# Selfish Mining Attacks Exacerbated by Elastic Hash Supply

Yoko Shibuya[1], Go Yamamoto[1], Fuhito Kojima[1], Elaine Shi[2], Shin'ichiro Matsuo[1], and Aron Laszka[3]

[1] NTT Research Inc, Palo Alto CA, USA
[2] Cornell University
[3] University of Houston

**Abstract.** Several attacks have been proposed against Proof-of-Work blockchains, which may increase the attacker's share of mining rewards (e.g., selfish mining, block withholding). A further impact of such attacks, which has not been considered in prior work, is that decreasing the profitability of mining for honest nodes incentivizes them to leave the attacked chain for a more profitable one (or to stop mining). The departure of honest nodes exacerbates the attack and may further decrease profitability and incentivize more honest nodes to leave. In this paper, we first present empirical analysis showing that there is statistically significant correlation between profitability of mining and the total hash rate, confirming that miners indeed respond to changing profitability. Second, we present a theoretical analysis showing that selfish mining under such elastic hash supply leads either to the collapse of a chain or to a stable equilibrium depending on the attacker's initial share.

**Keywords:** Blockchain · Selfish mining · Hash supply · Proof of Work.

## 1 Introduction

When blockchains were first introduced, it was believed that profitable attacks require at least 50% of the total mining power. However, several attacks have been found to go against proof-of-work blockchains, such as so-called selfish mining [2] and block withholding against mining pools [1]. A common goal of many such attacks is, at a high level, to increase the attacker's share of the mining rewards by reducing other miners' effective mining power. Prior work found that such attacks may be profitable even if the attacker's original share of the total mining power is less than 50%.

An important limitation of prior work is that they do not consider how honest miners react to changes in profitability when attacks occur. Most models assume that total hash supply in a chain is *fixed* and does not respond to changes in profitability of the chain. In practice, however, most miners are profit-oriented and choose which currency to mine (or to not mine at all) based on their profitability.

In this paper, we first document a real-world evidence of miner's profit-oriented behavior, using data from three different cryptocurrencies. We found positive and statistically significant correlation between total hash supply and

per-hash mining revenue, i.e., an evidence of *elastic* hash supply with respect to miners' revenue. We then provide a new analysis of selfish mining that takes into account the elasticity of hash supply. In an elegant work by Huberman et al. [3], the authors point out that Bitcoin mining is a free-entry, two-sided market. If there is a profit to be made, more miners will enter, which will then trigger the difficulty adjustment algorithm, making mining more difficult, and thus everyone's expected mining revenue decreases. In the equilibrium state, miners break even, i.e., the revenue that they earn from mining is equal to their cost. Inspired by this principle, we incorporate a free-entry condition in a model of selfish mining, and thus our analysis essentially characterizes the long-term effects of selfish mining to the eco-system in the equilibrium state.

To understand our analysis, let us first quickly review the classical analysis of the selfish mining attack: when a coalition (e.g., a mining pool) mines a new block $\mathsf{B}^*$ off the current longest chain denoted $\mathsf{chain}$, it does not release $\mathsf{B}^*$ immediately but withholds $\mathsf{B}^*$. Whenever an honest miner mines a block $\mathsf{B}$ also off $\mathsf{chain}$, the adversary releases the withheld block $\mathsf{B}^*$ immediately, and races in transmitting its block $\mathsf{B}^*$ to other miners. If the adversary has good control of the network (e.g., it controls some relays in the network) and can transmit its block $\mathsf{B}^*$ ahead of the honest block $\mathsf{B}$, it can convince other miners to mine off $\mathsf{B}^*$. In this case, the honest miners' work in mining $\mathsf{B}$ got erased. In other words, through selfish mining, an adversary controlling a coalition can erase some fraction of the honest mining power, and therefore the selfish coalition can gain an unfair share of the total rewards. In the worst case, assuming that the coalition can reliably win the race in transmission, then a $1/3$ coalition can erase $1/3$ of the honest mining power, and thus gain $1/2$ of the rewards.

The above classical analysis assumes that the total hash power participating in mining is fixed. Now let us consider what happens when honest miners may respond to profitability and freely enter and leave the system. During a selfish mining attack, because a fraction of the honest mining power is being erased, the erased fraction is essentially not gaining rewards. The immediate effect is that the cost of mining to earn each unit of reward becomes proportionally higher for honest miners; and if the honest miners' profitability plunges below zero, they start to leave the system. As honest miners leave, the impact of the attack on the remaining miners is magnified as a higher fraction of their mining power is now erased, which in turn drives more miners away. At the same time, as honest miners leave, the total mining power decreases. Therefore, the mining difficulty drops, and thus mining becomes cheaper—this second effect somewhat counteracts the decreased profitability for honest miners that stems from being the victim of selfish mining.

When hash power is elastic, what happens in the equilibrium state is driven simultaneously by the above two opposite effects. We show that for a wide range of parameter regimes, the first effect dominates and leads to a "collapse scenario"—specifically, selfish mining drives costs up for honest miners, and *all* honest miners end up leaving the system as a result. In some other parameter regimes, however, because the two effects somewhat counteract each other, the

system reaches a new equilibrium after some but not all honest nodes have left. In either scenario, the unfairness of selfish mining is significantly exacerbated by the elasticity of hashpower.

The rest of the paper is organized as follows. Section 2 shows our empirical evidence on the elasticity of hash supply. Section 3 describes our theoretical model of selfish mining under elastic hash supply. Section 4 concludes the paper.

## 2 Empirical Findings

This section presents new empirical facts on the elasticity of hash supply with respect to the miners' revenue. A few literature have worked on measuring elasticity of hash supply [4,5], but our paper distinguishes itself from the prior literature in three ways. First, we use various time-detrending methods from macroeconomics to deal with technological advancements in cryptocurrencies and related time trends in variables. Second, using time-detrending methods allows us to study longer time periods (5 years). Lastly, we study data from three different cryptocurrencies, which helps us to generalize our findings.

We start by explaining the data that we use in our study (Section 2.1). We then describe time-detrending methods and regression strategy (Section 2.2). Finally, we conclude the section by showing regression results (Section 2.3).

### 2.1 Data

We downloaded cryptocurrency data from three sources: Bitcoin data from Quandl, Ethereum data from Etherscan, and Ethereum Classic data from crypto-ethereum-classic public library on BigQuery. We use three variables in our regression analysis: daily price, network difficulty, and total hash rate of each cryptocurrency. Different currencies have different lengths of history, and thus we use data from 2017/1/1 to 2020/7/31 for Ethereum and Ethereum Classic, and from 2015/1/1 to 2020/7/31 for Bitcoin. We computed daily per-hash revenue from coinbase using daily price and network difficulty (and data on cryptocurrency halving). We focus on miner's revenue from coinbase and not from transaction fees because transaction fees have been randomly fluctuating over the recent years in these cryptocurrencies.

### 2.2 Empirical Analysis Strategy

**Time Detrending Methods** Technological advancements in cryptocurrency mining over the past 10 years pose a challenge for regression analysis since they add significant time trends to the variables. To remove the time trend components from variables, we apply three types of time-detrending filters that are commonly used in macroeconomics for detrending time series variables: Hodrick-Prescott (HP), Baxter-King (BK), and Christiano-Fitzgerald (CF) filters.[4]

[4] For HP filter, we use $\lambda = 10,000$. For BK filter we use (7, 90, 12) for high, low frequencies, and lead-lag length, respectively. For CF filter, we use (7, 90) for high and low frequency length.

Figure 1 shows the decomposition of the logarithm of total hash rate in the Bitcoin network over the past three years, using Hodrick-Prescott filter. The total hash rate of the Bitcoin network has an increasing trend over this time period, and the filter removes out the trend. In the later regression analysis, we use the cycle components of the variables after applying filters. We report regression results based on multiple time-detrending methods.

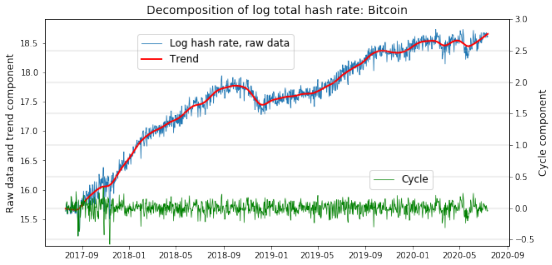


**Fig. 1.** Application of HP filter to raw hash-rate data from Bitcoin.

**Regression Equations** To estimate the elasticity of total hash rate with respect to per-hash revenue, we consider the following regression equation:

$$\Delta \log THR_{i,t} = \alpha_i \Delta \log MRC_{i,t} + \epsilon_{i,t}, \tag{1}$$

where THR stands for total hash rate, MRC stands for miners' per-hash revenue from coinbase. $i$ is an index representing the cryptocurrency (Bitcoin, Ethereum, or Ethereum Classic), and $t$ is an index for time (day). $\Delta$ variables are cycle components of the logged variables.[5] We include year-month fixed effect in the regression to take out some year/month fixed events such as regulation changes, which might not be taken out by time-detrending filters.

### 2.3 Results

Table 1 summarizes the results of running the above regression for the three cryptocurrencies. The main result of the regression analysis is that with any type of time-detrending filter, in any time period, and for any currency, the coefficients on $\Delta \log \mathrm{MRC}$ are *positive and statistically significant*. In other words, the total hash rate is elastic with respect to the miners' per-hash revenue from coinbase. The magnitude of the coefficient varies across different time-detrending methods and different currencies, but the elasticity ranges from 0.028 to 0.183.

[5] For regressions with Ethereum Classic data, we use daily difference in total hash rate as an independent variable. The reasons for this is that total hash rate of Ethereum Classic is volatile at high frequency, and does not exhibit any time trend over the sample period.

One percentage change in the miners' per-hash revenue from coinbase causes 0.027 to 0.183 percentage change in the total hash rate.

Regression with longer sample period for Bitcoin data gives us more interesting result. Table 2 summarizes the regression results for Bitcoin data with different sample periods. Interestingly, elasticity is higher and more statistically significantly positive in the recent period (2018–) compared to the beginning of the sample period (2015–2017). This shows the possibility that hash rate becomes more responsive to the miners' revenue as a currency grows.

**Table 1.** Regression results for three currencies in sample period 2017/1/1–2020/7/31

| | **Bitcoin** | | | **Ethereum** | | | **Ethereum Classic** | | |
|---|---|---|---|---|---|---|---|---|---|
| | HP | BK | CF | HP | BK | CF | HP | BK | CF |
| $\Delta \log \mathrm{MRC}$ | $0.175^{***}$ | $0.183^{***}$ | $0.181^{***}$ | $0.028^{***}$ | $0.033^{***}$ | $0.079^{***}$ | $0.041^{***}$ | $0.048^{***}$ | $0.027^{***}$ |
| | (5.53) | (8.83) | (1.30) | (3.69) | (5.08) | (12.54) | (3.20) | (3.12) | (2.57) |
| No. of obs. | 1308 | 1296 | 1308 | 1308 | 1296 | 1308 | 1308 | 1296 | 1308 |

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$, t-values in parentheses.

**Table 2.** Regression results for Bitcoin data with three different sample periods

| | **2015/1 - 2017/12** | | | **2018/1 - 2020/7** | | | **2015/1 - 2020/7** | | |
|---|---|---|---|---|---|---|---|---|---|
| | HP | BK | CF | HP | BK | CF | HP | BK | CF |
| $\Delta \log \mathrm{MRC}$ | $0.082^{*}$ | $0.108^{***}$ | $0.078^{***}$ | $0.163^{***}$ | $0.152^{***}$ | $0.194^{***}$ | $0.126^{***}$ | $0.132^{***}$ | $0.143^{***}$ |
| | (2.02) | (3.83) | (3.62) | (4.85) | (6.88) | (11.76) | (4.80) | (7.38) | (10.65) |
| No. of obs. | 1096 | 1084 | 1096 | 943 | 931 | 943 | 2039 | 2015 | 2039 |

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$, t-values in parentheses.

## 3 Model with Elastic Hash Supply

This section introduces our model of selfish mining with elastic hash supply. We first explain our baseline model without selfish mining and illustrate how total hash rate is determined endogenously in an equilibrium by free-entry condition. We then analyze the model with selfish mining, building on the seminal work by Eyal and Sirer [2]. Lastly, we discuss the stability of equilibria.

In our model of selfish mining with elastic hash supply, the equilibrium state is determined by the two opposing effects. An attack increases the cost of mining for honest miners and thus makes honest miners leave. At the same time, when some honest miners leave, the total mining power decreases and so does the cost of mining for honest miners. Which effect dominates depends on the attacker's initial share of mining power. We derive a threshold for the attacker's initial share such that (a) if the attacker's share is below the threshold, the system has

a stable equilibrium with a positive hash supply by honest miners; and (b) if the attacker's share is above the threshold, all honest miners leave and the system collapses. In either case, some or all honest miners leave the system, and thus the effect of selfish mining is significantly exacerbated under elastic hash supply.

**Baseline Model Without Selfish Mining** We consider a system with a group of honest miners (with mining power $H$) and an attacking pool (with mining power $M$). We let $B$ and $C$ denote block rewards and per-hash mining cost, respectively. We assume *elastic hash supply* in the system: the equilibrium mining power of honest miners ($H^*$) is determined such that honest miners make zero profit with mining power $H^*$. The attacking pool's mining power ($M$), block rewards ($B$) and cost ($C$) are assumed to be fixed and to satisfy $M < B/C$.

Without selfish mining attack, the honest miners' per-hash profit is

$$\mathcal{U}^N(H) = B\frac{1}{H+M} - C. \tag{2}$$

In an equilibrium, the elastic hash supply assumption implies $\mathcal{U}^N(H^*) = 0$. We can solve for $H^*$:

$$H^* = \frac{B}{C} - M > 0 \tag{3}$$

**Model With Selfish Mining** Now, we assume that the attacking pool performs selfish mining as defined by [2]. We can calculate[6] the expected mining reward per block discovery, including the hidden block discoveries, as

$$B_{\text{attacker}} = B\frac{\left(-2\,\alpha^4 + 5\,\alpha^3 - 4\,\alpha^2 + \alpha\right)\,\gamma + 4\,\alpha^4 - 9\,\alpha^3 + 4\,\alpha^2}{2\,\alpha^3 - 4\,\alpha^2 + 1}$$

for the attacking pool, and

$$B_{\text{honest}} = B\frac{\left(2\,\alpha^4 - 5\,\alpha^3 + 4\,\alpha^2 - \alpha\right)\,\gamma - 4\,\alpha^4 + 10\,\alpha^3 - 6\,\alpha^2 - \alpha + 1}{2\,\alpha^3 - 4\,\alpha^2 + 1}$$

for the honest miners, where we denote by $\alpha = \frac{M}{H+M}$ the fraction of the attacking pool's mining power out of the total mining power, and by $\gamma$ the ratio of honest miners that choose to mine on the attacking pool's block. The total effective mining power in the system under attack is $(B_{\text{honest}}+B_{\text{attacker}})(H+M)/B$. The honest miners' effective mining power is given by $B_{\text{honest}}(H+M)/B = \frac{B_{\text{honest}}}{(1-\alpha)B}H$ and the attacking pool's effective mining power is $B_{\text{attacker}}(H+M)/B = \frac{B_{\text{attacker}}}{\alpha B}M$.

Then, the honest miners' per-hash profit under selfish mining attack is

$$\mathcal{U}^S(H) = B\frac{B_{\text{honest}}}{(1-\alpha)B}\frac{B}{(B_{\text{attacker}} + B_{\text{honest}})(H+M)} - C. \tag{4}$$

[6] These calculations should coincide $B_{\text{attacker}} = B \cdot r_{\text{pool}}$ and $B_{\text{honest}} = B \cdot r_{\text{others}}$, where $r_{\text{pool}}$ and $r_{\text{others}}$ are from Equations (6) and (7) in [2].

In an equilibrium, honest miners' hash supply is again derived from $\mathcal{U}^S(H^*) = 0$:

$$\mathcal{U}^S(H^*) = B\frac{1}{M}\left\{\frac{\alpha^* \cdot B_{\text{honest}}(\alpha^*)}{(1-\alpha^*)(B_{\text{attacker}}(\alpha^*) + B_{\text{honest}}(\alpha^*))} - \beta\right\} = 0 \qquad (5)$$

for $\alpha^* = \frac{M}{H^*+M}$ and $\beta = M \cdot \frac{C}{B}$.

A natural question is whether the above equilibrium condition has a solution $H^* > M$. If not, then the system cannot find an equilibrium where honest miners stay in the system under selfish mining attack. This simple theorem answers that the attacker's hash rate must be bounded to avoid collapsing the system.

**Theorem 1.** *For any given $\gamma$, there exists $M_{max}$ such that a solution $H^*$ of $\mathcal{U}^S(H^*) = 0$ with $H^* > M(> 0)$ exists if and only if $M \leq M_{max}$.*

*Proof.* Let us define a function $f(\alpha) = \frac{\alpha \cdot B_{\text{honest}}(\alpha)}{(1-\alpha)(B_{\text{attacker}}(\alpha)+B_{\text{honest}}(\alpha))}$ . First, $f(\alpha)$ is continuous for $0 \leq \alpha \leq 1/2$ because the denominator of $f(\alpha)$ is $\alpha^3 - 2\alpha^2 - \alpha + 1$, and is strictly positive for $0 \leq \alpha \leq 1/2$. Since $f$ is continuous, there exists $\alpha_{\max} \in [0, 1/2]$ that achieves the maximum of $f(\alpha)$ for $0 \leq \alpha \leq 1/2$. Let $M_{\max} = \frac{B}{C} f(\alpha_{\max})$. If $M_{\max} < M$, then $\mathcal{U}^S(H) < 0$ for all $H$ such that $H > M$, so solution $H^*$ does not exist. To complete the proof it suffices to find a solution $H^* > M$ of $\mathcal{U}^S(H) = 0$ for constant $M$ that satisfies $0 < M \leq M_{\max}$. There exists some $\alpha^* \in (0, 1/2)$ such that $f(\alpha^*) = \beta$ because $f$ is continuous, $f(0) = f(1/2) = 0$, and $0 < \beta \leq \frac{C}{B} M_{\max} = f(\alpha_{\max})$. We find $H^*$ by solving $\alpha^* = \frac{M}{H^*+M}$, and $H^* > M$ because $\alpha^* < 1/2$. □

We can find $\alpha_{\max}$ by solving $f'(\alpha) = 0$:

$$\gamma = \frac{4\,\alpha^6 - 16\,\alpha^5 + 26\,\alpha^3 - 16\,\alpha^2 + 1}{2\,\alpha^6 - 8\,\alpha^5 - \alpha^4 + 14\,\alpha^3 - 10\,\alpha^2 + 2\,\alpha}. \qquad (6)$$

With elastic hash supply, selfish mining attacks reduce the profitability of honest miners, making honest miners leave the system, which in turn increases the attacker's share, further decreasing profitability for honest miners. If the attacking pool's share is large enough, the negative propagation effect forces all honest miners to leave the system. For example, when $\gamma = 1$, we find that $f(\alpha_{\max})$ is approximately 0.292. Since $H^* + M = \frac{B}{C}$ in the equilibrium without selfish mining attacks, this implies that if the attacking pool's share is larger than 29.2%, then the attack makes all the honest miners eventually leave the system. When $0 \leq \gamma \leq 1$, $f(\alpha_{max})$ is decreasing in $\gamma$, ranging from 0.3475 at $\gamma = 0$ to 0.2919 at $\gamma = 1$.

When the system does not collapse, we can find a stable equilibrium from the honest miners' response. It is straightforward to check[7] that Eq. (6) has only one solution in $0 \leq \alpha \leq 1/2$. This implies that we have only two equilibria $H^*_{\text{L}}$ and $H^*_{\text{H}}$ for $H^*$ when $M < M_{\max}$. We assume $H^*_{\text{L}} < H^*_{\text{H}}$ without loss of generality. Since $f'(\frac{M}{H^*_{\text{L}}+M}) < 0$ and $f'(\frac{M}{H^*_{\text{H}}+M}) > 0$, we obtain the following proposition.

[7] We omit the details due to the restriction of space.

**Proposition 1** *For any given $\gamma$ and $M < M_{max}$, there are two equilibria, $H^*_{\mathrm{H}}$ and $H^*_{\mathrm{L}}$ ($H^*_{\mathrm{H}} > H^*_{\mathrm{L}}$), where $H^*_{\mathrm{H}}$ is stable and $H^*_{\mathrm{L}}$ is unstable.*

Figure 2 illustrates the honest miners' per-hash revenue and cost, given parameters $B$, $C$, $\gamma$, and $M$.[8] Under the free entry condition, the equilibria correspond to points L and H where the revenue curve intersects the cost, i.e., points with zero profit. In this case, equilibrium H is stable, while L is not. When honest miners' mining power increases (decreases) by any small amount $\epsilon > 0$ from point L, positive (negative) profit will be generated and more honest miners will enter (leave) the system, ending up reaching equilibrium H (or an equilibrium $H = 0$[9]). On the other hand, when mining power increases (decreases) from point H, negative (positive) profit will be generated and honest miners leave (enter) the system. Therefore equilibrium H is the only stable equilibrium.

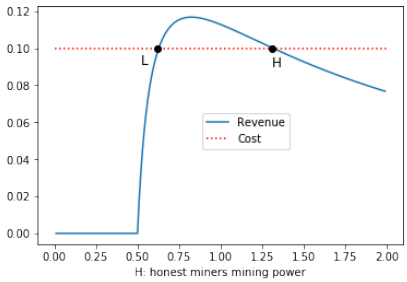


**Fig. 2.** Honest miners' revenue and cost

## 4 Conclusions

Majority of selfish mining literature assumes *fixed* total hash power. Our results show that *elastic* hash supply can significantly exacerbate the impact of selfish mining. We (i) showed empirically that hash supply is elastic with respect to the miners' per-hash revenue and (ii) theoretically derived a threshold such that if the attacker's initial share of the total mining power is above the threshold, all the honest miners will leave and the chain collapses. Limitations of our theoretical analysis lead us to future work. First, whether the equilibrium can be reached depends on the starting state. For example, if $H = 0$, then it will not be profitable for any individual honest miner to start mining if its hash power is less than $M$ (regardless of the relation between $M$ and $\frac{B}{C}$). Second, our analysis ignored transient effects, which may prevent reaching particular equilibria. For example, if difficulty adjustments are delayed, an attacker with $M \leq f(\alpha_{\max})\frac{B}{C}$ might be able to "chase away" honest miners before difficulty adjusts to incentivize them to stay. Our future work includes extending our model to dynamic analysis considering initial state and transient effects.

[8] We set $B = 0.2$, $C = 0.1$, $\gamma = 1$, and $M = 0.5$ for Figure 2.

[9] While $H = 0$ is an equilibrium, we do not consider cases where $H < M$ in our analysis since it is well known that such cases are unsustainable.

## References

1. Eyal, I.: The miner's dilemma. In: 36th IEEE Symposium on Security and Privacy (S&P). pp. 89–103. IEEE (2015)
2. Eyal, I., Sirer, E.G.: Majority is not enough: Bitcoin mining is vulnerable. In: 18th International Conference on Financial Cryptography and Data Security (FC). pp. 436–454. Springer (2014)
3. Huberman, G., Leshno, J., Moallemi, C.C.: An economic analysis of the bitcoin payment system. Columbia Business School Research Paper No. 17-92
4. Noda, S., Okumura, K., Hashimoto, Y.: An economic analysis of difficulty adjustment algorithms in proof-of-work blockchain systems. Available at SSRN: `http://dx.doi.org/10.2139/ssrn.3410460` (June 2019)
5. Noda, S., Okumura, K., Hashimoto, Y.: An economic analysis of difficulty adjustment algorithms in proof-of-work blockchain systems. In: 21st ACM Conference on Economics and Computation (EC). pp. 611–611 (2020)