Impossibility of Full Decentralization in Permissionless Blockchains

Yujin Kwon*, Jian Liu†, Minjeong Kim*, Dawn Song†, Yongdae Kim*

*KAIST
{dbwls8724,mjkim9394,yongdaek}@kaist.ac.kr

†UC Berkeley
jian.liu@eecs.berkeley.edu,dawnsong@cs.berkeley.edu

ABSTRACT

Bitcoin uses the *proof-of-work* (PoW) mechanism where nodes earn rewards in return for the use of their computing resources. Although this incentive system has attracted many participants, power has, at the same time, been significantly biased towards a few nodes, called *mining pools*. In addition, poor decentralization appears not only in PoW-based coins but also in coins that adopt *proof-of-stake* (PoS) and *delegated proof-of-stake* (DPoS) mechanisms.

In this paper, we address the issue of centralization in the consensus protocol. To this end, we first define (m, ε, δ)*-decentralization* as a state satisfying that 1) there are at least m participants running a node, and 2) the ratio between the total resource power of nodes run by the richest and the δ-th percentile participants is less than or equal to $1 + \varepsilon$. Therefore, when m is sufficiently large, and ε and δ are 0, (m, ε, δ)-decentralization represents *full decentralization*, which is an ideal state. To ascertain if it is possible to achieve good decentralization, we introduce conditions for an incentive system that will allow a blockchain to achieve (m, ε, δ)-decentralization. When satisfying the conditions, a blockchain system can reach full decentralization with probability 1, regardless of its consensus protocol. However, to achieve this, the blockchain system should be able to assign a positive Sybil cost, where the Sybil cost is defined as the difference between the cost for one participant running multiple nodes and the total cost for multiple participants each running one node. Conversely, we prove that if there is no Sybil cost, the probability of achieving (m, ε, δ)-decentralization is bounded above by a function of f_δ, where f_δ is the ratio between the resource power of the δ-th percentile and the richest participants. Furthermore, the value of the upper bound is close to 0 for small values of f_δ. Considering the current gap between the rich and poor, this result implies that it is almost impossible for a system without Sybil costs to achieve good decentralization. In addition, because it is yet unknown how to assign a Sybil cost without relying on a TTP in blockchains, it also represents that currently, a contradiction between achieving good decentralization in the consensus protocol and not relying on a TTP exists.

CCS CONCEPTS

• **Security and privacy → Economics of security and privacy**; **Distributed systems security**;

KEYWORDS

Blockchain; Consensus Protocol; Decentralization

1 INTRODUCTION

Traditional currencies have a centralized structure, and thus there exist several problems such as a single point of failure and corruption. For example, the global financial crisis in 2008 was aggravated by the flawed policies of banks that eventually led to many bank failures, followed by an increase in the distrust of these institutions. With this background, Bitcoin [36], which is the first decentralized digital currency, has received considerable attention; given that it is a decentralized cryptocurrency, there is no organization that controls the system, unlike traditional financial systems.

To operate the system without any central authority, Bitcoin uses *blockchain* technology. Blockchain is a public ledger that stores transaction history, while nodes record the history on the blockchain by generating blocks through a consensus protocol that provides a synchronized view among nodes. Bitcoin adopts a consensus protocol using the PoW mechanism in which nodes utilize their computational power in order to participate. Moreover, nodes receive coins as a reward for the use of their computational power, and this reward increases with the amount of computational power used. This incentive system has attracted many participants. At the same time, however, computational power has been significantly biased towards a few participants (i.e., mining pools). As a result, the decentralization of the Bitcoin system has become poor, thus deviating from its original goal [2, 19, 20].

Since the success of Bitcoin, many cryptocurrencies have been developed. They have attempted to address several drawbacks of Bitcoin, such as low transaction throughput, waste of energy owing to the utilization of vast computational power, and poor decentralization. Therefore, some cryptocurrencies use consensus mechanisms different from PoW, such as PoS and DPoS, in which nodes should have stakes for participation instead of a computing resource. While these new consensus mechanisms have addressed several of the drawbacks of Bitcoin, the problem of poor decentralization still remains unsolved. For example, similar to PoW systems, stakes are

AFT '19, October 21–23, 2019, Zurich, Switzerland
© 2019 Association for Computing Machinery.
ACM ISBN 978-1-4503-6732-5/19/10…$15.00
https://doi.org/10.1145/3318041.3355463

also significantly biased towards a few participants. This has caused concern about poor decentralization in PoS and DPoS coins.

Currently, many coins suffer from two problems that degrade the level of decentralization: 1) an insufficient number of independent participants because of the coalition of participants (e.g., mining pools) and 2) a significantly biased power distribution among them. Therefore, many developers have attempted to create a well-decentralized system [4, 5]. In addition, researchers such as Micali have noted that "incentives are the hardest thing to do" and believe that inappropriate incentive systems may cause blockchain systems to be significantly centralized [8]. This implies that it is currently an open problem as to whether we can design an incentive system that allows for good or full decentralization to be achieved.

Full decentralization. In this paper, the conditions for full decentralization are studied for the first time. To this end, we define (m, ε, δ)*-decentralization* as a state that satisfies that 1) *the number of participants running nodes in a consensus protocol is not less than m* and 2) *the ratio between the effective power of the richest and the δ-th percentile participants is not greater than* $1 + \varepsilon$, where the effective power of a participant represents the total resource power of the nodes run by that participant. The case when m is sufficiently large and ε and δ are 0 represents full decentralization in which everyone has the same power. To investigate if a high level of decentralization is possible, we model a blockchain system (Section 3), and find four sufficient conditions of the incentive system such that the blockchain system *converges in probability* to (m, ε, δ)-decentralization. If an incentive system that satisfies these four conditions exists, the blockchain system can achieve (m, ε, δ)-decentralization with probability 1, regardless of the underlying consensus protocol. The four conditions are: 1) *at least m nodes earn rewards*, 2) *it is not more profitable for participants to delegate their resource power to fewer participants than it is to run their own nodes*, 3) *it is not more profitable for a participant to run multiple nodes than to run one node*, and 4) *the ratio between the resource power of the richest and the δ-th percentile nodes converges in probability to a value of less than* $1 + \varepsilon$.

Impossibility. Based on these conditions, we find an incentive system that enables a system to achieve full decentralization. *In this incentive system, for the third condition to be met, the cost for one participant running multiple nodes should be greater than the total cost for multiple participants each running one node. The difference between the former cost and the latter cost is called a Sybil cost in this paper.* This implies that a system where Sybil costs exist can be fully decentralized with probability 1.

When a system does not have Sybil costs, there is no incentive system that satisfies the four conditions (Section 5). More specifically, the probability of reaching (m, ε, δ)-decentralization is bounded above by a function $G(f_\delta)$ that is close to 0 for a small ratio f_δ between the resource power of the δ-th percentile and the richest participants. This implies that achieving good decentralization in a system without Sybil costs depends totally on the rich-poor gap in the real world. As such, the larger the rich-poor gap, the closer the probability is to zero. To determine the approximate ratio f_δ in actual systems, we investigate hash rates in Bitcoin and observe that f_0 $(\delta = 0)$ and f_{15} $(\delta = 15)$ are less than 10^{-8} and 1.5×10^{-5}, respectively. In this case, f_0 indicates the ratio between the resource power of the poorest and the richest participants.

Unfortunately, it is not yet known how permissionless blockchains that have no *real identity management* can have Sybil costs. Indeed, to the best of our knowledge, all permissionless blockchains that do not rely on a TTP do not currently have any Sybil costs. Taking this into consideration, *it is almost impossible for permissionless blockchains to achieve good decentralization, and there is a contradiction between achieving good decentralization in the consensus protocol and not relying on a TTP.* The existence of mechanisms to enforce a Sybil cost in permissionless blockchains is left as an open problem. The solution to this issue would be the key to determining how blockchains can achieve a high level of decentralization.

Protocol analysis in the top 100 coins. Next, to find out what condition each system does not satisfy, we extensively analyze incentive systems of all existing PoW, PoS, and DPoS coins among the top 100 coins in CoinMarketCap [49] according to the four conditions (Section 6). According to this analysis, PoW and PoS systems do not have both enough participants running nodes and an even power distribution among the participants. However, unlike PoW and PoS coins, DPoS coins can have an even power distribution among a fixed number of participants when Sybil costs exist. If the Sybil costs do not exist, however, rational participants would run multiple nodes for higher profits. In that case, DPoS systems cannot guarantee that any participants possess the same power.

Data analysis in top 100 coins. To validate the result of the protocol analysis and our theory, we also conduct data analysis of the same list of coins using three metrics: the number of block generators, the Gini coefficient, and Shannon entropy (Section 7). Through this empirical study, we can observe the expected rational behaviors in most existing coins. In addition, we *quantitatively confirm* that the coins do not currently achieve good decentralization. As a result, this data analysis not only investigates the actual level of decentralization, but also empirically confirms the analysis results of incentive systems. We discuss the debate surrounding incentive systems and whether we can relax the conditions for full decentralization (Section 8). Finally, we conclude and provide two directions to go (Section 10).

2 IMPORTANCE OF DECENTRALIZATION

Decentralization is an essential factor that should be inherent in the design of blockchain systems. However, most of the computational power of PoW-based systems is currently concentrated in only a few nodes, called *mining pools*,[1] where individual miners gather together for mining. This causes concern not only about the level of decentralization, but also about the security of systems since the mining-power distribution is a critical aspect to be considered in the security of PoW systems. In general, when a participant has large amounts of resource power, their behavior will significantly influence others in the consensus protocol. In other words, the more resources a participant has, the greater their influence on the system. Therefore, the resource power distribution implicitly represents the level of decentralization in the system.

[1] More specifically, this refers to centralized mining pools. Even though there are decentralized mining pools, given that centralized pools are major pools, we will, hereafter, simply refer to them as mining pools.

At this point, we can consider the following questions: "What can influential participants do in practice?" and "Can this behavior harm other nodes?" Firstly, there are attacks such as double spending and selfish mining, which can be executed by attackers with over 50% and 33% of the resource power, respectively. These attacks would result in significant financial damage [10]. In addition, in a consensus protocol combined with PBFT [7], malicious behavior of nodes that possess over 33% resource power can cause the consensus protocol to become stuck. It would certainly be more difficult for such attacks to be executed through collusion with others if the resource power is more evenly distributed. In addition, nodes participating in the consensus protocol verify transactions and generate blocks. More specifically, when generating a block, nodes choose which transactions to include in that block. Therefore, they can choose only the advantageous transactions while ignoring the disadvantageous transactions. For example, participants can exclude transactions issued by rivals in the process of generating blocks and, if they possess large amounts of power, validation of these transactions will often be delayed because the malicious participant has many opportunities to choose the transactions that will be validated. Even though the rivals can also retaliate against them, the damage from the retaliation depends on the power gap between the malicious participants and their rivals.

Furthermore, transaction issuers are required to pay transaction fees. The fees are usually determined by economic interactions [50]. This implies that the fees can depend on the behavior of block generators. For example, if they verify only transactions that have fees above a specific amount, the overall transaction fees can increase because users would have to pay a higher fee for their transactions to be validated. Considering this, the more the system is centralized, the closer it may become to oligopolies.

In fully decentralized systems, however, it would be significantly more difficult for the above problems to occur. Moreover, the system would certainly be fair to everyone. This propels the desire to achieve a fully decentralized system. Even though there have been many discussions and attempts to achieve good decentralization, existing systems except for a few coins [19, 25] have rarely been analyzed. This paper not only studies the possibility of full decentralization, but also extensively investigates the existing coins.

3 SYSTEM MODEL

In this section, we model a consensus protocol and an incentive system. Moreover, we introduce the notation used throughout this paper (see Tab. 1).

Consensus protocol. A blockchain system has a consensus protocol where player p_i participates and generates blocks by running their own nodes. The set of all nodes in the consensus protocol is denoted by $\mathcal{N}$, and that of the nodes run by player p_i is denoted by $\mathcal{N}_{p_i}$. Moreover, we define $\mathcal{P}$ as the set of all players running nodes in the consensus protocol (i.e., $\mathcal{P} = \{p_i | \mathcal{N}_{p_i} \neq \emptyset\}$). Therefore, $|\mathcal{N}|$ is not less than $|\mathcal{P}|$. In particular, if a player has multiple nodes, $|\mathcal{N}|$ would be greater than $|\mathcal{P}|$.

For nodes to participate in the consensus protocol, they should possess specific resources, and their influence significantly depends on their resource power. The resource power in consensus protocols using PoW and PoS mechanisms is in the form of computational power and stakes, respectively. Node $n_i \in \mathcal{N}$ possesses resource power $\alpha_{n_i}(> 0)$. Moreover, $\bar{\boldsymbol{\alpha}}$ denotes the vector of the resource power for all nodes (i.e., $\bar{\boldsymbol{\alpha}} = (\alpha_{n_i})_{n_i \in \mathcal{N}}$). We also denote the resource power owned by player p_i as α_{p_i} and the set of players with positive resource power as $\mathcal{P}_\alpha$ (i.e., $\mathcal{P}_\alpha = \{p_i | \alpha_{p_i} > 0\}$). Here, we note that these two sets, $\mathcal{P}_\alpha$ and $\mathcal{P}$, can be different because when players delegate their own power to others, they do not run nodes but possess the resource power (i.e., the fact that $\alpha_{p_i} > 0$ does not imply that $\mathcal{N}_{p_i} \neq \emptyset$). For clarity, we describe a mining pool as an example. In the pool, there are workers and an operator, where the workers own their resource power but delegate it to the operator without running a full node. Therefore, pool workers belong to $\mathcal{P}_\alpha$ but not $\mathcal{P}$ while the operator belongs to both $\mathcal{P}_\alpha$ and $\mathcal{P}$.

In fact, the influence of player p_i on the consensus protocol depends on the total resource power of the nodes run by the player rather than just its resource power α_{p_i}. Therefore, we define EP_{p_i}, the *effective power* of player p_i as $\sum_{n_i \in \mathcal{N}_{p_i}} \alpha_{n_i}$. Again, considering the preceding example of mining pools, the operator's effective power is the sum of the resource power of all pool workers while the workers have zero effective power. The maximum and δ-th percentile of $\{EP_{p_i} | p_i \in \mathcal{P}\}$ are denoted by $EP_{\max}$ and EP_δ, respectively, and $\bar{\boldsymbol{\alpha}}_{\mathcal{N}_{p_i}}$ represents a vector of the resource power of the nodes owned by player p_i (i.e., $\bar{\boldsymbol{\alpha}}_{\mathcal{N}_{p_i}} = (\alpha_{n_i})_{n_i \in \mathcal{N}_{p_i}}$). Note that $EP_{\max}$ and EP_{100} are the same. In addition, we consider the average time to generate one block as a *time unit* in the system. We use the superscript t to express time t. For example, $\alpha^t_{n_i}$ and $\bar{\boldsymbol{\alpha}}^{\boldsymbol{t}}$ represent the resource power of node n_i at time t and the vector of the resource power possessed by the nodes at time t, respectively.

Incentive system. To incentivize players to participate in the consensus protocol, the blockchain system must have an incentive system. The incentive system would assign rewards to nodes, depending on their resource power. Here, we define the utility function $U_{n_i}(\alpha_{n_i}, \bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}})$ of the node n_i as the expected net profit per time unit, where $\bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}}$ represents the vector of other nodes' resource power and the net profit indicates earned revenues with all costs subtracted. Specifically, the utility function $U_{n_i}(\alpha_{n_i}, \bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}})$ of node n_i can be expressed as

$$U_{n_i} = E[R_{n_i} | \bar{\boldsymbol{\alpha}}] = \begin{cases} \sum_{R_{n_i}} R_{n_i} \times \Pr(R_{n_i} | \bar{\boldsymbol{\alpha}}) & \text{if } R_{n_i} \text{ is discrete} \\ \int_{R_{n_i}} R_{n_i} \times \Pr(R_{n_i} | \bar{\boldsymbol{\alpha}}) & \text{otherwise,} \end{cases}$$

where R_{n_i} is a random variable with probability distribution $\Pr(R_{n_i} | \bar{\boldsymbol{\alpha}})$ for a given $\bar{\boldsymbol{\alpha}}$. This equation for U_{n_i} and R_{n_i} indicates that U_{n_i} is the arithmetic mean of the random variable R_{n_i} for given $\bar{\boldsymbol{\alpha}}$. In addition, while function U_{n_i} indicates the expected net profit that node n_i can earn for the time unit, random variable R_{n_i} represents all possible values of the net profit that node n_i can obtain for the time unit. For clarity, we give an example of the Bitcoin system, whereby R_{n_i} and $\Pr(R_{n_i} | \bar{\boldsymbol{\alpha}})$ are defined as:

$$R_{n_i} = \begin{cases} 12.5 \text{ BTC} - c_{n_i} & \text{if } n_i \text{ generates a block} \\ -c_{n_i} & \text{otherwise,} \end{cases}$$

$$\Pr(R_{n_i} = a | \bar{\boldsymbol{\alpha}}) = \begin{cases} \frac{\alpha_{n_i}}{\sum_{n_j \in \mathcal{N}} \alpha_{n_j}} & \text{if } a = 12.5 \text{ BTC} - c_{n_i} \\ 1 - \frac{\alpha_{n_i}}{\sum_{n_j \in \mathcal{N}} \alpha_{n_j}} & \text{otherwise,} \end{cases}$$

where c_{n_i} represents all costs associated with running node n_i during the time unit. This is because a node currently earns 12.5 BTC as the block reward, and the probability of generating a block is proportional to its computing resource. Moreover, R_{n_i} cannot be greater than a constant R_{max}, determined in the system. In other words, the system can provide nodes with a limited value of rewards at a given time. Indeed, the reward that a node can receive for a time unit cannot be infinity, and problems such as inflation would occur if the reward were significantly large.

In addition, if nodes can receive more rewards when they have larger resource power, then players would increase their resources by spending a part of the earned profit. In that case, for simplicity, we assume that all players increase their resource power per earned net profit R_{n_i} at rate r every time. For example, if a node earns a net profit $R^t_{n_i}$ at time t, the node's resource power would increase by $r \cdot R^t_{n_i}$ after time t.

We also define the *Sybil cost function* $C(\bar{\boldsymbol{\alpha}}_{\mathcal{N}_{p_i}})$ as an additional cost that a player should pay per time unit to run multiple nodes compared to the total cost of when those nodes are run by different players. The cost $C(\bar{\boldsymbol{\alpha}}_{\mathcal{N}_{p_i}})$ would be 0 if $|\mathcal{N}_{p_i}|$ is 1 (i.e., the player p_i runs one node). Moreover, the case where $C(\bar{\boldsymbol{\alpha}}_{\mathcal{N}_{p_i}}) > 0$ for any set $\mathcal{N}_{p_i}$ such that $|\mathcal{N}_{p_i}| > 1$ indicates that the cost for one player to run $M(> 1)$ nodes is always greater than the total cost for M players each running one node. Note that this definition does not just imply that it is expensive to run many nodes, the cost of which is usually referred to as Sybil costs in the consensus protocol [9], *this function implies that the total cost for running multiple nodes depends on whether one player runs those nodes.*

Finally, we assume that all players are rational. Thus, they act in the system for higher utility. More specifically, if there is a coalition of players in which the members can earn a higher profit, they delegate their power to form such a coalition (formally, it is referred to as a cooperative game). In addition, if it is more profitable for a player to run multiple nodes as opposed to one node, the player would run multiple nodes.

Table 1: List of parameters.

Notation	Definition
p_i	Player of index i
$\mathcal{P}$	The set of players running nodes in the consensus protocol
n_i	Node of index i
$\mathcal{N}$	The set of nodes in the consensus protocol
$\mathcal{N}_{p_i}$	The set of nodes owned by p_i
α_{n_i}, α_{p_i}	The resource power of node n_i and player p_i
$\bar{\boldsymbol{\alpha}}$	The vector of resource power α_{n_i} for all nodes
$\mathcal{P}_\alpha$	The set of players with positive resource power
EP_{p_i}	The effective power of nodes run by p_i
EP_{max}, EP_δ	The maximum and δ-th percentile of effective power of players running nodes
$\bar{\boldsymbol{\alpha}}_{\mathcal{N}_{p_i}}$	The vector of resource power of nodes run by p_i
$\alpha^t_{n_i}$	The resource power of n_i at time t
$\bar{\boldsymbol{\alpha}}^{\boldsymbol{t}}$	The vector of resource power at time t
$\bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}}$	The vector of resource power of nodes other than n_i
$U_{n_i}(\alpha_{n_i}, \bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}})$	Utility function of n_i
R_{n_i}	Random variable for a net reward of n_i per time unit
R_{max}	The maximum value of random variable R_{n_i}
r	Increasing rate of resource power per the net profit
$C(\bar{\boldsymbol{\alpha}}_{\mathcal{N}_{p_i}})$	Sybil cost function of p_i

4 CONDITIONS FOR FULL DECENTRALIZATION

In this section, we study the circumstances under which a high level of decentralization can be achieved. To this end, we first formally define (m, ε, δ)-decentralization and introduce the sufficient conditions of an incentive system that will allow a blockchain system to achieve (m, ε, δ)-decentralization. Then, based on these conditions, we find such an incentive system.

4.1 Full Decentralization

The level of decentralization largely depends on two elements: the number of players running nodes in a consensus protocol and the distribution of effective power among the players. In this paper, full decentralization refers to the case where a system satisfies that 1) the number of players running nodes is as large as possible and 2) the distribution of effective power among the players is even. Therefore, if a system does not satisfy one of these requirements, it cannot become fully decentralized. For example, in the case where only two players run nodes with the same resource power, only the second requirement is satisfied. As another example, a system may have many nodes run by independent players with the resource power being biased towards a few nodes. Then, in this case, only the first requirement is satisfied. Clearly, both of these cases have poor decentralization. Note that, as described in Section 2, blockchain systems based on a peer-to-peer network can be manipulated by partial players who possess in excess of 50% or 33% of the effective power. Next, to reflect the level of decentralization, we formally define (m, ε, δ)-decentralization as follows.

Definition 4.1 ((m, ε, δ)-Decentralization). For $1 \leq m$, $0 \leq \varepsilon$, and $0 \leq \delta \leq 100$, a system is **(m, ε, δ)-decentralized** if it satisfies that

1. The size of $\mathcal{P}$ is not less than m (i.e., $|\mathcal{P}| \geq m$),
2. The ratio between the effective power of the richest player, EP_{max}, and the δ-th percentile player, EP_δ, is less than or equal to $1 + \varepsilon$ (i.e., $\frac{EP_{\mathsf{max}}}{EP_\delta} \leq 1 + \varepsilon$).

In Def. 4.1, the first requirement indicates that not only there are players that possess resources, but also that at least m players should run their own nodes. In other words, too many players do not combine into one node (i.e., many players do not delegate their resources to others.). Note that delegation decreases the number of players running nodes in the consensus protocol. The second requirement ensures an even distribution of the effective power among players running nodes. Specifically, for the richest and the δ-th percentile players running nodes, the gap between their effective power should be small. According to Def. 4.1, it is evident that as m increases and ε and δ decrease, the level of decentralization increases. Therefore, $(m, 0, 0)$-decentralization for a sufficiently large m indicates full decentralization where there is a sufficiently large number of independent players and everyone has the same power.

4.2 Sufficient Conditions

Next, we introduce four sufficient conditions of an incentive system that will allow a blockchain system to achieve (m, ε, δ)-decentralization with probability 1. We first revisit the two requirements of (m, ε, δ)-decentralization. For the first requirement in Def. 4.1, the size of $\mathcal{N}$ should be greater than or equal to m because the size of $\mathcal{P}$ is never greater than that of $\mathcal{N}$. This can be achieved by assigning rewards to at least m nodes. This approach is presented in Condition 1 (GR-m). In addition, it should not be more profitable for too many players to combine into a few nodes than it is when they run their nodes directly. If delegating is more profitable than not delegating, many players with resource power would delegate their power to a few players, resulting in $|\mathcal{P}| < m$. Condition 2 (ND-m) states that it should not be more profitable for nodes run by independent (or different) players to combine into fewer nodes when the number of all players running nodes is not greater than m.

Condition 1 (**Giving Rewards (GR-*m*)**). *At least m nodes should earn net profit. Formally, for any $\bar{\boldsymbol{\alpha}}$, $|\mathcal{N}^+| \geq m$, where*

$$\mathcal{N}^+ = \{n_i \in \mathcal{N} \,|\, U_{n_i}(\alpha_{n_i}, \bar{\boldsymbol{\alpha}}_{\boldsymbol{-n_i}}) > 0\}.$$

This condition states that some players can earn a reward by running a node, which makes the number of existing nodes equal to or greater than m. Meanwhile, if the system does not give net profit, rational players would not run a node because the system requires a player to possess a specific resource (i.e., $\alpha_{n_i} > 0$) in order to run a node unlike other peer-to-peer systems such as Tor. Simply put, players should invest their resource power elsewhere for higher profits instead of participating in a consensus protocol with no net profit, which is called an opportunity cost [18]. As a result, to reach (m, δ, ϵ)-decentralization, it is also necessary for a system to give net profit to some nodes.

Condition 2 (**Non-Delegation (ND-*m*)**). *Nodes run by different players do not combine into fewer nodes unless the number of all players running their nodes is greater than m. Before defining it formally, we denote a set of nodes run by different players by $\mathcal{S}^d$. That is, for any $n_i, n_j \in \mathcal{S}^d$, the two players running n_i and n_j are different. We also let s^d denote a proper subset of $\mathcal{S}^d$ such that $|\mathcal{P}(\mathcal{N}\backslash\mathcal{S}^d \cup s^d)| < m$, where*

$$\mathcal{P}(\mathcal{N}\backslash\mathcal{S}^d \cup s^d) = \{p_i \in \mathcal{P} \,|\, \exists n_i \in (\mathcal{N}\backslash\mathcal{S}^d \cup s^d) \text{ s.t. } n_i \in \mathcal{N}_{p_i}\}.$$

Then, for any set of nodes $\mathcal{S}^d$,

$$\sum_{n_i \in \mathcal{S}^d} U_{n_i}(\alpha_{n_i}, \bar{\boldsymbol{\alpha}}_{\boldsymbol{-n_i}}) \geq$$
$$\max_{\substack{s^d \subsetneq \mathcal{S}^d \\ \bar{\boldsymbol{\alpha}}_{\boldsymbol{d}} \in s^d_\alpha}} \Big\{ \sum_{\alpha_{n_i} \in \bar{\boldsymbol{\alpha}}_{\boldsymbol{d}}} U_{n_i}(\alpha_{n_i}, \boldsymbol{\alpha}^{-}_{\boldsymbol{-n_i}}(\mathcal{S}^d \backslash s^d)) \Big\}, \tag{1}$$

where,

$$s^d_\alpha = \Big\{ \bar{\boldsymbol{\alpha}}_{\boldsymbol{d}} = (\alpha_{n_i})_{n_i \in s^d} \,\Big|\, \sum_{\alpha_{n_i} \in \bar{\boldsymbol{\alpha}}_{\boldsymbol{d}}} \alpha_{n_i} = \sum_{n_i \in \mathcal{S}_d} \alpha_{n_i} \Big\},$$

and $\boldsymbol{\alpha}^{-}_{\boldsymbol{-n_i}}(\mathcal{S}^d \backslash s^d) = (\alpha_{n_j})_{n_j \notin \mathcal{S}^d \backslash s^d, n_j \neq n_i}$.

The set $\mathcal{P}(\mathcal{N}\backslash\mathcal{S}^d \cup s^d)$ represents all players running nodes that do not belong to $\mathcal{S}^d \backslash s^d$. In Eq. (1), the left-hand side represents the total utility of the nodes in $\mathcal{S}^d$ that are individually run by different players. Here, given that $\mathcal{S}^d \subseteq \mathcal{N}$, we note that $\bar{\boldsymbol{\alpha}}_{\boldsymbol{-n_i}}$ includes the resource power of the nodes in $\mathcal{S}^d$ except for node n_i. The right-hand side represents the maximum total utility of the nodes in s^d when the nodes in $\mathcal{S}^d$ are combined into fewer nodes belonging to s^d by delegation of resource power of players. Note that $|s^d| < |\mathcal{S}^d|$ because $s^d \subsetneq \mathcal{S}^d$. Therefore, Eq. (1) indicates that the utility in the case where multiple players delegate their power to fewer players is not greater than that for the case where the players directly run nodes. As a result, ND-m prevents delegation that results in the number of players running nodes being less than m, and the first requirement of (m, ε, δ)-decentralization can be met when GR-m and ND-m hold.

Next, we consider the second requirement in Def. 4.1. One way to achieve an even distribution of effective power among players is to cause the system to have an even resource power distribution among nodes while each player has only one node. Note that in this case where each player has only one node, an even distribution of their effective power is equivalent to an even resource power distribution among nodes. Condition 3 (NS-δ) states that, for any player with above the δ-th percentile effective power, running multiple nodes is not more profitable than running one node. In addition, to reach a state where the richest and the δ-th percentile nodes possess similar resource power, the ratio between the resource power of these two nodes should *converge in probability* to a value of less than $1 + \varepsilon$. This is presented in Condition 4 (ED-(ε, δ)).

Condition 3 (**No Sybil nodes (NS-*δ*)**). *For any player with effective power not less than EP_δ, participation with multiple nodes is not more profitable than participation with one node. Formally, for any player p_i with effective power $\alpha \geq EP_\delta$,*

$$\max_{\substack{\{\mathcal{N}_{p_i} : |\mathcal{N}_{p_i}| > 1\} \\ \bar{\boldsymbol{\alpha}}_{\boldsymbol{\mathcal{N}_{p_i}}} \in \mathcal{S}^{p_i}_\alpha}} \Big\{ \sum_{\alpha_{n_i} \in \bar{\boldsymbol{\alpha}}_{\boldsymbol{\mathcal{N}_{p_i}}}} U_{n_i}\Big(\alpha_{n_i}, \boldsymbol{\alpha}^{+}_{\boldsymbol{-n_i}}(\mathcal{N}_{p_i})\Big) - C(\bar{\boldsymbol{\alpha}}_{\boldsymbol{\mathcal{N}_{p_i}}}) \Big\}$$
$$\leq U_{n_j}(\alpha_{n_j} = \alpha, \bar{\boldsymbol{\alpha}}_{\boldsymbol{-\mathcal{N}_{p_i}}}), \tag{2}$$

where node $n_j \in \mathcal{N}_{p_i}$, the set $\bar{\boldsymbol{\alpha}}_{\boldsymbol{-\mathcal{N}_{p_i}}} = (\alpha_{n_k})_{n_k \notin \mathcal{N}_{p_i}}$, $\boldsymbol{\alpha}^{+}_{\boldsymbol{-n_i}}(\mathcal{N}_{p_i}) = \bar{\boldsymbol{\alpha}}_{\boldsymbol{-\mathcal{N}_{p_i}}} \| (\alpha_{n_k})_{n_k \in \mathcal{N}_{p_i}, n_k \neq n_i}$, and

$$\mathcal{S}^{p_i}_\alpha = \Big\{ \bar{\boldsymbol{\alpha}}_{\boldsymbol{\mathcal{N}_{p_i}}} = (\alpha_{n_i})_{n_i \in \mathcal{N}_{p_i}} \,\Big|\, \sum_{\alpha_{n_i} \in \bar{\boldsymbol{\alpha}}_{\boldsymbol{\mathcal{N}_{p_i}}}} \alpha_{n_i} = \alpha \Big\}.$$

In Eq. (2), the left and right-hand sides represent the maximum utility of the case where a player runs multiple nodes of which the total resource power is α, and the utility of the case where the player runs only one node n_j with resource power α, respectively. Therefore, Eq. (2) indicates that a player with equal to or greater than the δ-th percentile effective power can earn the maximum utility when running one node.

Condition 4 (**Even Distribution (ED-(ε, δ))**). *The ratio between the resource power of the richest and the δ-th percentile nodes should* ***converge in probability*** *to a value less than $1 + \varepsilon$. Formally, when $\alpha^t_{\max}$ and α^t_δ represent the maximum and the δ-th percentile of $\{\alpha^t_{n_i} | n_i \in \mathcal{N}^t\}$, respectively,*

$$\lim_{t \to \infty} \Pr\Big[\frac{\alpha^t_{\max}}{\alpha^t_\delta} \leq 1 + \varepsilon \Big] = 1.$$

The above condition indicates that when enough time is given, the ratio between the resource power of the richest and the δ-th percentile nodes reaches a value less than $1 + \varepsilon$ with probability 1. We note that $\alpha^t_{n_i}$ changes over time, depending on the behavior of each player. In particular, if it is profitable for a player to increase their effective power, $\alpha^t_{n_i}$ would be a random variable related to

$R^t_{n_i}$ because a player would reinvest part of their net profit $R^t_{n_i}$ to increase their resources. More specifically, in that case, $\alpha^t_{n_i}$ increases to $\alpha^t_{n_i} + rR^t_{n_i}$ after time t as described in Section 3.

As a result, these four conditions allow blockchain systems to reach (m, ε, δ)-decentralization with probability 1, as is presented in the following theorem. The proof of the theorem is omitted because it follows the above logic.

Theorem 4.2. *For any initial state, a system satisfying GR-m, ND-m, NS-δ, and ED-(ε, δ) converges in probability to (m, ε, δ)-decentralization.*

4.3 Possibility of Full Decentralization in Blockchain

To determine whether blockchain systems can achieve full decentralization, we study the existence of an incentive system satisfying these four conditions for a sufficiently large m, $\delta = 0$, and $\varepsilon = 0$. We provide an example of an incentive system that satisfies the four conditions, thus allowing full decentralization to be achieved.

It is also important to increase the total resource power involved in the consensus protocol from the perspective of security. This is because if the total resource power involved in the consensus protocol is small, an attacker can easily subvert the system. Therefore, to prevent this, we construct $U_{n_i}(\alpha_{n_i}, \bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}})$ as an increasing function of α_{n_i}, which implies that players continually increase their resource power. In addition, we construct random variable R_{n_i} with probability $\Pr(R_{n_i} | \bar{\boldsymbol{\alpha}})$ as follows:

$$R_{n_i} = \begin{cases} B_r & \text{if } n_i \text{ generates a block} \\ 0 & \text{otherwise} \end{cases}, \tag{3}$$

$$\Pr(R_{n_i} = a \mid \bar{\boldsymbol{\alpha}}) = \begin{cases} \frac{\sqrt{\alpha_{n_i}}}{\sum_{n_j \in \mathcal{N}} \sqrt{\alpha_{n_j}}} & \text{if } a = B_r \\ 1 - \frac{\sqrt{\alpha_{n_i}}}{\sum_{n_j \in \mathcal{N}} \sqrt{\alpha_{n_j}}} & \text{otherwise} \end{cases}, \tag{4}$$

$$U_{n_i}(\alpha_{n_i}, \bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}}) = \frac{B_r \cdot \sqrt{\alpha_{n_i}}}{\sum_{n_j \in \mathcal{N}} \sqrt{\alpha_{n_j}}}, \tag{5}$$

where the superscript t representing time t is omitted for convenience. This incentive system indicates that when a node generates a block, it earns the block reward B_r, and the probability of generating a block is proportional to the square root of the node's resource power. Under these circumstances, we can easily check that the utility function U_{n_i} is a mean of R_{n_i}.

Next, we show that this incentive system satisfies the four conditions. Firstly, the utility satisfies GR-m for any m because it is always positive. ND-m is also satisfied because the following equation is satisfied: This can be easily proven by using the fact that the utility is a concave function.

$$\sum_{i=1}^{m} U_{n_i}(\alpha_{n_i}, \bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}}) > U_{n_i}\Big(\sum_{i=1}^{m} \alpha_{n_i} \,\Big|\, (\alpha_{n_j})_{j>m}\Big)$$

Thirdly, to make NS-0 true, we can choose a proper Sybil cost function C of Eq. (2), which satisfies the following:

$$\sum_{i=1}^{M} U_{n_i}(\alpha_{n_i}, \bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}}) - U_{n_i}\Big(\sum_{i=1}^{M} \alpha_{n_i} \,\Big|\, (\alpha_{n_j})_{j>M}\Big) \le C((\alpha_{n_i})_{i \le M})$$

Under this Sybil cost function, the players would run only one node. Finally, to show that this incentive system satisfies ED-$(0, 0)$, we use the following theorem, whose proof is presented in the full version [28].

Theorem 4.3. *Assume that R_{n_i} is defined as follows:*

$$R_{n_i} = \begin{cases} f(\bar{\boldsymbol{\alpha}}) & \text{if } n_i \text{ generates a block} \\ 0 & \text{otherwise} \end{cases},$$

where $f : \mathbb{R}^{|\mathcal{N}|} \mapsto \mathbb{R}^+$. If $U_{n_i}(\alpha_{n_i}, \bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}})$ is a strictly increasing function of α_{n_i} and the following equation is satisfied for all $\alpha_{n_i} > \alpha_{n_j}$, ED-$(\varepsilon, \delta)$ is satisfied for all ε and δ.

$$\frac{U_{n_i}(\alpha_{n_i}, \bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}})}{\alpha_{n_i}} < \frac{U_{n_j}(\alpha_{n_j}, \bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_j}})}{\alpha_{n_j}} \tag{6}$$

On the contrary, if $U_{n_i}(\alpha_{n_i}, \bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}})$ is a strictly increasing function of α_{n_i} and Eq. (6) is not satisfied for all $\alpha_{n_i} > \alpha_{n_j}$, ED-$(\varepsilon, \delta)$ cannot be met for all $0 \le \varepsilon < \frac{EP^0_{\max}}{EP^0_{\delta}} - 1$ and $0 \le \delta < 100$.

Thm. 4.3 states that when the utility is a strictly increasing function of α_{n_i} and Eq. (6) is satisfied under the assumption that the block reward is constant for a given $\bar{\boldsymbol{\alpha}}$, an even power distribution is achieved. Meanwhile, if Eq. (6) is not met, the gap between rich and poor nodes cannot be narrowed. Specifically, for the case where $\frac{U_{n_i}(\alpha_{n_i}, \bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}})}{\alpha_{n_i}}$ is constant, the large gap between rich and poor nodes can be continued[2]. Moreover, the gap would widen when $\frac{U_{n_i}(\alpha_{n_i}, \bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}})}{\alpha_{n_i}}$ is a strictly increasing function of α_{n_i}. In fact, here $\frac{U_{n_i}(\alpha_{n_i}, \bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}})}{\alpha_{n_i}}$ can be considered as an increasing rate of resource power of a node. Thus, Eq. (6) indicates that the resource power of a poor node increases faster than that of a rich node.

Now, we describe why the incentive system defined by Eq. (3), (4), and (5) satisfies ED-$(0, 0)$. Firstly, Eq. (3) is a form of R_{n_i} described in Thm. 4.3, and Eq. (5) implies that U_{n_i} is a strictly increasing function of α_{n_i}. Therefore, ED-$(0, 0)$ is met by Thm. 4.3 because Eq. (5) satisfies Eq. (6) for all $\alpha_{n_i} > \alpha_{n_j}$. As a result, the incentive system defined by Eq. (3), (4), and (5) satisfies the four sufficient conditions, *implying that full decentralization is possible under a proper Sybil cost function C*. Moreover, Thm. 4.3 describes the existence of infinitely many incentive systems that can facilitate full decentralization. Interestingly, we have found that an incentive scheme similar to this is being considered by the Ethereum foundation, who have also indicated that *real identity management* can be important [5]. This finding is in accordance with our results.

5 IMPOSSIBILITY OF FULL DECENTRALIZATION IN PERMISSIONLESS BLOCKCHAINS

In the previous section, we showed that blockchain systems can be fully decentralized under an appropriate Sybil cost function C, where the Sybil cost represents the additional costs for a player running multiple nodes when compared to the total cost for multiple players each running one node. In order for a system to implement the Sybil cost, we can easily consider real identity management where a trusted third party (TTP) manages the *real identities* of players. When real identity management exists, it is certainly possible to implement a Sybil cost. However, the existence of a TTP contradicts the concept of decentralization, and thus, we cannot adopt such identity management for good decentralization. Currently, it is not yet known how permissionless blockchains without such identity

[2] Formally speaking, the probability of achieving an even power distribution among players is less than 1, and in Thm. 5.3, we will address how small the probability is.

management can implement Sybil cost. In fact, many cryptocurrencies are based on permissionless blockchains, and many people want to design permissionless blockchains on the basis of their nature. Unfortunately, as far as we know, the Sybil cost function C of all permissionless blockchains is currently zero. Taking this into consideration (i.e., $C = 0$), we examine whether blockchains without Sybil costs can achieve good decentralization in this section.

5.1 Almost Impossible Full Decentralization

To determine whether it is possible for a system without Sybil costs to achieve full decentralization, we describe the following theorem.

Theorem 5.1. *Consider a system without Sybil costs (i.e., $C = 0$). Then, the probability of the system achieving (m, ε, δ)-decentralization is always less than or equal to*

$$\max_{s\in\mathcal{S}} \Pr[\text{System } s \text{ reaches } (m,\varepsilon,\delta)\text{-decentralization}], \quad \textit{where}$$

$\mathcal{S}$ is the set of all systems satisfying GR-$|\mathcal{N}|$, ND-$|\mathcal{P}_\alpha|$, and NS-0.

GR-$|\mathcal{N}|$ means that all nodes can earn net profit, and the satisfaction of both ND-$|\mathcal{P}_\alpha|$ and NS-0 indicates that all players run only one node without delegating. **The above theorem implies that the maximum probability for a system, which satisfies GR-$|\mathcal{N}|$, ND-$|\mathcal{P}_\alpha|$, and NS-0, to reach (m, ε, δ)-decentralization is equal to the global maximum probability.** Moreover, according to Thm. 5.1, there is a system satisfying GR-$|\mathcal{N}|$, ND-$|\mathcal{P}_\alpha|$, NS-0, and ED-(ε, δ) *if and only if* there is a system that converges in probability to (m, ε, δ)-decentralization. In other words, the fact that a system satisfying GR-$|\mathcal{N}|$, ND-$|\mathcal{P}_\alpha|$, NS-0, and ED-(ε, δ) should exist is **sufficient and necessary** to create a system converging in probability to (m, ε, δ)-decentralization.

The proof of Thm. 5.1 is presented in the full version [28]. In the proof, we use the fact that the system can optimally change the state (i.e., the effective power distribution among players above the δ-th percentile) for (m, ε, δ)-decentralization when the system can recognize the current state (i.e., the current effective power distribution among players above the δ-th percentile). Then we prove that, to learn the current state, players above the δ-th percentile should run only one node, or coalition of some players should be more profitable. In that case, to make a system most likely to reach (m, ε, δ)-decentralization, resources of rich nodes should not increase through delegation of others. Considering this, we can derive Thm. 5.1.

According to Thm. 5.1, to find out if a system without Sybil costs can reach a high level of decentralization, it is sufficient to determine the maximum probability for a system satisfying GR-$|\mathcal{N}|$, ND-$|\mathcal{P}_\alpha|$, and NS-0 to reach (m, ε, δ)-decentralization. Therefore, we first find a utility function that satisfies GR-$|\mathcal{N}|$, ND-$|\mathcal{P}_\alpha|$, and NS-0 through the following lemma.

Lemma 5.2. *When the Sybil cost function C is zero, GR-$|\mathcal{N}|$, ND-$|\mathcal{P}_\alpha|$, and NS-0 are met if and only if*

$$U_{n_i}(\alpha_{n_i}, \bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}}) = F\Big(\sum_{n_j\in\mathcal{N}} \alpha_{n_j}\Big)\cdot\alpha_{n_i}, \ \textit{where } F:\mathbb{R}^+ \mapsto \mathbb{R}^+. \quad (7)$$

Eq. (7) implies that the utility function is linear when the total resource power of all nodes is given. Under this utility (i.e., net profit), a player would run one node with its own resource power because delegation of its resource and running multiple nodes are not more profitable than running one node with its resource power. Lem. 5.2 is proven using a proof by induction, and it is presented in the full version [28].

We then consider whether Eq. (7) can satisfy ED-(ε, δ). Note that when ED-(ε, δ) is satisfied, the probability of achieving (m, ε, δ)-decentralization is 1. Therefore, it is sufficient to answer the following question: "What is the probability of a system defined by Eq. (7) to reach (m, ε, δ)-decentralization?" Thm. 5.3 gives the answer by providing the upper bound of the probability. Before describing the theorem, we introduce several notations. Given that players, in practice, start running their nodes in the consensus protocol at different times, $\mathcal{P}$ would differ depending on the time. Thus, we use notations $\mathcal{P}^t$ and $\mathcal{P}^t_\delta$ to reflect this, where $\mathcal{P}^t_\delta$ is defined as:

$$\mathcal{P}^t_\delta = \{p_i \in \mathcal{P}^t | EP^t_{p_i} \geq EP^t_\delta\}.$$

That is, $\mathcal{P}^t_\delta$ indicates the set of all players who have above the δ-th percentile effective power at time t. Moreover, we define α_{MAX} and f_δ as

$$\alpha_{\mathsf{MAX}} = \max\left\{\alpha^{t_i^0}_{p_i} \,\middle|\, p_i \in \lim_{t\to\infty} \mathcal{P}^t\right\},$$

$$f_\delta = \min\left\{ \frac{\alpha^{t^0_{ij}}_{p_i}}{\alpha^{t^0_{ij}}_{p_j}} \,\middle|\, p_i, p_j \in \lim_{t\to\infty}\mathcal{P}^t_\delta,\ t^0_{ij} = \max\{t^0_i, t^0_j\}\right\},$$

where t^0_i denotes the time at which player p_i starts to participate in a consensus protocol. The parameter α_{MAX} indicates the initial resource power of the richest player among the players who remain in the system for a long time. Furthermore, f_δ represents the ratio between the δ-th percentile and the largest initial resource power of the players who remain in the system for a long time. Taking these notations into consideration, we present the following theorem.

Theorem 5.3. *When the Sybil cost function C is zero, the following holds for any incentive system that satisfies Eq. (7):*

$$\lim_{t\to\infty}\Pr\left[\frac{EP^t_{\max}}{EP^t_\delta} \leq 1+\varepsilon\right] < G^\varepsilon\left(f_\delta, \frac{rR_{\max}}{\alpha_{\mathsf{MAX}}}\right), \quad (8)$$

where $\lim_{f_\delta\to 0} G^\varepsilon(f_\delta, \frac{rR_{\max}}{\alpha_{\mathsf{MAX}}})$ and $\lim_{\alpha_{\mathsf{MAX}}\to\infty} G^\varepsilon(f_\delta, \frac{rR_{\max}}{\alpha_{\mathsf{MAX}}})$ are 0.

This theorem implies that the probability of achieving (m, ε, δ)-decentralization is less than $G^\varepsilon(f_\delta, \frac{rR_{\max}}{\alpha_{\mathsf{MAX}}})$. Here, note that $rR_{\max}$ represents the maximum resource power that can be increased by a player per time unit. Given that $\lim_{f_\delta\to 0} G^\varepsilon(f_\delta, \frac{rR_{\max}}{\alpha_{\mathsf{MAX}}}) = 0$, *the upper bound would be smaller when the rich-poor gap in the current state is larger.* In addition, the fact that $\lim_{\alpha_{\mathsf{MAX}}\to\infty} G^\varepsilon(f_\delta, \frac{rR_{\max}}{\alpha_{\mathsf{MAX}}})$ implies that the greater the difference between the resource power of the richest player and the maximum value that can be increased by a player per time unit, the smaller the upper bound.

In fact, to make a system more likely to reduce the rich-poor gap, poor nodes should earn a small reward with a high probability for some time, while rich nodes should get the reward $R_{\max}$ with a small probability. This is proved in the proof of Thm. 5.3, which is presented in the full version [28]. Note that, in that case, rich nodes would rarely increase their resources, but poor nodes would often increase their resources.

To determine how small $G^\varepsilon(f_\delta, \frac{rR_{\max}}{\alpha_{\mathsf{MAX}}})$ is for a small value of f_δ, we adopt a Monte Carlo method. This is because a large degree of complexity is required to compute a value of $G^{\varepsilon}(f_{\delta}, \frac{rR_{\max}}{\alpha_{\mathsf{MAX}}})$ directly. Fig. 1 displays the value of $G^{\varepsilon}(f_{\delta}, \frac{rR_{\max}}{\alpha_{\mathsf{MAX}}})$ with respect to f_{δ} and ε when $\frac{rR_{\max}}{\alpha_{\mathsf{MAX}}}$ is 0.1. For example, we can see that $G^0(10^{-4}, 0.1)$ is about 10^{-5}, and this implies that a state where the ratio between resource power of the δ-th percentile player and the richest player is 10^{-4} can reach $(m, 0, \delta)$-decentralization with a probability less than 10^{-5} even if infinite time is given. Note that $\varepsilon = 9, 99$, and 999 indicate that the effective power of the richest player is 10 times, 100 times, and 1000 times that of the δ-th percentile player in (m, ε, δ)-decentralization, respectively.

Fig. 1 shows that the probability of achieving (m, ε, δ)-decentralization is smaller when f_{δ} and ε are smaller. From Fig. 1, one can see that the value of $G^{\varepsilon}(f_{\delta}, \frac{rR_{\max}}{\alpha_{\mathsf{MAX}}})$ is significantly small for a small value of f_{δ}. This result means that the probability of achieving good decentralization is close to 0 if there is a large gap between the rich and poor, and the resource power of the richest player is large (i.e., the ratio $\frac{rR_{\max}}{\alpha_{\mathsf{MAX}}}$ is not large[3]). The values of $G^{\varepsilon}(f_{\delta}, \frac{rR_{\max}}{\alpha_{\mathsf{MAX}}})$ when $\frac{rR_{\max}}{\alpha_{\mathsf{MAX}}}$ is 10^{-2} and 10^{-4} are presented in the full version [28], and the values are certainly smaller than those presented in Fig. 1.

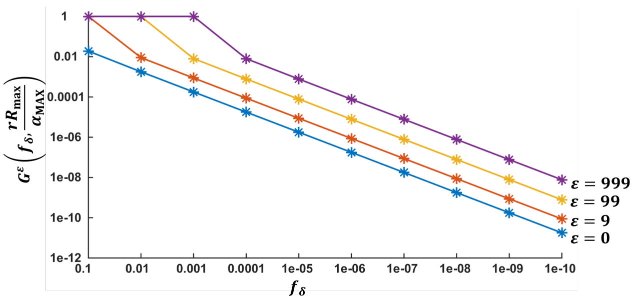

Figure 1: In this figure, when $\frac{rR_{\max}}{\alpha_{\mathsf{MAX}}}$ is 0.1, $G^{\varepsilon}(f_{\delta}, \frac{rR_{\max}}{\alpha_{\mathsf{MAX}}})$ (y-axis) is presented with respect to f_{δ} (x-axis) and ε.

To determine how small the ratio f_{δ} is at present, we use the hash rate of all users in Slush mining pool [44] in Bitcoin as an example. We find miners with hash rates lower than 3.061 GH/s and greater than 404.0 PH/s at the time of writing. Referring to these data, we can see that the ratio f_0 (i.e., the ratio between the resource power of the poorest and richest players) is less than $\frac{3.061\times10^9}{404.0\times10^{15}}$ ($\approx 7.58 \times 10^{-9}$). We also observe that the 15-th percentile and 50-th percentile hash rates are less than 5.832 TH/s and 25.33 TH/s, respectively. Therefore, the ratios f_{15} and f_{50} are less than approximately 1.44×10^{-5} and 6.27×10^{-5}, respectively. This example indicates that the rich-poor gap is significantly large. Moreover, we observe an upper bound of $\frac{rR_{\max}}{\alpha_{\mathsf{MAX}}}$ in the Bitcoin system. Given that the block reward is 12.5 BTC ($\approx$ \$65, 504), the maximum value of $rR_{\max}$ is approximately 384 TH. This maximum value can be derived, assuming that a player reinvests all earned rewards to increase their hash power. Then, an upper bound of $\frac{rR_{\max}}{\alpha_{\mathsf{MAX}}}$ would be 9.5×10^{-4}, which is certainly less than the value of 0.1 used in Fig. 1. **As a result, Thm. 5.3 implies that, currently, it is almost impossible for a system without Sybil costs to achieve good decentralization. In other words, the achievement of good decentralization in the consensus protocol and a non-reliance on a TTP, which are required for good decentralization of systems, contradict each other.**

5.2 Intuition and Implication

Here, we describe intuitively why a permissionless blockchain, which does not rely on any TTP, cannot reach good decentralization. Because a player with great wealth can possess more resources, the initial distribution of the resource power in a system depends significantly on the distribution of wealth in the real world when the system does not have any constraint of participation and can attract many players. Therefore, if wealth is equally distributed in the real world and many players are incentivized to participate in the consensus protocol, full decentralization can be easily achieved, even in permissionless blockchains where anyone can join without any permission processes. However, according to many research papers and statistics, the rich-poor gap is significant in the real world [22, 43, 47]. In addition, the wealth inequality is well known as one of the most glaring deficiencies in today's capitalist society, and resolving this problem is difficult.

In a permissionless blockchain, players can freely participate without any restrictions, and large wealth inequality would appear initially. Therefore, for the system to achieve good decentralization, its incentive system should be designed to gradually narrow the rich-poor gap. To this end, we can consider the following incentive system: Nodes receive net profit in proportion to the square root of their resource power on average (e.g., Eq. (5)). This incentive system can result in the resource power distribution among nodes being more even (see Section 4.3). However, this alone cannot satisfy NS-δ when there is no Sybil cost (i.e., $C = 0$). Therefore, to satisfy NS-δ, we can establish that the expected net profit decreases when the number of existing nodes increases. For example, B_r in Eq. (5) can be a decreasing function of the number of existing nodes. In this case, players with large resources would not run Sybil nodes because when they do so, their utilities decrease with the increase in the number of nodes. However, this approach has a side effect in that players ultimately delegate their power to a few other players in order to earn higher profits. This is because this rational behavior on the part of the players decreases the number of nodes. As a result, the above example intuitively describes that *the four conditions are contradictory when a Sybil cost does not exist*[4], and whether a permissionless blockchain can achieve good decentralization depends completely on how wide the gap is between the rich and the poor in the real world. This finding is supported by Thm. 5.3.

Conversely, if we can establish a method of implementing Sybil costs without relying on a TTP in blockchains, we would be able to resolve the contradiction between achieving good decentralization in the consensus protocol and non-reliance on a TTP. This allows for designing a blockchain that achieves good decentralization. We leave this as an open problem.

[3] The ratio $\frac{rR_{\max}}{\alpha_{\mathsf{MAX}}}$ does not need to be small.

[4] This does not imply the impossibility of full decentralization. It only implies that the probability of achieving full decentralization is less than 1.

5.3 Question and Answer

In this section, to further clarify the implications of our results, we present questions that academic reviewers or blockchain engineers have considered in the past and provide answers to them.

[Q1] "Creating more nodes does not increase your mining power, so why is this a problem?" Firstly, note that decentralization is significantly related to *real identities*. That is, when the number of independent players is large and the power distribution among them is even, the system has good decentralization. In this paper, we *do not claim* that the higher the number of Sybil nodes, the lower the level of decentralization. We simply assert that a system should have knowledge of the current power distribution among players to achieve good decentralization, and a system without real identity management can know the distribution when each player runs only one node. Moreover, we prove that, to achieve good decentralization as far as possible, all players should run only one node (Thm. 5.1).

[Q2] "Would a simple puzzle for registering as a block-submitter not be a possible Sybil cost, without identity management?" According to the definition of Sybil cost (Section 3), the cost to run one node should depend on whether a player runs another node. More specifically, the cost to run one node for a player who has other nodes should be greater than that for a player with no other nodes. Therefore, the proposed scheme cannot constitute a Sybil cost. Again, note that the Sybil cost described in this paper is different from that usually mentioned in PoW and PoS systems [9].

[Q3] "If mining power is delivered in proportion to the resources one has available (which would be an ideal situation in permissionless systems), achievement of good decentralization is clearly an impossibility. This seems rather self-evident." Naturally, a system would be centralized in its initial state because the rich-poor gap is large in the real world and only a few players may be interested in the system in the early stages. Considering this, our work investigates *whether there is a mechanism to achieve good decentralization.* Note that our goal is to reduce the gap between the effective power of the rich and poor, not the gap between their resource power. In other words, even if the rich possess significantly large resource power, the decentralization level can still be high if the rich participate in the consensus protocol with only part of their resource power and so not large effective power. To this end, we can consider a utility function, which is a decreasing function for a large input (e.g., a concave function). However, this function cannot still achieve good decentralization because it does not satisfy NS-δ. Note that, with a mechanism satisfying the four conditions, a system can *always* reach good decentralization regardless of the initial state. Unfortunately, our finding is that there is no mechanism satisfying the four conditions, which implies that the probability of achieving good decentralization is less than 1. To make matters worse, Thm. 5.3 states that the probability is bounded above by a value close to 0. *As a result, this implies that it is almost impossible for us to create a system with good decentralization without any Sybil cost, even if infinite time is given.*

[Q4] "I think when the rich invest a lot of money in a system, the system can become popular. So, if the large power of the rich is not involved in the system, can it become popular?" In this paper, we focus on the decentralization level in a consensus protocol, which performs a role as the government of a system. Therefore, good decentralization addressed in this paper implies a fair government rather than indicating that there are no rich or poor in the entire system. If the rich invest a lot of money in business (e.g., an application based on the smart contract) running on the system instead of the consensus protocol, the system may have a fair government and become popular. Indeed, the efforts to make a fair government also appear in the real world since people are extremely afraid of an unfair system in which the rich influence the government through bribes.

6 SUMMARY OF PROTOCOL ANALYSIS

To determine if what condition each system satisfies or not, we analyze the incentive systems of the top 100 coins extensively according to the four conditions. In this section, we summarize the protocol analysis (see the full version [28] for more details), and focus on the analysis of the coins with PoW, PoS, and DPoS mechanisms, which are the major consensus mechanisms of non-permissioned blockchains. Tab. 2 presents the results of the analysis, where the black circle (●) and the half-filled circle (◐) indicate the full and partial satisfaction of the corresponding condition, respectively. The empty circle (○) indicates that the corresponding condition is not satisfied at all. In addition, we mark each coin system with a triangle (▲) or an X (✗) depending on whether it partially implements or does not implement a Sybil cost, respectively. Here, partial Sybil cost means that the payment of the Sybil cost can be avoided by pretending that the multiple nodes run by one player are run by different players (i.e., players with different real identities). Note that PoW, PoS, and DPoS coins cannot have perfect Sybil costs because they are non-permissioned blockchains.

Proof of Work. Most PoW systems are designed to give nodes a block reward proportional to the ratio of the computational power of each node to the total power. In addition, there are electric bills that are dependent on the computational power, as well as the other costs associated with running a node, such as a large memory for the storage of blockchain data. The other cost required to run a node is independent of the computational power. Considering this, we can express a utility (i.e., an expected net profit) $U_{n_i}(\alpha_{n_i}, \bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}})$ of node n_i as follows:

$$U_{n_i}(\alpha_{n_i}, \bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}}) = B_r \cdot \frac{\alpha_{n_i}}{\sum_{n_j} \alpha_{n_j}} - c_1 \cdot \alpha_{n_i} - c_2. \tag{9}$$

In Eq. (9), B_r represents the block reward (e.g., 12.5 BTC in the Bitcoin system) that a node can earn for a time unit, and $c_1(> 0)$ and $c_2(> 0)$ represent the electric bill per computational power and the other costs incurred during the time unit, respectively. In particular, the cost c_2 is independent of the computational power. The values of the three coefficients, B_r, c_1, and c_2, determine whether the four conditions are satisfied.

Firstly, in order for the system to satisfy GR-m for any m, it should be able to assign rewards to nodes with small computational power. Considering Eq. (9) for appropriate values of B_r, there is $\bar{\boldsymbol{\alpha}} = (\alpha_{n_i})_{n_i \in \mathcal{N}}$ such that $U_{n_i}(\alpha_{n_i}, \bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}}) > 0$ for all nodes n_i. However, there also exists α_{n_i} such that $U_{n_i}(\alpha_{n_i}, \bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}}) < 0$ for a given $\bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}}$, which implies that the PoW system cannot satisfy GR-m for some values of m. In practice, CPU miners cannot earn net profit in the Bitcoin system. As special cases, in IOTA and

Table 2: Analysis of incentive systems

Coin name	Con 1	Con 2	Con 3	Con 4	Sybil cost
All PoW&PoS†	◐	○	●	○	✗
IOTA/ BridgeCoin/ Nano	○	○	●	●	✗
Cardano	◐	◐	◐	◐	✗
DPoS-1	◐	◐	◐*	◐	▲
DPoS-2	◐	◐	◐	◐	✗

† = except for IOTA, BridgeCoin, Cardano, and Nano; ●= fully satisfies the condition; ◐= partially satisfies the condition; ○= does not satisfy the condition; ▲= has partial Sybil costs; ✗= does not have Sybil costs;

BridgeCoin, there is no block reward because coin mining does not exist or has already been completed. These systems do not satisfy GR-m at all because the utility U_{n_i} is negative for all $\bar{\boldsymbol{\alpha}}$.

In addition, PoW systems cannot satisfy ND-m. This is because when m players run their own nodes, they must pay the additional cost of $(m-1)\cdot c_2$ as compared to the case where they run only one node by cooperating with one another. This cooperation is commonly observed in the form of centralized mining pools. Of course, because the variance of rewards decrease when players join the pools, many of them may join these pools. However, although there are decentralized pools (e.g., P2Pool [37] and SMARTPOOL [31]) in which players can reduce the variance of rewards and run a full node, most players do not join these decentralized pools owing to the cost of running a full node[5].

Meanwhile, for the aforementioned reason, the systems can satisfy NS-δ. Finally, PoW systems with an incentive system defined by Eq. (9) cannot satisfy ED-(ε, δ). Considering Thm. 4.3, we can easily derive this. *As a result, we expect that the current PoW systems have neither a sufficient number of independent players nor an even power distribution among the players.* On the other hand, IOTA and Bridgecoin, which do not have any incentives, satisfy both NS-δ and ED-(ε, δ) as trivial cases because rational players would not run nodes.

Proof of Stake. In PoS systems, nodes receive block rewards proportional to their stake. Therefore, in these systems, we can express the utility U_{n_i} as follows:

$$U_{n_i}(\alpha_{n_i}, \bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}}) = B_r \cdot \frac{\alpha_{n_i}}{\sum_j \alpha_{n_j}} - c \quad \text{if } \alpha_{n_i} \geq S_b. \tag{10}$$

B_r and c in Eq. (10) represent the block reward that a node can earn for a time unit and the cost required to run one node, respectively. S_b indicates the least amount of stakes required to run one node. Therefore, Eq. (10) implies that only nodes with stakes above S_b can be run and earn a reward proportional to their stake fraction.

Similar to PoW systems, PoS systems only satisfy GR-m for some m (i.e., partially satisfy GR-m) because there exists a large value of $\sum \alpha_{n_j}$ such that $U_{n_i}(\alpha_{n_i}, \bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}}) < 0$. In addition, it is more profitable for multiple players to run one node through cooperation when compared to running each different node. For example, if a player has a stake below S_b, rewards cannot be earned by running nodes in the consensus protocol. However, the player can receive a reward by delegating their stake to others. In addition, if multiple players run only one node, they can reduce the cost required to run nodes. Therefore, PoS systems do not satisfy ND-m. These behaviors are observed through PoS pools [38, 45] or leased PoS [30] in practice. This fact also implies that it is less profitable for one player to run multiple nodes than it is to run one node; thus, PoS systems satisfy NS-δ. Finally, considering Thm. 4.3, the system with Eq. (10) cannot satisfy ED-(ε, δ).

As shown in Tab. 2, the results are similar to those for PoW coins. *Therefore, as with PoW coins, PoS coins would have a restricted number of independent players and a biased power distribution among them.* Similar to IOTA and BridgeCoin, Nano does not provide incentives to run nodes. Therefore, the result of Nano is the same with IOTA and BridgeCoin. In addition, Cardano is planning to implement an incentive system different from that of the usual PoS systems [4]. The system has the goal that there should be k nodes with similar resource power for a given k. In fact, this incentive system has a similar property to DPoS systems, which will be described below.

Delegated Proof of Stake. DPoS systems are significantly different from PoW and PoS systems. In the systems, stake holders elect block generators through a voting process, where the voting power is proportional to the stake owned by the stake holders (i.e., voters). Then, the block generators have an equal opportunity to generate blocks and earn the same block rewards. Therefore, when we arrange $\bar{\boldsymbol{\alpha}} = \{\alpha_{n_i} | 1 \leq i \leq n\}$ in descending order, we can express the utility U_{n_i} in DPoS systems as follows:

$$U_{n_i}(\alpha_{n_i}, \bar{\boldsymbol{\alpha}}_{-\boldsymbol{n_i}}) = \begin{cases} B_r - c & \text{if } i \leq N_{\mathsf{dpos}} \\ -c & \text{otherwise} \end{cases}, \tag{11}$$

where B_r is a block reward that a node can earn on average per a time unit, and c represents the cost associated with running one node. In addition, N_{dpos} is a constant number given by the DPoS system. Eq. (11) implies that only N_{dpos} nodes with the most votes can earn rewards by generating blocks. However, not all DPoS systems have the same incentive scheme as Eq. (11). For example, EOS with $N_{\mathsf{dpos}} = 21$ gives small rewards to nodes ranked within the 100-th place [12]. Although incentive systems different from Eq. (11) exist, we describe the analysis results of the DPoS coins with respect to Eq. (11) because their properties are similar.

Firstly, the DPoS system attracts players who can obtain high voting power because it provides them with a block reward. Meanwhile, rational players who are unable to obtain high voting power cannot earn any rewards. Therefore, the system partially satisfies GR-m. Moreover, it is rational for multiple players with small stakes to delegate their stakes to one player by voting for that player, which is why this system is called a *delegated* PoS system. On the other hand, players with high stakes would run their own nodes by voting for themselves. For example, if two players have sufficiently high stakes and run two nodes, they can earn a total value of $2(B_r - c)$ as net profit. However, when they run only one node, they earn only $B_r - c$. As a result, it is rational only for those players with small stakes to delegate all their resource power to others, and ND-m is partially satisfied.

Next, we consider NS-δ. As described above, a player with small stakes would not run multiple nodes, but instead would delegate their stakes to others. For a player with high stakes, this is divided into two cases: when weak identity management exists and when

[5] One can see that the percentage of resource power possessed by the decentralized pools is significantly small.

it does not. Weak identity management implies that nodes should reveal a pseudo-identity such as a public URL or a social ID. Firstly, in the latter case (DPoS-2), the player with high stakes can earn a higher profit by running multiple nodes because there is no Sybil cost. Therefore, a DPoS system without identity management partially satisfies NS-δ because only players with high stakes would run multiple nodes. Meanwhile, when the system (DPoS-1) includes weak identity management, voters can partially recognize whether different nodes are run by the same player. Therefore, the voters can avoid voting for these multiple nodes run by the same player because they may want to achieve good decentralization in the system. This means that it is not more profitable for one player to run multiple nodes than it is to run one node (i.e., Sybil costs exist), and these DPoS systems satisfy NS-δ. Note that because the identity management is not perfect, a rich player can still run multiple nodes by *creating multiple pseudo-identities.* Thus, strictly speaking, systems with weak identity management still do not fully satisfy NS-δ. However, because it is certainly more expensive for a rich player to run multiple nodes in DPoS-1 systems when compared to DPoS-2 systems, we mark such systems with ◐* for NS-δ in Tab. 2. Currently, EOS, TRON, Steem, and Steem Dollars have weak identity management (i.e., belong to DPoS-1).

Finally, we examine whether the DPoS system satisfies ED-(ε, δ). To this end, we consider two cases: when a delegate shares the block reward with voters (e.g., TRON [48] and Lisk [11]), and when they do not share (e.g., EOS[6]). In the former case, if a delegator receives V votes, the voters who voted for the delegator can, in general, earn reward $\frac{B_r}{V} - f$ per vote, where f represents a fee per vote paid to the delegator. Note that the larger V is, the smaller the reward is that the voters earn. Thus, when voters are biased towards a delegator, some voters can move their vote to other delegators for higher profits. In the latter case, delegators would increase their effective power by voting for themselves with more stakes to maintain or increase their ranking, and a more even power distribution among delegators would be achieved according to Thm. 4.3. Therefore, in the two cases, the power distribution among delegators can converge to an even distribution. However, the wealth gap between nodes obtaining small voting power and nodes obtaining high voting power would increase, thus implying that the probability of poor nodes generating blocks becomes smaller gradually. Consequently, the DPoS system partially satisfies ED-(ε, δ).

Tab. 2 presents the analysis result for the DPoS coins according to the four conditions. *DPoS systems may potentially ensure even power distribution among a limited number of players when weak identity management exists. However, the system has a limited number of players running nodes in the consensus protocol, which implies that they cannot have good decentralization.*

7 SUMMARY OF EMPIRICAL STUDY

We quantitatively analyze the data for PoW, PoS, and DPoS coins not only to establish the degree to which they are currently centralized, but also to validate four conditions. In this section, we describe the results for the most popular three coins each in PoW, PoS, and DPoS systems (see the full version [28] for the entire analysis result).

7.1 Methodology

We considered the past 10,000 blocks before Oct. 15, 2018, for PoW and PoS systems and the past 100,000 blocks before Oct. 15, 2018, for DPoS systems since some DPoS systems do not renew the list of block generators within 10,000 blocks. We parsed addresses of block generators from each blockchain explorer for 68 coins.

We determined the number NB_{A_i} of blocks generated by each address A_i, where the set of all addresses is denoted by $\mathcal{A}$. We then constructed a dataset $\mathcal{NB} = \left\{NB_{A_i} | A_i \in \mathcal{A}\right\}$ and rearranged $\mathcal{NB}$ and $\mathcal{A}$ in descending order of NB_{A_i}. Then, we analyzed the dataset using three metrics: the total number of addresses ($|\mathcal{A}|$), the Gini coefficient, and the entropy (H). Regarding the security in blockchain systems, it is meaningful to analyze not only how evenly the total power is distributed but also how evenly 50% and 33% of the power are distributed. Therefore, we also measure the level of decentralization for 50% and 33% power in the systems using the three metrics. To do this, we first define subset $\mathcal{A}^x$ of the address set $\mathcal{A}$, and subset $\mathcal{NB}^x$ of the data set $\mathcal{NB}$ as follows:

$$\mathcal{A}^x = \left\{A_i \in \mathcal{A} \,\middle|\, \frac{\sum_{j=1}^{i-1} NB_{A_i}}{\sum_{A_i \in \mathcal{A}} NB_{A_i}} < x\right\},$$
$$\mathcal{NB}^x = \{NB_{A_i} | A_i \in \mathcal{A}^x\},$$

where $0 \leq x \leq 1$. Here, note that if x is 0, the two sets are empty, and if x is 1, they are equal to $\mathcal{A}$ and $\mathcal{NB}$, respectively. The Gini coefficient and the entropy (H) are then defined as:

$$Gini(\mathcal{NB}^x) = \frac{\sum_{A_i, A_j \in \mathcal{A}^x} |NB_{A_i} - NB_{A_j}|}{2|\mathcal{A}| \sum_{A \in \mathcal{A}^x} NB_{A_i}},$$

$$H(\mathcal{NB}^x) = -\sum_{A_i \in \mathcal{A}^x} \frac{NB_{A_i}}{\sum_{A_i \in \mathcal{A}^x} NB_{A_i}} \log_2\left(\frac{NB_{A_i}}{\sum_{A_i \in \mathcal{A}^x} NB_{A_i}}\right).$$

If the deviation of $\mathcal{NB}^x$ is small, the Gini value is close to 0. Otherwise, the value is close to 1. The entropy depends on both $|\mathcal{A}^x|$ and the Gini coefficient. As $|\mathcal{A}^x|$ gets larger and the Gini coefficient gets smaller, the entropy gets larger. Therefore, entropy implicitly represents the level of decentralization, and large entropy implies a high level of decentralization. In fact, because a player can have multiple addresses, the measured values may not accurately represent the actual level of decentralization. However, since entropy is a concave function of the relative ratio of NB_{A_i} to the total number of generated blocks (i.e., $\frac{NB_{A_i}}{\sum_{A_i \in \mathcal{A}^x} NB_{A_i}}$), the results show an upper bound of the current level of decentralization. Therefore, if the measured values of entropy are low, the current systems do not have good decentralization.

7.2 Data Analysis

7.2.1 Quantitative analysis. Tab. 3 represents the results for the most popular three coins each in PoW, PoS, and DPoS systems.

Firstly, one can see that there is an insufficient number of block generators in PoW, PoS, and DPoS coins except for Qtum. In particular, $|\mathcal{A}^{\frac{1}{2}}|$ and $|\mathcal{A}^{\frac{1}{3}}|$ in PoW and PoS except for Qtum are quite small. The reason why Qtum has relatively many block generators is that it did not have staking pools *yet.* Note that this increases the number of block generators. However, we observe that there have been some requests for Qtum staking pools and intentions

[6] A debate exists as to whether delegates should share their rewards with voters or not [13, 29].

Table 3: Data analysis

Type	Coin name	100 %			50%			33%		
		$\lvert\mathcal{A}\rvert$	Gini	H	$\lvert\mathcal{A}^{\frac{1}{2}}\rvert$	$\text{Gini}^{\frac{1}{2}}$	$\text{H}^{\frac{1}{2}}$	$\lvert\mathcal{A}^{\frac{1}{3}}\rvert$	$\text{Gini}^{\frac{1}{3}}$	$\text{H}^{\frac{1}{3}}$
PoW	Bitcoin	62	0.8192	3.89	4	0.1143	1.98	3	0.1103	1.57
	Ethereum	65	0.8634	3.38	3	0.1402	1.53	2	0.0415	1.00
	Bitcoin Cash	15	0.5729	3.06	3	0.2572	1.51	2	0.0859	0.12
PoS	Tezos	245	0.8391	5.54	9	0.1061	3.13	6	0.1168	2.55
	Qtum	1853	0.7404	8.07	32	0.5923	4.12	7	0.2512	2.69
	Waves	110	0.8606	4.24	4	0.1545	1.93	3	0.1628	1.51
DPoS	EOS (21)	22	0.0447	4.43	11	0.0002	3.46	7	0.0003	2.81
	TRON (27)	28	0.0358	4.79	14	0.0009	3.81	9	0.0008	3.17
	Lisk (101)	101	0.0023	6.66	51	0.0011	5.67	34	0.0010	5.09

Table 4: Resource Power in DPoS Coins

Coin name	Delegates			100 %			50%			33%		
	$\lvert\mathcal{N}^{\text{D}}\rvert$	Gini^{D}	H^{D}	$\lvert\mathcal{N}\rvert$	Gini	H	$\lvert\mathcal{N}^{\frac{1}{2}}\rvert$	$\text{Gini}^{\frac{1}{2}}$	$\text{H}^{\frac{1}{2}}$	$\lvert\mathcal{N}^{\frac{1}{3}}\rvert$	$\text{Gini}^{\frac{1}{3}}$	$\text{H}^{\frac{1}{3}}$
EOS	21	0.048	4.39	439	0.846	6.47	28	0.063	4.80	18	0.047	4.16
TRON	27	0.198	4.54	165	0.849	4.84	12	0.258	3.29	6	0.324	2.23
Lisk	101	0.031	6.65	1179	0.907	6.99	52	0.013	5.70	35	0.011	5.13

to run a business for the pools [39–42]. Therefore, we expect that staking pools will become more popular, resulting in a decrease in the number of block generators. Indeed, Tezos and Waves, already allowing the delegation of stakes, have a smaller number of block generators. For DPoS systems, they have $|\mathcal{A}|$ similar to N_{dpos}, which is presented in parentheses in Tab. 3. In addition, $|\mathcal{A}^{\frac{1}{2}}|$ and $|\mathcal{A}^{\frac{1}{3}}|$ are close to $\frac{N_{\text{dpos}}}{2}$ and $\frac{N_{\text{dpos}}}{3}$, respectively. This indicates that only a small number of players have been block generators even though block generators are frequently elected, implying that the barriers to becoming a block generator are quite high.

Next, we describe the power distribution among nodes. As shown in Tab. 3, PoW and PoS coins certainly have a high value of the Gini coefficient, which implies that they have a significantly biased power distribution. Meanwhile, DPoS coins have a low Gini coefficient. This is because the elected block generators have the same opportunity to generate blocks in the DPoS systems.

In fact, results for DPoS coins in Tab. 3 does not present the resource power of the nodes, where the resource power indicates the number of stakes delegated to each node, because the probability of generating blocks is not proportional to the resource power in DPoS systems. Thus, to present the distribution of resource power among nodes, we analyze the instantaneous number of stakes delegated to each node through block explorers. Tab. 4 represents the distribution of stakes used to vote for nodes as of Nov. 19, 2018.

In Tab. 4, $|\mathcal{N}^{\text{x}}|$, Gini^{x}, and H^{x} represent the size of $\mathcal{N}^{\text{x}}$, Gini coefficient, and entropy for $\mathcal{N}^{\text{x}}$, respectively. The columns labeled *Delegates*, *100%*, *50%*, and *33%* provide information regarding the number of nodes, the Gini coefficient, and the entropy for the delegates ($\mathcal{N}^{D}$), and for the nodes whose total resource power is 100% ($\mathcal{N}$), 50% ($\mathcal{N}^{\frac{1}{2}}$), and 33% ($\mathcal{N}^{\frac{1}{3}}$), respectively. Gini^{D} is low for all DPoS systems, indicating that delegates possess similar resource power. In Section 6, we explained that DPoS systems can converge in probability to the state where delegates have similar resource power. Here, the reason Gini^{D} of TRON is relatively high compared to the others is that the node [51] operated by the TRON foundation is ranked in the first place by a relatively large margin. However, we observe that delegates, except for this node, possess almost the same resource power in TRON. Conversely, the value of Gini for all nodes is high, implying a large gap between the rich and the poor nodes. Moreover, Tab. 4 shows that the resource power is significantly biased toward the delegates.

As a result, the quantitative data analysis validates our theory and the analysis result of the incentive systems in Section 6.

7.2.2 Multiple nodes run by the same player. In DPoS systems that do not have weak identity management, a rich player can easily earn a higher profit by running multiple nodes. However, because they do not have any real identity management, it can be difficult to detect this rational behavior in practice. Nevertheless, in the full version [28], we describe that one player runs multiple nodes in several coins: GXChain, Ark, and Asch.

8 DISCUSSION

8.1 Debate on Incentive Systems

Recently, there was an interesting debate on the incentive system of Algorand [8, 18, 21]. Micali said that incentives are the hardest thing to do, and that existing incentivization has led to poor decentralization. Our study supports this notion by proving that it is impossible to design incentive systems for permissionless blockchains such that good decentralization is achieved.

Can we then create a permissionless blockchain to achieve good decentralization without any incentive system? The case where the incentive system does not exist is represented by $U_{n_i} = -c$, where c is the cost associated with running one node. This satisfies the second requirement of Def. 4.1 because NS-δ and ED-(ε, δ) are met as a trivial case. Meanwhile, the first two conditions, GR-m and ND-m, cannot be satisfied. As examples, we can consider BridgeCoin, IOTA, and Byteball, which do not have incentive systems and have difficulty in attracting the participation of many players. BridgeCoin has only one player, and IOTA is also controlled by just one player, the IOTA foundation [23, 24]. Byteball is another system that adopts DAG, and there are only four players. These examples show that blockchain systems with no incentive system cannot have a sufficient number of players.

However, our study considered only the incentives inside the system, and not incentives outside the system. Therefore, if there are some incentives that players can obtain outside the blockchain system, they can participate in the system. For example, IBM is a validator in Stellar, which does business using Stellar, and BrainBlocks [3] provides a payment platform related to Nano. This incentivizes IBM and BrainBlocks to participate in each system. Note that that fact does not ensure that these systems reach good decentralization. Indeed, both of these systems have poor decentralization [25, 35, 46]. In other words, they do not have a sufficient number of players and have a biased power distribution. Besides, through these cases, we can empirically see that organizations related to the coin system (e.g., the coin foundation or companies that do business with the coin) control the blockchain system, which may deviate from the philosophy of permissionless blockchains.

Note that we do not assert that blockchains without an incentive mechanism would always suffer from poor decentralization. Indeed, we can also find other peer-to-peer systems such as Tor and BitTorrent that attract many players without an incentive system. Of course, these systems are significantly different from a blockchain because they do not require resources such as computational power and stakes unlike a blockchain. In this paper, we do remain neutral on this debate.

8.2 Relaxation of Conditions from Consensus Protocol

We proved that an incentive system in permissionless blockchains cannot simultaneously satisfy the four conditions. Nevertheless, if there is a consensus protocol that relaxes part of the four conditions, we can expect to be able to design an incentive system such that good decentralization is achieved. However, it seems to be quite difficult to design such consensus protocols. In the full version [28], we explain the reason why the design of a consensus protocol relaxing the conditions is difficult by considering two methods of designing such protocols: 1) designing non-outsourceable puzzles and 2) finding non-delegable or non-divisible resources.

9 RELATED WORK

Attacks. Eyal et al. [16] proposed selfish mining, which an attacker possessing over 33% of the computing power can execute in PoW-based systems. They mentioned that this attack causes rational miners to join the attacker, eventually decreasing the level of decentralization. Eyal [14] and Kwon et al. [26] modeled a game between two pools. When considering block withholding attacks, the game is equivalent to *the prisoner's dilemma*, and the attacks cause rational miners to leave their mining pools, and instead, directly run nodes in a consensus protocol [14]. Contrary to this positive result, a fork after withholding attack between two pools leads to a pool-size game, where a larger pool can earn extra profits, and thus, the Bitcoin system can become more centralized. Furthermore, two existing works analyzed the Bitcoin system in a transaction-fee regime where transaction fees in block rewards are not negligible [6, 52]. They described that this regime incentivizes large miner coalitions and make a system more centralized.

Analysis. Many papers have already examined centralization in the Bitcoin system. First, Gervais et al. described centralization of the Bitcoin system in terms of various aspects such as services, mining, and incident resolution processes [20]. Miller et al. observed a topology in the Bitcoin network and found that approximately 2% of high-degree nodes acquire three quarters of the mining power [34]. Moreover, Feld et al. analyzed the Bitcoin network, focusing on its autonomous systems (ASes), and showed that routable peers are concentrated only in a few ASes [17]. Recently, Gencer et al. analyzed the Bitcoin and Ethereum systems from the perspective of decentralization [19]. Kwon et al. analyzed a game in which two PoW coins with compatible mining algorithms exist [27]. They showed that fickle mining behavior between two coins can reduce the decentralization level of the lower-valued one of the two coins. In addition, Kim et al. analyzed the Stellar system and concluded that the system is significantly centralized [25].

Solutions. There are several works that address the issue of poor decentralization in blockchains. Many works [15, 32, 33, 53] have proposed non-outsourceable puzzles to prevent mining pools from being popular. However, they cannot fully prevent the delegation. As another solution, Luu et al. proposed an efficient decentralized mining pool, SMARTPOOL, where individual miners who directly run nodes in the consensus protocol can consistently earn profits [31]. However, this still does not incentivize players to run nodes directly (see Section 6). Another work [1] proposed a proof-of-human-work requiring labor from players with CAPTCHA as a human-work puzzle. As mentioned by [1], although the gap among labor abilities of people is relatively small by nature, rich players can hire more workers to solve more puzzles. Lastly, we are aware of a recent paper [4] in which the authors addressed a similar problem to our paper. Brünjes et al. proposed a reward scheme, which causes a system to reach a state where k staking pools with similar resource power exist. They assumed our third condition, NS-δ (i.e., all players can run only one node), and thus, it seems difficult for their incentive system to achieve good decentralization in practice. As described in previous sections, there is an incentive system that satisfies only GR-m, ND-m, and ED-(ε, δ).

10 CONCLUSION AND DIRECTION

Developers are facing difficulties in designing blockchain systems to achieve good decentralization. Our study answers the question of why it is significantly difficult to design a system that achieves good decentralization, by proving that the achievement of good decentralization in the consensus protocol and non-reliance on a TTP contradict each other. More specifically, we prove that when the ratio between the resource power of the poorest and richest players is close to 0, the upper bound of the probability that systems without a Sybil cost will achieve full decentralization is close to 0. This result indicates that if we cannot narrow the gap between the rich and the poor in the real world or assign a Sybil cost without relying on a TTP, a high level of decentralization in systems will not occur forever with a high probability. Furthermore, through the protocol and data analysis, we observed the phenomena consistent with our theory. From our result, we propose one direction to achieve good decentralization of the system; developing a method that can assign Sybil costs without relying on a TTP in blockchains.

REFERENCES

[1] Jeremiah Blocki and Hong-Sheng Zhou. 2016. Designing proof of human-work puzzles for cryptocurrency and beyond. In *Theory of Cryptography Conference*. Springer, 517–546.

[2] Joseph Bonneau, Andrew Miller, Jeremy Clark, Arvind Narayanan, Joshua A Kroll, and Edward W Felten. 2015. Sok: Research perspectives and challenges for bitcoin and cryptocurrencies. In *Security and Privacy (SP), 2015 IEEE Symposium on*. IEEE.

[3] Brainblocks 2019. Free and instant NANO payments. https://brainblocks.io/. (2019). [Online; accessed 1-May-2019].

[4] Lars Brünjes, Aggelos Kiayias, Elias Koutsoupias, and Aikaterini-Panagiota Stouka. 2018. Reward sharing schemes for stake pools. *arXiv preprint arXiv:1807.11218* (2018).

[5] Vitalik Buterin and Glen Weyl. 2018. Liberation Through Radical Decentralization. https://medium.com/@VitalikButerin/liberation-through-radical-decentralization-22fc4bedc2ac. (2018). [Online; accessed 23-Nov-2018].

[6] Miles Carlsten, Harry Kalodner, S Matthew Weinberg, and Arvind Narayanan. 2016. On the instability of bitcoin without the block reward. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM.

[7] Miguel Castro, Barbara Liskov, et al. 1999. Practical Byzantine fault tolerance. In *OSDI*, Vol. 99.

[8] Debate 2017. No Incentive? Algorand Blockchain Sparks Debate at Cryptography Event. https://www.google.com/amp/s/www.coindesk.com/

no-incentive-algorand-blockchain-sparks-debate-cryptography-event/amp/. (2017). [Online; accessed 11-Nov-2018].

[9] Dmitrii Zhelezov. 2018. PoW, PoS and DAGs are NOT consensus protocols. https://medium.com/coinmonks/a-primer-on-blockchain-design-89605b287a5a. (2018). [Online; accessed 28-Mar-2019].

[10] Double spending 2018. Bitcoin Gold suffers double spend attacks, $17.5 million lost. https://www.zdnet.com/article/bitcoin-gold-hit-with-double-spend-attacks-18-million-lost/. (2018). [Online; accessed 14-Nov-2018].

[11] Earnlisk 2018. EARN LISK. https://earnlisk.com/. (2018). [Online; accessed 26-Nov-2018].

[12] EOS incentive 2018. How Reward Distribution in EOSIO Works. https://blog.springrole.com/how-reward-distribution-in-eosio-works-936e292dfbab. (2018). [Online; accessed 18-Nov-2018].

[13] Eosnewyork 2018. Thomas Cox of Block.One Confirms Vote-Buying Will Be Against EOS.IO Proposed Constitution. https://steemit.com/eos/@eosnewyork/block-one-confirms-vote-buying-will-be-against-eos-io-proposed-constitution. (2018). [Online; accessed 26-Nov-2018].

[14] Ittay Eyal. 2015. The Miner's Dilemma. In *Symposium on Security and Privacy*. IEEE.

[15] Ittay Eyal and Emin Gün Sirer. 2014. How to Disincentivize Large Bitcoin Mining Pools. http://hackingdistributed.com/2014/06/18/how-to-disincentivize-large-bitcoin-mining-pools/. (2014). [Online; accessed 2-Nov-2018].

[16] Ittay Eyal and Emin Gün Sirer. 2014. Majority Is Not Enough: Bitcoin Mining Is Vulnerable. In *International Conference on Financial Cryptography and Data Security*. Springer.

[17] Sebastian Feld, Mirco Schönfeld, and Martin Werner. 2014. Analyzing the Deployment of Bitcoin's P2P Network under an AS-level Perspective. *Procedia Computer Science* 32 (2014), 1121–1126.

[18] Alexis Gauba and Zubin Koticha. [n. d.]. The Need for an Incentive Scheme in Algorand. https://blockchainatberkeley.blog/the-need-for-an-incentive-scheme-in-algorand-6fe9db45f2a7. ([n. d.]). [Online; accessed 11-Nov-2018].

[19] Adem Efe Gencer, Soumya Basu, Ittay Eyal, Robbert van Renesse, and Emin Gün Sirer. 2018. Decentralization in Bitcoin and Ethereum Networks. (2018).

[20] Arthur Gervais, Ghassan O Karame, Vedran Capkun, and Srdjan Capkun. 2014. Is Bitcoin a Decentralized Currency? *IEEE security & privacy* 12, 3 (2014), 54–60.

[21] Yossi Gilad, Rotem Hemo, Silvio Micali, Georgios Vlachos, and Nickolai Zeldovich. 2017. Algorand: Scaling Byzantine Agreements for Cryptocurrencies. In *Proceedings of the 26th Symposium on Operating Systems Principles*. ACM.

[22] Inequality 2018. Global Inequality. https://inequality.org/facts/global-inequality/. (2018). [Online; accessed 12-Oct-2018].

[23] Iota-centralization 2017. IOTA is centralized. https://medium.com/@ercwl/iota-is-centralized-6289246e7b4d. (2017). [Online; accessed 11-Nov-2018].

[24] Iota-milestone 2018. The Tangle: an illustrated introductionl. https://blog.iota.org/the-tangle-an-illustrated-introduction-79f537b0a455. (2018). [Online; accessed 11-Nov-2018].

[25] Minjeong Kim, Yujin Kwon, and Yongdae Kim. 2019. Is Stellar As Secure As You Think? *arXiv preprint arXiv:1904.13302* (2019).

[26] Yujin Kwon, Dohyun Kim, Yunmok Son, Eugene Vasserman, and Yongdae Kim. 2017. Be Selfish and Avoid Dilemmas: Fork After Withholding (FAW) Attacks on Bitcoin. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. ACM.

[27] Yujin Kwon, Hyoungshick Kim, Jinwoo Shin, and Yongdae Kim. 2019. Bitcoin vs. Bitcoin Cash: Coexistence or Downfall of Bitcoin Cash? *arXiv preprint arXiv:1902.11064* (2019).

[28] Yujin Kwon, Jian Liu, Minjeong Kim, Dawn Song, and Yongdae Kim. 2019. Impossibility of Full Decentralization in Permissionless Blockchains. *arXiv preprint arXiv:1905.05158* (2019).

[29] Daniel Larimer. 2018. Proposal for EOS Resource Renting & Rent Distribution. https://medium.com/@bytemaster/proposal-for-eos-resource-renting-rent-distribution-9afe8fb3883a. (2018). [Online; accessed 26-Nov-2018].

[30] Lpos 2018. Waves Docs. https://docs.wavesplatform.com/en/platform-features/leased-proof-of-stake-lpos.html. (2018). [Online; accessed 28-Oct-2018].

[31] Loi Luu, Yaron Velner, Jason Teutsch, and Prateek Saxena. 2017. SMART POOL: Practical Decentralized Pooled Mining. 2017 (2017).

[32] Andrew Miller, Ari Juels, Elaine Shi, Bryan Parno, and Jonathan Katz. 2014. Permacoin: Repurposing Bitcoin Work for Data Preservation. In *2014 IEEE Symposium on Security and Privacy (SP)*. IEEE, 475–490.

[33] Andrew Miller, Ahmed Kosba, Jonathan Katz, and Elaine Shi. 2015. Nonoutsourceable Scratch-Off Puzzles to Discourage Bitcoin Mining Coalitions. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. ACM, 680–691.

[34] Andrew Miller, James Litton, Andrew Pachulski, Neal Gupta, Dave Levin, Neil Spring, and Bobby Bhattacharjee. 2015. Discovering bitcoin's public topology and influential nodes. *et al.* (2015).

[35] Mystellar.tools 2018. mystellar.tools. https://mystellar.tools/explorer/network/. (2018). [Online; accessed 15-Nov-2018].

[36] Satoshi Nakamoto. 2008. Bitcoin: A peer-to-peer electronic cash system. (2008).

[37] P2Pool 2019. DECENTRALIZED BITCOIN MINING POOL. http://p2pool.org/. (2019). [Online; accessed 18-Apr-2019].

[38] PoS pool 2018. SIMPLE POS POOL. https://simplepospool.com/#. (2018). [Online; accessed 28-Oct-2018].

[39] PoS request 2017. QTUM Staking Pools. https://cryptopanic.com/news/43154/QTUM-Staking-Pools. (2017). [Online; accessed 20-Nov-2018].

[40] PoS request 2017. Smart contract and stake delegation. https://forum.qtum.org/topic/229/smart-contract-and-stake-delegation. (2017). [Online; accessed 20-Nov-2018].

[41] Qtum pool 2018. Proof of Stake. https://www.poolofstake.io/wp-content/uploads/2018/09/Pool_of_Stake_whitepaper.pdf. (2018). [Online; accessed 20-Nov-2018].

[42] Qtum pool 2018. Staking pool and QTUM MVP. https://bitcointalk.org/index.php?topic=4391854.0. (2018). [Online; accessed 20-Nov-2018].

[43] Douglas Sikorski. 2015. The Rich-Poor Gap: A Synopsis. (2015).

[44] Slush 2018. SLUSHPOOL. https://slushpool.com/stats/?c=btc. (2018). [Online; accessed 27-Oct-2018].

[45] Stakeminers 2018. STAKEMINERS.COM. https://stakeminers.com/index.php. (2018). [Online; accessed 19-Nov-2018].

[46] Stellarbeat.io 2018. stellarbeat.io. https://stellarbeat.io/. (2018). [Online; accessed 15-Nov-2018].

[47] Chad Stone, Danilo Trisi, Arloc Sherman, and Brandon Debot. 2015. A guide to statistics on historical trends in income inequality. *Center on Budget and Policy Priorities* 26 (2015).

[48] Tokengoodies 2018. TRX VOTING REWARDS CALCULATOR. https://www.tokengoodies.com/. (2018). [Online; accessed 26-Nov-2018].

[49] Top 100 coins 2018. TOP 100 Cryptocurrencies By Market Capitalization. https://coinmarketcap.com/coins/. (2018). [Online; accessed 11-Sep-2018].

[50] Transaction fees 2018. Big transaction fees are a problem for bitcoin — but there could be a solution. https://www.cnbc.com/2017/12/19/big-transactions-fees-are-a-problem-for-bitcoin.html. (2018). [Online; accessed 13-Nov-2018].

[51] TRON node 2018. Sesameseed. https://www.sesameseed.org/. (2018). [Online; accessed 20-Nov-2018].

[52] Itay Tsabary and Ittay Eyal. 2018. The Gap Game. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. ACM.

[53] Gongxian Zeng, Siu Ming Yiu, Jun Zhang, Hiroki Kuzuno, and Man Ho Au. 2017. A Nonoutsourceable Puzzle Under GHOST Rule. In *2017 15th Annual Conference on Privacy, Security and Trust (PST)*. IEEE, 35–358.