10-1-2019

Bitcoin: An Impossibility Theorem for Proof-of-Work based Protocols

Jacob Leshno

Philipp Strack

BITCOIN: AN IMPOSSIBILITY THEOREM FOR PROOF-OF-WORK BASED PROTOCOLS

By

Jacob Leshno and Philipp Strack

February 2019

COWLES FOUNDATION DISCUSSION PAPER NO. 2204

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Bitcoin: An Impossibility Theorem for Proof-of-Work based Protocols

Jacob Leshno & Philipp Strack

February 11, 2019*

Abstract

A key part of decentralized consensus protocols is a procedure for random selection, which is the source of the majority of miners cost and wasteful energy consumption in Bitcoin. We provide a simple economic model for random selection mechanism and show that *any* PoW protocol with natural desirable properties is outcome equivalent to the random selection mechanism used in Bitcoin.

*This paper was submitted on February 14, 2019 to the EC'19 conference (ACM Conference on Economics and Computation).

1 Introduction

A key part of decentralized consensus protocols is a procedure for random selection. In Bitcoin (Nakamoto, 2008), the ledger is periodically update by a randomly selected server in network. In fact, the random selection is so central to the Bitcoin protocol that the servers that maintain Bitcoin are called miners because a server who updates the ledger is said to have "mined a block" (a block is a batch of transaction data).

One of Bitcoin's central innovations is a method for verifiably selecting a random miner in a decentralized manner. This entrails two central restrictions: (i) the system cannot rely on a trusted randomization device, and (ii) any computer can join the network, implying that miner are not identifiable. The random selection is achieved through the use of cryptography as follows: Each miner assembles a block of transaction data, which includes a free-set field called nonce. The block can be added to the ledger if applying a cryptographic hash function to the block yields a value that is below a difficulty threshold. A miner that finds such a block is said to have mined a block. Miners are rewarded financially when mining a block, and therefore compete to mine blocks.

Under the assumption that the cryptographic hash function is irreversible, each selection of a nonce yields a mined block with a fixed probability. A miner's probability of mining the next block is proportional to the number of nonces attempted. Since each attempt requires calculation of the hash function, the number of attempts is in turn proportional to the miner's share of the total computational power of the Bitcoin network. The difficulty threshold adjusts to keep the overall average block mining rate constant. Bitcoin and similar systems are often called called Proof-of-Work (PoW) protocols, as each miner increases his probability of being selected by executing costly computations.

Bitcoin's PoW protocol have been successful in establishing a reliable system, but there has been much interest in replacing it. One reason is the high monetary and environmental cost of wasteful computation. The work done by miners serves no purpose other than providing a random selection. As the popularity and value of Bitcoin increased, more miners compete for the rewards. Currently the total electricity used by miners exceeds that of some small countries, and the majority of this consumption is spent on computing the hash in attempt to mine a block.

This note analyzes random selection as a game theoretic problem. This allows us to abstract away from the particular protocol used in Bitcoin and consider general proof-of-work protocols. We define three desirable properties of PoW protocols: anonymity, robustness to Sybil attacks and robustness to merging. Anonymity states that the protocol discriminates between miners only based on the computations they performed. Robustness to Sybil attacks ensures that it is never beneficial for a miner to pretend to be a number of different miners. Robustness to merger states that no group of miners can increase their joint probability of mining a block by joining forces and becoming a single miner without performing more computations.

Our main result shows that any PoW protocol with these properties is outcome equivalent to

the random selection used in Bitcoin. More precisely, each miner is selected with a probability that is proportional to the number of computations she performed. This result is driven only by game-theoretic constraints and not by constraints on computations underlying the protocol. Thus, in order to improve upon the Bitcoin PoW mechanism it is necessary to use a different form of decentralization that violates on of the properties we introduced.

The result has drastic implications for the design of decentralized cryptocurrencies and the plethora of alternative blockchain protocols that propose different cryptographic methods to improve on the PoW mechanism of Bitcoin.[1] Such improvements cannot be obtained only from a change in the cryptographic method while maintaining anonymity, robustness to Sybil attacks and merging. In order for alternatives such as Proof-of-Stake to provide better performance these must be able to identify miners (violate anonymity), or restrict the entry of unidentified miners (which allows the protocol to violate robustness to Sybil attacks), or provide the miners with incentives to merge and therefore limit the decentralization of the system.

This note is structured as follows: Section 2 defines a random selection mechanisms based on investment levels (i.e. the number of computations performed by each miner) and provides a characterization of all random selection mechanisms that are anonymous and robust to Sybil attakcs and merging. Section 3 makes the connection between random selection based on computational tasks and our definition of random selection mechanisms. We conclude in Section 4 and comment on how Proof-of-Stake can allow other forms of random selection by violating our axioms.

2 Random Selection Mechanisms

Denote by Δ^n a random selection from the set of n agents $\Delta^n = \left\{z \in \mathbb{R}^n_+ \colon \sum_{i=1}^n z_i = 1\right\}$. We furthermore denote by $N = \{1, \ldots, n\}$ the set of agents and by i a typical agent. We begin by introducing the main object of interest in our study – the random selection mechanism:

Definition 1 (Random Selection Mechanism)**.** *A random selection mechanism p is described by a family of functions $p^n : \mathbb{R}^n_+ \to \Delta^n$ indexed by $n \in \mathbb{N}$ such that the probability with which agent $i \in N$ is selected at the investment profile $x = (x_1, \ldots, x_n)$ equals*

$$p^n_i(x_1, \ldots, x_n),$$

which is non-decreasing in x_i.

In the context of Bitcoin, random selection is achieved by picking the agent who is first able to produce a block with a sufficiently low hash value. The number of costly computations x_i agent i is willing to invest in mining this block is mapped to a probability of being the first to compute

[1] As of February 2019, Wikipedia lists more than 20 different PoW cryptocurrencies, see table in https://en.wikipedia.org/wiki/Proof-of-work_system.

such a hash and thus being selected. We will explain in Section 3 in greater detail how block mining in Bitcoin and other proof of work based crypto-currencies are a special case of a random selection mechanism. Conceptually, the abstraction to selection mechanisms is useful as it allows us to analyze the implication of economic constraints on decentralized systems like Bitcoin without the need to model the cryptographic and computational details.

The first constraint we impose is anonymity. It states that every agent is treated the same by the mechanism, i.e. if two agents exchange their identities they are still treated the same. For example in the Bitcoin protocol all miners are treated the same, i.e. they all face the same requirement to be selected to mine the next block.

Axiom 1 (Anonymity)**.** *A selection mechanism is anonymous if it is invariant under permutations, i.e. for every n and every permutation $\pi : \mathbb{R}^n_+ \to \mathbb{R}^n_+$ it satisfies $\pi(p^n(x)) = p^n(\pi(x))$.*

For notational simplicity we will state our other axioms for anonymous mechanisms.

Axiom 2 (Robustness to Sybil Attacks)**.** *An anonymous random selection mechanism is robust to Sybil attacks if for every $x \in \mathbb{R}^n_+$ and every $y \in \mathbb{R}^k_+$ with $\sum_{j=1}^k y_j = x_1$*

$$p_1^n(x_1, \ldots, x_n) \geq \sum_{j=1}^{k} p_j^{n+k-1}(y_1, \ldots, y_k, x_2, x_3, \ldots, x_n).$$

Thus, robustness to Sybil attacks states that the selection probability of agent 1 in a situation with n agents is weakly less than the sum of the selection probabilities of agent 1 and 2 in an $n+1$ agent situation where agent 1's total investments are split between the first two agents. Intuitively, this restriction ensures that no agent can pretend to be a group of different agents and increase his winning probability without investing more.

Axiom 3 (Robust to Centralization)**.** *An anonymous random selection mechanism is robust to centralization if for every $x \in \mathbb{R}^n_+$ and every $y \in \mathbb{R}^k_+$ with $\sum_{j=1}^k y_j = x_1$*

$$p_1^n(x_1, \ldots, x_n) \leq \sum_{j=1}^{k} p_j^{n+k-1}(y_1, \ldots, y_k, x_2, x_3, \ldots, x_n).$$

Robustness to centralization imposes the complementary requirement to robustness to Sybil attacks. No group of agents can merge and increase their joint winning probability without investing more. A mechanism which is not robust to centralization will, by definition, provide some agents with incentives to merge as agents with larger investments have a relatively higher winning probability. This might lead such a selection mechanism to be, in the long-run, controlled by relatively few agents which in some context (like cryptocurrencies) is non-desirable as it increases the risk of attacks and manipulation.

Proposition 1. *A random selection mechanism p is anonymous, robust to Sybil attacks, and robust to centralization if and only if*

$$p_i^n(x) = \frac{x_i}{\sum_{j=1}^n x_j}\,. \tag{1}$$

Proof. We begin by showing the axioms imply the functional form (1) in the case where the investment of each agent is rational $x \in \mathbb{Q}_+^n$. Consider an arbitrary vector of investments $x \in \mathbb{Q}_+^n$ and w.l.o.g assume that all investments are expressed with respect to a common denominator $b \in \mathbb{N}$, i.e. there exitst $a \in \mathbb{N}^n$ such that $x_i = {}^{a_i}\!/_{b}$. We begin by splitting the first agent into a_1 agents each of which makes an investment of ${}^{1}\!/_{b}$. As a consequence of the robustness to Sybil attacks and centralization it follows that the joint winning probability of the first a_1 agents after this split equals the original winning probability of the first agent

$$p_1^n(x) = \sum_{j=1}^{a_1} p_j^{n+a_1-1}\left(\frac{1}{b},\ldots,\frac{1}{b},\frac{a_2}{b},\ldots,\frac{a_n}{b}\right)\,.$$

In the next step we merge the last $n-1$ agents into a single agent. Again by the robustness to Sybil attacks and centralization the winning probability of the last agent in the new situation equals the joint winning probability of the last $n-1$ agents in the old situation. As the winning probabilities sum up to 1 the joint winning probability of the first a_1-1 agents remains unaffected and we have that

$$p_1^n(x) = \sum_{j=1}^{a_1} p_j^{a_1+1}\left(\frac{1}{b},\ldots,\frac{1}{b},\frac{\sum_{i=2}^n a_i}{b}\right)\,.$$

In the next step we split the $a+1$ agent into $\sum_{i=2}^n a_i$ agents each investing $\frac{1}{b}$. Again, by the robustness to Sybil attacks and centralization this implies that

$$p_1^n(x) = \sum_{j=1}^{a_1} p_j^{|a|}\left(\frac{1}{b},\ldots,\frac{1}{b}\right)\,,$$

where $|a| = \sum_{i=1}^n a_i$. It follows from anonymity that each of the agents wins with equal probability of ${}^{1}\!/_{|a|}$, and thus

$$p_1^n(x) = \frac{a_1}{|a|} = \frac{a_1/b}{|a|/b} = \frac{x_1}{\sum_{j=1}^n x_j}\,.$$

To extend this result from $\mathbb{Q}_+^n$ to $\mathbb{R}_+^n$ we first show that the result extends to vectors where the first coordinate is chosen from $\mathbb{R}_+$ instead of $\mathbb{Q}_+$. Consider an arbitrary $x_{-1} \in \mathbb{Q}_+^{n-1}$ and $x_1 \in \mathbb{R}_+$. Choose two sequences $w^r, v^r \in \mathbb{Q}_+$ such that w^r converges to x_1 from above and v^r converges to x_1 from below when $r \to \infty$. By monotonicity we have that

$$\frac{v^r}{v^r + \sum_{j=2}^n x_j} \leq p_1^n(x_1, x_{-1}) \leq \frac{w^r}{w^r + \sum_{j=2}^n x_j}\,.$$

As the lower bound and the upper bound converge to the same limit it follows that $p_1^n(x) = \frac{x_1}{|x|}$ for all x with $x_1 \in \mathbb{R}_+$ and $x_{-1} \in \mathbb{Q}_+^{n-1}$. By anonymity $p_2^2(x) = \frac{x_2}{|x|}$ for all x with $x_{-2} \in \mathbb{Q}_+^{n-1}$ and $x_2 \in \mathbb{R}_+$. Thus, for $x_{-2} \in \mathbb{Q}_+^{n-1}$ and $x_2 \in \mathbb{R}_+$ we have that

$$p_1^n(x) = 1 - p_2^n(x) - \sum_{k=3^n} p_k^n(x) = 1 - \frac{x_2}{|x|} - \sum_{k=3}^{n} \frac{x_2}{|x|} = \frac{x_1}{|x|}\,.$$

Applying the above argument with an upper and lower bound again yields that for all x with $(x_1, x_2) \in \mathbb{R}_+^2$ and $(x_3, \ldots, x_n) \in \mathbb{Q}_+^{n-2}$ we have that $p_1^n(x) = \frac{x_1}{|x|}$. Applying the same argument sequentially for each agent $k \geq 3$ yields that $p_1^n(x) = \frac{x_1}{|x|}$ for all $x \in \mathbb{R}_+^n$. By permuting the role of agent 1 and agent i and anonymity we have that $p_i^n(x) = \frac{x_i}{\sum_{j=1}^n x_j}$ for all $i \in \{1, .., n\}$ and all $x \in \mathbb{R}_+^n$.

We are left to verify that the functional form (1) satisfies our assumptions. Clearly (1) is monotonic and anonymous. Furthermore, we have that for every $y \in \mathbb{R}_+^k$ with $|y| = x_1$

$$\sum_{j=1}^{k} p_j^{n+k-1}\big(y_1, \ldots, y_k, x_2, x_3, \ldots, x_n\big) = \sum_{j=1}^{k} \frac{y_j}{|y| + \sum_{i=2}^n x_i} = \frac{x_1}{\sum_{i=1}^n x_i} = p_1^n(x)\,,$$

which shows that the functional form (1) is robust to Sybil attacks and centralization and completes the proof. □

Equation 1 states that the probability with which an agent is selected is proportional to the share of computations she performed. For example, it describes the probability that a miner is selected to mine the next block in Bitcoin: Miners attempt to mine the next Bitcoin block once the previous block was published (we abstract from some technical details and assume blocks are transmitted instantaneously to all miners) by attempting different values of a nonce and computing their hashes. Under common cryptographic assumptions, no miner can do better than guess a random nonce and each nonce entails the same probability of being selected (to mine the next block). Thus, the probability with which an agent is selected in the Bitcoin protocol equals the number of hashes she computed relative to the total number of hashes computed before the next block is mined.

The proof of Proposition 1 shows that the monotonicity of the selection mechanism is not necessary if one restricts attention to the case where investments are rational numbers. In any practical application where quantities invested can be finitely encoded the restriction to rational number is vacuously satisfied and thus the monotonicity assumption plays no role.

3 Mining in the Bitcoin and other Proof of Work Protocols

This section establishes the link between random selection mechanisms (analyzed in Section 2) and cryptographic protocols for randomly selecting an agent based upon the computations performed by each agent, i.e. proof of work (PoW) based protocols. To ease the exposition we sometimes follow the language commonly used in the context of Bitcoin and refer to agents as "miners", and being selected as "mining a block". Throughout we focus on the economic incentives — cost and benefits — of mining a block intentionally abstracting away from many computational and cryptographic details.

Consider a situation where n agents (miners), indexed by $i \in N$, compete to be selected to mine the next block. Each agent assigns a value of 1 to being selected to mine the next block.[2] Denote by S the set of strategies available to each miner. A strategy $s_i \in S$ describes a complete contingent plan of what the miner will do until the next block is mined. For example, which computations miner i will perform and which hardware she will use to perform them and so on. We denote by

$$\gamma_i \, c(s_i) \geq 0$$

the expected cost miner i incurs when using the strategy s_i. The cost could be energy cost associated with the computations performed according to s_i, but also the cost of renting computational power from a cloud service such as Amazon AWS. Through the parameter $\gamma_i > 0$ we explicitly allow the miners to have different costs to account for the fact that they might have access to different hardware and might face different energy prices.

Consider a family of functions $\phi^n : S^n \to \Delta^n$ such that for each $n \in \mathbb{N}$ and each vector of strategies $s \in S^n$ the probability with which miner i is selected equals

$$\phi_i^n(s_1, \ldots, s_n) \, .$$

We assume that there exists a recommended strategy $\sigma : \mathbb{R}_+ \to S$ that recommends for each budget of computations x an agent is willing to perform on the next block a strategy that is strictly optimal independent of the strategies used by other agents, i.e. for all $x_i \in \mathbb{R}_+, i \in N, s_{-i} \in S^{n-1}$ and $s_i \in S$ such that $c(s_i) \leq x_i$

$$\phi(\sigma(x), s_{-i}) > \phi(s_i, s_{-i}) \, .$$

Definition 2 (Proof of Work Protocol)**.** *We call a tuple* (S, c, f, σ) *a Proof of Work protocol.*

A few properties of the PoW protocol are important for maintaining a reliable decentralized system. To prevent dependency on other systems, the PoW selection should not rely on an external source of identity verification. To maintain decentralization, the PoW should allow any potential

[2] This is without loss of generality as we can rescale the agents cost and benefits by dividing through the value this agent assigns to mining a block.

miner to be able to enter and participate in the random selection. In particular, new miners should be free to join, and small miners or new miners should not be at a disadvantage. These motivate the following axioms, which are the counterparts of the axioms of Section 2.

As stated above, in a decentralized PoW protocol miners are anonymous and there is no registry of miners identities. The selection function ϕ can distinguish between players only through the results of their computation, which are fully determined by their strategies. Therefore, ϕ depends only on the strategies chosen by players and not their identities:

Axiom 4 (Anonymity)**.** *A PoW protocol is anonymous if ϕ is invariant under permutations, i.e. for every n and every permutation $\pi : \mathbb{R}_+^n \to \mathbb{R}_+^n$ it satisfies $\pi(\phi^n(x)) = \phi^n(\pi(x))$.*

The lack of identifiable identities also implies that the selection mechanism cannot know whether multiple players are controlled by a single entity. Allowing any potential participant to join without authentication allows existing players to participate under many different identities, and potentially engage in Sybil attacks. We therefore ask that the PoW protocol is robust to Sybil attacks:

Axiom 5 (Robustness to Sybil Attacks)**.** *An anonymous PoW protocol is robust to Sybil attacks if for every $s \in S^n$ and every $\tilde{s} \in S^k$ with $\sum_{j=1}^k c(\tilde{s}_j) = c(s_1)$*

$$\phi_1^n\big(s_1, \ldots, s_n\big) \geq \sum_{j=1}^{k} \phi_j^{n+k-1}\big(\tilde{s}_1, \ldots, \tilde{s}_k, s_2, s_3, \ldots, s_n\big)\,.$$

The security of the blockchain can be jeopardized when a single miner controls a large fraction of the computational power in the network. Nakamoto (2008) argues that the Bitcoin is secure as long as no miners holds more than half of the mining power in the network (a miner with more than half of the total mining power can reverse transactions). We therefore ask that the PoW protocol does not create incentives for miner consolidation.

Axiom 6 (Robust to Centralization)**.** *An anonymous PoW protocol is robust to centralization if for every $s \in S^n$ and every $\tilde{s} \in S^k$ with $\sum_{j=1}^k c(\tilde{s}_j) = c(s_1)$*

$$\phi_1^n\big(s_1, \ldots, s_n\big) \leq \sum_{j=1}^{k} \phi_j^{n+k-1}\big(\tilde{s}_1, \ldots, \tilde{s}_k, s_2, s_3, \ldots, s_n\big)\,.$$

The above axioms describe properties of the PoW protocol. Next, we describe how miners will behave in such a protocol. An *equilibrium* of the game played between n miners is a strategy profile $s \in S^n$ such that no miner i can benefit from deviating to another strategy $s_i' \in S$

$$\phi_i^n(s_i, s_{-i}) - \gamma_i\, c(s_i) \geq \phi_i^n(s_i', s_{-i}) - \gamma_i\, c(s_i')\,.$$

Our main theorem below shows that, maybe surprisingly, our previous axioms are enough to

uniquely pin down the winning probability of each miner *only* as a function of the number of computations each miner performed:

Theorem 1. *Consider a PoW protocol that is anonymous, robust to Sybil attacks and centralization then in any equilibrium $s = (s_1, \ldots, s_n)$ miner i mines the next block with probability*

$$\frac{c(s_i)}{\sum_{j=1}^{n} c(s_j)}\,.$$

Proof. Fix an anonymous PoW protocol (S, c, f, σ) that is robust to Sybil attacks and centralization. For every n define $p^n : \mathbb{R}_+^n \to \Delta^n$ by $p^n(x) = \phi(\sigma(x_1), \ldots, \sigma(x_2))$. Because a larger x_i allows $\sigma(x_i)$ to select among more strategies, $p(x)$ is non-decreasing in x_i and thus p is a selection mechanism according to Definition 1. This mechanism, it is anonymous as ϕ is invariant. Furthermore, as the PoW protocol is robust to Sybil attacks we have that for every $x \in \mathbb{R}_+^n$ and every $y \in \mathbb{R}_+^k$ with $\sum_{j=1}^{k} y_j = x_1$

$$\begin{aligned} p_1^n(x) = \phi_1^n(\sigma(x_1), \ldots, \sigma_n(x_n)) &\geq \sum_{j=1}^{k} \phi_j^{n+k-1}\big(\sigma(y_1), \ldots, \sigma(y_k), \sigma(x_2), \sigma(x_3), \ldots, \sigma(x_n)\big) \\ &= \sum_{j=1}^{k} p_j^{n+k-1}\big(y_1, \ldots, y_k, x_2, x_3, \ldots, x_n\big)\,. \end{aligned}$$

Thus, the selection mechanism p is robust to Sybil attacks. The same argument establishes that p is robust to centralization. Thus, by Proposition 1 we have that

$$\phi_i^n(\sigma(x_1), \ldots, \sigma(x_n)) = p_i^n(x) = \frac{x_i}{\sum_{j=1}^{n} x_j}\,. \tag{2}$$

Now, consider an equilibrium $s \in S^n$, by the strict optimality of σ it follows that $s_i = \sigma(c(s_i))$. Plugging into (2) yields that $\frac{c(s_i)}{\sum_{j=1}^{n} c(s_j)}$ and completes the proof. □

This result carries a few implications. First, our results characterize the random selection mechanism regardless of the computational tasks and the miners strategy space. This implies that a different form of competition between miners cannot arise from a different specification of the computational tasks (for example using a different hash function). Other PoW protocols yield the same economic competition and computational expenditure. For example, the exactly same selection mechanism arises in a PoW protocol where the system has access to synchronized clocks and selects the miner that produces the lowest hash within a prespecified time frame (e.g. every 10 minutes). Our model intentionally abstracted away from many practical frictions that restrict real-world PoW protocols, such as the lack of access to a synchronized clock, potentially asynchronous ledgers among the miners, etc. As we derived our impossibility theorem without imposing any such friction it follows that relaxing any of these practical restrictions through system design or

cryptographic improvements of the protocol will not lead to a protocol that improves upon Bitcoin in terms of energy spent on mining.

Second, while it is not surprising that miners will spend more resources in attempt to increases their chances, the theorem gives a specific function form for the competition between miners. The winning probability of a miner depends only on his his investment and the aggregate investment. These determine the equilibrium investment level and the wasteful expenditure on mining. Thus, reducing this wasteful expenditure in *any* PoW protocol requires violating one of our axioms.

4 Conclusion

We hope that our results will be helpful in clarifying the trade-offs between PoW systems and alternative designs. Proof-of-Stake systems violate our anonymity axiom, while maintaining a weaker version of anonymity. Our anonymity axiom is strong, it requires that all miners are treated equally regardless of their history within the system. Proof-of-Stake make use of the miner history within the system (and potentially disadvantages new entrants without a history), violating our assumptions and thus enabling different random selection mechanisms.

We think that analyzing the set of selection mechanisms that are achievable under weaker anonymity and robustness requirements is an important question for future research which could help guide the design of future crypto-currencies. We hope that the formalism we introduced in this paper will be helpful in this endeavour.

References

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.