. (1982), Abd-El-Malek et al. (2005), David et al. (2017) and Dinsdale-Young et al. (2019)). Thus, in that case, if the fraction of staking pools run by honest agents is at least  $\frac{2}{3}$ , then full security is achieved.

Requiring the probability of the blockchain security to surpass the threshold  $1 - \theta$  imposes a threshold on the cost of pool running in the equilibrium. We denote the corresponding game by  $\mathcal{G}^\theta$  and the cost threshold by  $c^\theta$ . It is defined by the following equation:

$$c^\theta := \inf_c \frac{F(c)H}{F(c)H + M} \geq 1 - \theta. \quad (18)$$

In this setting, to have a unique threshold equilibrium, we need

$$c^\theta < \frac{\lambda R}{F(c^\theta)H + M} - \frac{(1 - \lambda)R}{(1 - F(c^\theta))H}. \quad (19)$$

We obtain a result similar to the one of Theorem 1:

**Theorem 3**

There exists a unique threshold equilibrium  $c^* > c^\theta$  to the game  $\mathcal{G}^e$  if and only if

$$\lambda > \frac{\frac{c^\theta}{R}(F(c^\theta)H + M)(1 - F(c^\theta))H + F(c^\theta)H + M}{H + M} \text{ and } T > \frac{\lambda R}{H + M}. \quad (20)$$

The proof is analogous to the proof of Theorem 1.

The results from the Section 7 on return competition are also translated directly in this setting. In particular, if the agents are allowed to set any  $\lambda \in [0, 1]$  as their own pool return, the blockchain security fails. We denote this game by  $\mathcal{G}_0^e$ . Formally, we obtain the following result:

**Proposition 8**

In any equilibrium of the game  $\mathcal{G}_0^e$  no honest agent runs a pool and the blockchain is disrupted.

The proof is analogous of the proof of Proposition 3 and exploits the fact that individual honest agents have zero measure. Therefore, any unilateral deviation by an agent does not affect the probability that the blockchain operates correctly. outcome.

### 8.3 Discrete Case

In this subsection, we show how the analysis can be recasted in the discrete framework—albeit in a much more complicated form. In particular, we show that the same lower bound condition on  $\lambda$  for having positive threshold equilibrium as in Theorem 1 can be obtained if we assume that the number of honest agents is a large integer  $n > 0$ , and replace the expected number of honest agents who run pools with  $nF(c^*)$ . The latter approximately holds by Chernoff concentration bounds. This observation adds to the robustness of the continuum approach, as it shows that the result obtained in Theorem 1 is not a byproduct of our assumption on having infinitely small agents. Rather, this assumption allows us to derive clean results more easily.

Let  $m \in \mathbb{N}$  denote the number of malicious agents. By the assumption on  $c^*$ , we have  $k + m$  pools, where  $k$  is distributed as binomial random variable with parameters  $n$  and  $F(c^*)$ . In expectation, every pool will obtain  $\frac{n-k}{k+m}$  delegated stakes. Hence, every pool has a total expected stake of  $1 + \frac{n-k}{k+m} = \frac{n+m}{k+m}$ . If a pool is chosen (i.e. it becomes the new proposer), then all its delegators receive a reward. In expectation, the probability that a pool is chosen is  $\frac{1}{k+m}$  and the reward for all delegators who delegated to this pool is  $(1 - \lambda)r$ .

The pool owners always face private cost  $c^*$  and obtain the reward  $\lambda r$  with probability  $\frac{1}{k+m}$ . The expected utility of the single delegator is  $\frac{1}{k+m}(1 - \lambda)\frac{r}{d}$ , where  $d$  is the number of delegators to that particular pool.

Let  $d_1, \dots, d_{k+m}$  be non-negative numbers, such that  $\sum_i d_i = n - k$ , where  $d_i$  denotes the number of delegators to pool  $i$ . The ex-post utility of a delegator of a baking pool  $i$ , when pool  $i$  has  $d_i$  delegators:

$$\frac{1 + d_i}{n + m} (1 - \lambda) \frac{r}{d_i},$$

and for the pool owner,

$$\frac{1 + d_i}{n + m} \lambda r - c_i.$$

We model  $d_i$  as a binomial random variable,  $d_i \sim \text{Bin}(n - k, \frac{1}{k+m})$ .

Next, we replace random variables with their means in the ex-post indifference equation, i.e., instead of  $d_i$ , we insert  $E[d_i] = \frac{n-k}{k+m}$  and instead of  $k$ , we insert  $E[k] = F(c^*)n$ . Then, we obtain:

$$\begin{aligned} \frac{1 + d_i}{n + m} (1 - \lambda) \frac{r}{d_i} &= \frac{1 + d_i}{n + m} \lambda r - c^* \\ \frac{1 + \frac{n-F(c^*)n}{F(c^*)n+m}}{n + m} (1 - \lambda) \frac{r}{\frac{n-F(c^*)n}{F(c^*)n+m}} &= \frac{1 + \frac{n-F(c^*)n}{F(c^*)n+m}}{n + m} \lambda r - c^* \\ c^* &= \frac{1}{F(c^*)n + m} \lambda r - \frac{1}{n - F(c^*)n} (1 - \lambda)r. \end{aligned}$$

The left hand side is equal to 0 for  $c^* = 0$  and it is increasing in  $c^*$ . The right hand side is a decreasing function in  $c^*$  and therefore, it should be non-negative for  $c^* = 0$ , to have a positive solution in  $c^*$ . That is,

$$\frac{1}{F(c^*)n + m} \lambda r \geq \frac{1}{n - F(c^*)n} (1 - \lambda)r.$$

For  $c^* = 0$ , we take into account that  $F(0) = 0$  for a distribution function  $F$ . Thus, we obtain:

$$\begin{aligned} \frac{1}{m} \lambda &\geq \frac{1}{n} (1 - \lambda) \\ \lambda &\geq \frac{m}{n + m}. \end{aligned}$$

## 9 Conclusion

In this paper, we initiated the study of the formation of staking pools from game-theoretic and mechanism design perspectives. Our insights can help to design reward distribution rules that improve the blockchain security and fairness of reward distribution. This study might open many further research avenues on how staking pools can be designed optimally for blockchains. For instance, one might introduce quantity constraints such as a leverage constraint on staking pools which limits the share of delegators towards the

pool owner. Whether such a constraint further improves the security of the blockchain is left for future research. One might also allow the history of staking and running staking pools to play a role in dynamic versions of the game and thus, an agent’s reputation to behave honestly may be taken into account in such staking pool formation games.

## References

- Abd-El-Malek, M., Ganger, G. R., Goodson, G. R., Reiter, M. K., and Wylie, J. J. (2005). Fault-scalable Byzantine fault-tolerant services. *ACM SIGOPS Operating Systems Review*, 39(5):59–74.
- Allen, D. W. E., Berg, C., Lane, A. M., and Potts, J. (2017). The economics of crypto-democracy. In *LINKDEM@IJCAI*.
- Amoussou-Guenou, Y., Biais, B., Potop-Butucaru, M., and Tucci-Piergiovanni, S. (2020). Rational vs Byzantine players in consensus-based blockchains. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS ’20*, page 43–51, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.
- Benhaim, A., Falk, B. H., and Tsoukalas, G. (2021). Scaling blockchains: Can elected committees help? *DecisionSciRN: Other Decision-Making in Computational Design & Technology (Topic)*.
- Biais, B., Bisière, C., Bouvard, M., and Casamatta, C. (2019). The blockchain folk theorem. *Review of Financial Studies*, 32(5):1662–1715.
- Bloembergen, D., Grossi, D., and Lackner, M. (2019). On rational delegations in liquid democracy. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(1):1796–1803.
- Börger, T. (2004). Costly voting. *American Economic Review*, 94(1):57–66.
- Brünjes, L., Kiayias, A., Koutsoupias, E., and Stouka, A.-P. (2020). Reward sharing schemes for stake pools. In *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 256–275.
- Castanheira, M. (2003). Victory margins and the paradox of voting. *European Journal of Political Economy*, 19(4):817–841.
- Chen, X., Papadimitriou, C., and Roughgarden, T. (2019). An axiomatic approach to block rewards. In *Proceedings of the 1st ACM Conference on Advances in Financial*

- Technologies*, AFT '19, page 124–131, New York, NY. Association for Computing Machinery.
- Cousaert, S., Xu, J., and Matsui, T. (2021). SoK: Yield aggregators in DeFi. <https://arxiv.org/pdf/2105.13891.pdf>.
- Damgård, I., Gersbach, H., Maurer, U., Nielsen, J. B., Orlandi, C., and Pedersen, T. P. (2020). Concordium white paper. [https://assets-global.website-files.com/6172c154c6a83e2a4713d1af/61cc25e0b3273f28f6a75dba\\_Concordium-White-Paper-June2021.pdf](https://assets-global.website-files.com/6172c154c6a83e2a4713d1af/61cc25e0b3273f28f6a75dba_Concordium-White-Paper-June2021.pdf).
- David, B., Gaži, P., Kiayias, A., and Russell, A. (2017). Ouroboros praos: An adaptively-secure, semi-synchronous proof-of-stake protocol. Cryptology ePrint Archive, Report 2017/573. <https://ia.cr/2017/573>.
- Dinsdale-Young, T., Magri, B., Matt, C., Nielsen, J. B., and Tschudi, D. (2019). Afgjort: A partially synchronous finality layer for blockchains. Cryptology ePrint Archive, Report 2019/504. <https://ia.cr/2019/504>.
- Gersbach, H. (2009). Democratic mechanisms. *Journal of the European Economic Association*, 7(6):1436–1469.
- Gersbach, H., Mamageishvili, A., and Schneider, M. (2021a). Vote delegation and misbehavior. In Caragiannis, I. and Hansen, K. A., editors, *Algorithmic Game Theory - 14th International Symposium, SAGT 2021, Aarhus, Denmark, September 21-24, 2021, Proceedings*, volume 12885 of *Lecture Notes in Computer Science*, page 411. Springer.
- Gersbach, H., Mamageishvili, A., and Tejada, O. (2021b). The effect of handicaps on turnout for large electorates with an application to Assessment Voting. *Journal of Economic Theory*, 195:105228.
- Goodman, L. (2014). Tezos — A self-amending crypto-ledger white paper. <https://tezos.com/whitepaper.pdf>.
- Halaburda, H., He, Z., and Li, J. (2021). An economic model of consensus on distributed ledgers. Working Paper 29515, National Bureau of Economic Research.
- Hayden Adams, Noah Zinsmeister, D. R. (2020). Uniswap v2 core. <https://uniswap.org/whitepaper.pdf>.
- Herrera, H., Morelli, M., and Palfrey, T. (2014). Turnout and power sharing. *Economic Journal*, 124(574):F131–F162.

- John, K., Rivera, T., and Saleh, F. (2020). Economic implications of scaling blockchains: Why the consensus protocol matters. <https://ssrn.com/abstract=3750467> or <http://dx.doi.org/10.2139/ssrn.3750467>.
- Kiayias, A., Koutsoupias, E., Kyropoulou, M., and Tselekounis, Y. (2016). Blockchain mining games. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, EC '16, page 365–382, New York, NY. Association for Computing Machinery.
- Lamport, L., Shostak, R., and Pease, M. (1982). The Byzantine generals problem. *ACM Transactions on Programming Languages and Systems*, 4(3):382–401.
- Ledyard, J. O. (1984). The pure theory of large two-candidate elections. *Public Choice*, 44(1):7–41.
- Leonardos, S., Reijnders, D., and Piliouras, G. (2020). Weighted voting on the blockchain: Improving consensus in proof of stake protocols. *International Journal of Network Management*, 30(5):e2093.
- Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. <https://bitcoin.org/bitcoin.pdf>.
- Osgood, R. (2016). The future of democracy: Blockchain voting. <https://www.cs.tufts.edu/comp/116/archive/fall2016/rosgood.pdf>.
- Palfrey, T. R. and Rosenthal, H. (1983). A strategic calculus of voting. *Public Choice*, 41(1):7–53.
- Wood, G. (2016). Polkadot: Vision for a heterogeneous multi-chain framework. <https://polkadot.network/PolkaDotPaper.pdf>.
- Yakovenko, A. (2017). Solana: A new architecture for a high performance blockchain v0.8.13. <https://solana.com/solana-whitepaper.pdf>.