



OPEN

Statistical detection of selfish mining in proof-of-work blockchain systems

Sheng-Nan Li^{1,2}, Carlo Campajola^{1,2,3,4} & Claudio J. Tessone^{1,2}✉

The core of many cryptocurrencies is the decentralised validation network operating on proof-of-work technology. In these systems, validation is done by so-called miners who can digitally sign blocks once they solve a computationally-hard problem. Conventional wisdom generally considers this protocol as secure and stable as miners are incentivised to follow the behaviour of the majority. However, whether some strategic mining behaviours occur in practice is still a major concern. In this paper we target this question by focusing on a security threat: a selfish mining attack in which malicious miners deviate from protocol by not immediately revealing their newly mined blocks. We propose a statistical test to analyse each miner's behaviour in five popular cryptocurrencies: Bitcoin, Litecoin, Monacoin, Ethereum and Bitcoin Cash. Our method is based on the realisation that selfish mining behaviour will cause identifiable anomalies in the statistics of miner's successive blocks discovery. Secondly, we apply heuristics-based address clustering to improve the detectability of this kind of behaviour. We find a marked presence of abnormal miners in Monacoin and Bitcoin Cash, and, to a lesser extent, in Ethereum. Finally, we extend our method to detect coordinated selfish mining attacks, finding mining cartels in Monacoin where miners might secretly share information about newly mined blocks in advance. Our analysis contributes to the research on security in cryptocurrency systems by providing the first empirical evidence that the aforementioned strategic mining behaviours do take place in practice.

Blockchains are decentralised and distributed systems, where sequential, verified data is stored sequentially in blocks of a chain and data transmission is secured from manipulation through cryptography. Among all blockchain-based technologies, cryptocurrencies are the most famous ones. The original crypto consensus mechanism is called "Proof-of-Work" (PoW) and is employed in the majority of cryptocurrency systems^{1,2}. The consistency of a PoW system's ledger is maintained by all participants solving hash puzzles, a process usually called "block mining". In order to solve the puzzles, attempts have to be made through brute force and therefore, *a priori*, the probability of finding a solution is proportional to the number of tries per unit of time a miner is able to perform, measured in hashes per second (H/s). Each miner is then rewarded by a nominal amount of cryptocurrency if they are the first acknowledged miner to find a valid block in the longest chain of the network. This type of rewarding system provides an incentive for miners to contribute their resources to the system, and is essential to the cryptocurrency's decentralised nature. According to this mechanism, the more mining power (resources) a miner invests, the bigger their chance to mine the next block first³: as a result miners often join in mining pools to share their mining powers, thus reducing the variance of their rewards.

PoW mining protocols in principle are tailored to be resistant towards multiple kinds of attacks, but several potentially harmful strategies have been analysed in the literature and some of them have been shown to be profitable under proper conditions. Some attacks might influence the information propagation in the peer-to-peer network, as is the case for Sybil attacks, eclipse attacks⁴ and routing attacks⁵; others could threaten data consistency, such as double-spending attacks⁶ or block withholding attacks⁷, which are the focus of this paper. According to the PoW protocol, when miners find a block they should submit it to their peer nodes unconditionally. However, in a block withholding attack, miners could decide to not submit the block, or to postpone submitting it. The former, which is also named as sabotage, has no direct benefit for the attacker but can harm the other miners; the latter however, which is also known as selfish mining (SM), can be shown to be potentially profitable for the attacker.

¹Blockchain and Distributed Ledger Technologies, Faculty of Business, Economics and Informatics, University of Zurich, 8050 Zurich, Switzerland. ²UZH Blockchain Center, Faculty of Business, Economics and Informatics, University of Zurich, 8050 Zurich, Switzerland. ³DLT Science Foundation, London, UK. ⁴Institute of Finance and Technology, University College London, London, United Kingdom. ✉email: claudio.tessone@uzh.ch

The selfish mining (SM) attack was first described by Eyal and Sirer⁸ in 2014 as follows: the selfish mining pool keeps its mined blocks private, secretly creating a private branch. However, their private branch will not remain ahead of the public branch indefinitely. Consequently, a selfish miner judiciously reveals blocks from the private branch to the public to ensure a profit from the attack. The SM strategy is illustrated in Fig. 1. Let's classify the miners in a stylised P2P mining network into two groups: selfish (red node) and honest (green and black nodes). At t_1 , a selfish miner mines a block (in red) appended to the longest chain terminating at t_0 , but withholds it and secretly starts mining on this new private branch. Then, if the selfish miner finds the next block (t_{2A}), the attack is successful and they can choose whether to publish the new chain or continue mining selfishly; however, if an honest miner also finds a block (in green) at the same height (t_{2B}), the selfish miner will immediately publish its secret block, triggering a competition. Later (t_3, t_{3A}), if the selfish miner finds the next block after its own block, it is also a successful attack, whereas if (t_{3B}) an honest miner finds the next block after the selfish miner's block, the selfish miner still enjoys the revenue of the first block. The only negative outcome for the selfish miner is (t_{3C}), where an honest miner finds the next block after the honest miner's block, resulting in the selfish miner not getting any reward. The original work by Eyal and Sirer demonstrates that a selfish pool that controls more than 1/3 of the total mining power will always be able to gain mining rewards in excess of its proportion of mining power. Moreover, for a mining pool with high connectivity and good control of information flow, the power threshold for its profitability could be close to zero, which implies that by leveraging on the information asymmetry even the smaller pool has the potential to earn profits by doing selfish mining. Consequently, if the higher rewards of selfish pools encourage more miners to join, it may eventually lead to the selfish pool holding the majority of mining power, and collapse the decentralised nature of the cryptocurrency ecosystem.

The formulation of the SM attack has drawn a lot of attention and many extended mining strategies have been proposed, such as *stubborn mining*⁹ and the *publish- n strategy*¹⁰. Many of these extended strategies use Markov Decision Processes (MDP) solving to compute optimal selfish mining strategies and have managed to lower the profitability threshold of running a SM attack from 25% hash power⁸, to 23.21%¹¹, or even 21.48%¹². Meanwhile, various countermeasures have been proposed against SM attacks. Zhang¹³ has categorised the existing defence methods into two approaches: 1) making fundamental changes to the block validity rules, for example adopting the *ZeroBlock*¹⁴ timestamp-free solution which requires that each block must be generated and received by the network within a maximum acceptable time interval, or 2) lowering the chance of honest miners working on the selfish miner's chain during a forked situation, as in the case of the *Freshness Preferred* defence¹⁵ which uses unforgeable timestamps issued by a trusted party, providing an incentive for miners to immediately publish newly mined blocks. To replace the original Bitcoin Fork-Resolving Policy (FRP), denoted by *length FRP*, Zhang¹³ proposed *weighted FRP*: it asks miners to compare the weight of the chains instead of their length, where the *weight* is the number of "in time" blocks in the chain, and a block is considered "in time" based on an upper bound on the block propagation time. However selfish miners' timely reaction to another competitive block, and the high cost of changing the blockchain's fundamental design, might be obstacles to efficiently implementing defences against SM attacks. More essentially, the problem of how to detect these selfish miners and quantify the size of the attack is a more urgent problem for the already running blockchain platforms. At the end of 2020, Nicolas analysed the benefits and limitations of 6 SM detection models¹⁶. From his summary, one can easily find that most of the existing detection methods have not been tested on real blockchain systems. Hou et al.¹⁷ used a deep reinforcement learning framework, called *SquirRL*, to evaluate both single and multiple agent selfish mining attacks in Bitcoin, Monacoin and Litecoin. The empirical data they scraped is the estimated hourly total hash power from real cryptocurrency systems. Some stochastic modelling of the PoW mechanism^{18,19} has verified the profitability of SM attack and shown the attackers can leverage on their location in the P2P network. However, none of these previous studies has detected selfish miners in any real blockchain platforms. The question of whether selfish mining exists in practice or not is largely left unanswered so far.

Although selfish mining attacks have not been empirically discovered by academic research, Monacoin, a cryptocurrency developed in Japan, reportedly has suffered a selfish mining attack that caused damage estimated

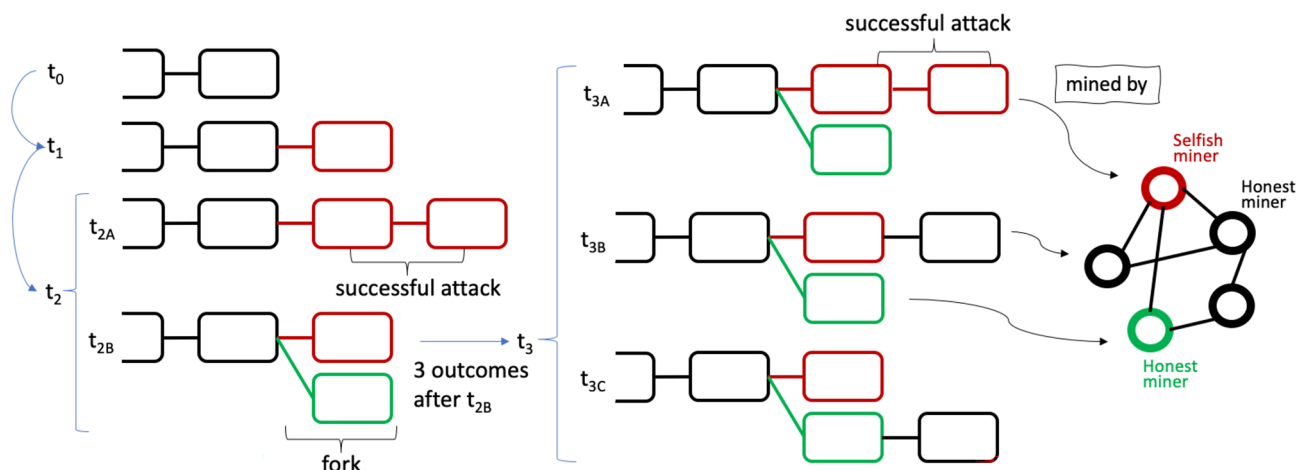


Figure 1. Visualisation of selfish mining strategy.

at roughly \$90,000²⁰. Therefore, the empirical evidence on whether miners do deviate from the mining protocols in practice is important to the security and stability of cryptocurrencies²¹, and it is necessary to direct research towards refining detection methods.

In our previous work^{22,23}, we had tried to identify the selfish miners in real cryptocurrency systems by proposing the Miner Sequence Bootstrapping method (MSB), the core of which is to shuffle simulations of the sequence of miners' block discoveries. The MSB method is more strict with users with low computing power. In this paper, we propose a more comparable and interpretable statistical test to evaluate miners' behaviour in five popular PoW-based cryptocurrencies: Bitcoin, Litecoin, Ethereum, Monacoin and Bitcoin Cash. The logic informing our detection method is that selfish miners' behaviour of selectively revealing their mined blocks in sequences would cause abnormal occurrences of successive block discovery, diverging from the expected behaviour under a null hypothesis of statistical independence between mining outcomes. We define these abnormal occurrences as *selfish anomalies*. As we show in the following, under the null hypothesis of "honest" mining the number of sequences of two blocks in a row mined by the same miner over a given time interval is characterised by a *type II binomial distribution of order 2*²⁴, which we use to construct a statistical test to detect mining anomalies. In order to optimise the detection results, we also apply heuristics-based address clustering techniques on all UTXO blockchain datasets. Furthermore, we extended the object of our method from single miner (mining pool) to pairs of miners that may constitute a *mining cartel*, in which miners secretly share information related to blocks discovery before publishing. Our main contributions are listed as follows :

1. To the best of our knowledge, our empirical research on selfish mining attacks and mining cartels in real cryptocurrency systems is presented for the first time. Mining attack detection is important for maintaining blockchain security and could be a fundamental index for cryptocurrencies ranking in the future.
2. In our mining behaviour detection test, we use a *type II binomial distribution of order 2* to compute each miner's probability of successive block discovery. This can be widely applied in various competitive consensus protocols, including but not limited to PoW and PoS.
3. Our results show that in some cryptocurrencies abnormal miners do secretly collaborate in mining cartels; this could raise concerns about concentration of mining power which has been ignored by most of previous studies.
4. We highlight the importance of heuristic address clustering for empirical studies in real blockchain systems, especially for user behaviour analysis.
5. Our empirical analyses also reveal that mathematical or economical models that focus on cost-benefit analysis could fail to detect some behaviours, as participants of cryptocurrencies might have bounded rationality or be risk seeking.

Results

Dashboard of datasets

The mining difficulty adjustment in the PoW protocol ensures a fixed average time between each block, called "block time". Since Bitcoin (BTC), Litecoin (LTC), Monacoin (MONA), Ethereum (ETH) and Bitcoin Cash (BCH) have different block times, in order to have compatible datasets we split the blockchain in different time intervals tailored to maintain a similar number of blocks (~ 5000) in each sample. This amounts to monthly (BTC and BCH), weekly (LTC), 5 days (MONA) and daily (ETH) splits. In Fig. 2 we show the number of blocks mined in each time interval for the five cryptocurrencies from the genesis block until the end of our dataset on December 2020. One can find that the mining markets of all the five cryptocurrencies have unstable stages of block mining with different lengths after launch. However, because of the difficulty adjustment, the amount of blocks in each stated period is similar in five coins (as shown in Fig. 2a). In our Ethereum and Monacoin datasets we only have information about each block miner's address, while miners' addresses were tagged to named mining pools in the Bitcoin, Litecoin and Bitcoin Cash datasets.

We further show the revenue (amount of mined blocks) distributions in each period in Fig. 3. The "Unknown" miner are some mining addresses whose identities cannot be traced back to any known pool. It is worth mentioning that some of the unknown mining addresses might be owned by named pools (e.g. to hide their activities such as selfish mining). In detail, one can observe that in BTC and LTC as time flows more and more blocks were mined by named pools, while in BCH there are more than 20% of blocks that are mined by "Unknown" miners all the time. Comparing MONA and ETH where we lack the information about mining pool identity, we find that in MONA the revenue distributions among miners' addresses is more volatile.

In addition, according to the "PoW" mechanism, the fair proportion of blocks a miner may discover during a time period (i.e. their blocks share) is equal to their share of mining power. Since we lack better estimators of hash rates, for the rest of this paper we use a miner's share of blocks as a proxy for its mining power.

Address clustering

After finding the "Unknown" miners and large fluctuation of hashing power distribution, to enhance the datasets we adopt known methodologies to cluster addresses controlled by the same miner^{25–28}. All the heuristic methods we applied exploit inherent properties of UTXO-based transactions, which can include multiple inputs and multiple outputs and generate patterns that allow to cluster addresses together. This was not possible on the account-based blockchain of Ethereum, where only one input and one output can appear in the same transaction. Further details of the three applied methodologies (H_1 , H_2 , H_p) are provided in the Methods section.

We apply the most basic method, H_1 , to cluster the miners' addresses in the Monacoin dataset. The distribution of mined blocks share among different entities (addresses or clusters) is shown in Fig. 4, and the outcome of clustering in MONA is shown in the subplot of network in Fig. 4. In the latter dots are addresses, connected

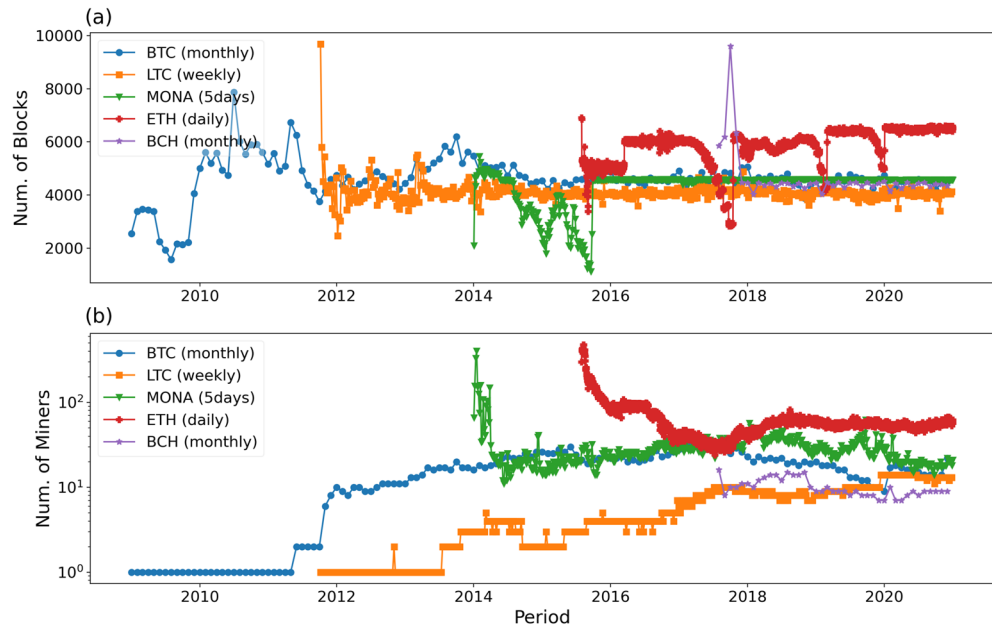


Figure 2. Number of miners and blocks during each period in five cryptocurrencies.

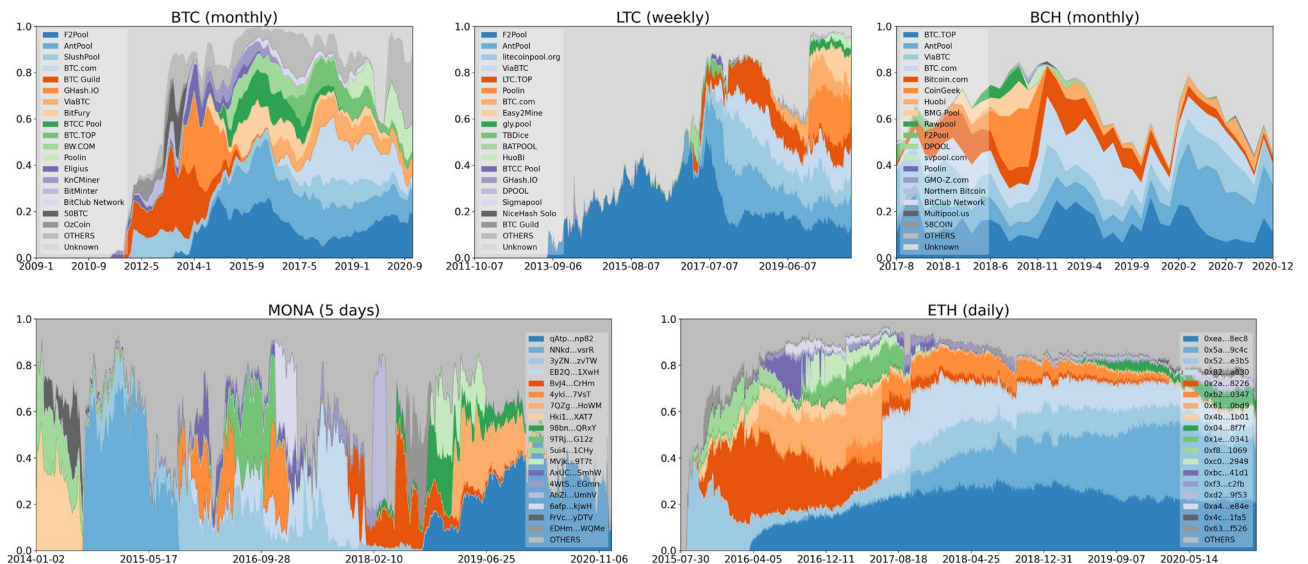


Figure 3. Periodic hashing power (share of blocks) distribution in BTC, LTC, BCH, MONA and ETH.

and marked in the same colour if they belong to the same cluster, and the highlighted communities are the larger clusters with more than 10 addresses. It can be seen that the H_1 methodology aggregates miner’s addresses into clusters of different size, effectively changing the estimation of hashing power attributed to them.

For the Bitcoin, Litecoin and Bitcoin Cash datasets where we already knew the named-pools of part of the blocks, we try all the three mentioned heuristics to tag the “Unknown” miners to named pools, with the priority order as $H_1 > H_2 > H_p$. In other words, in each among the BTC, LTC, and BCH datasets, firstly we apply H_1 to cluster the miners’ addresses, and then each “Unknown” miner whose address could be clustered together with all the addresses of a named mining pool will be tagged to this named pool. We then do the same using the H_2 and H_p methods to complement the tagging.

The result of this procedure is shown in Fig. 5, where it is clear that although it’s difficult for address clustering heuristics to tag all the “Unknown” miners, a significant fraction of blocks can be attributed to a tagged pool, which is important for a more accurate estimation of miners’ actual computing (hashing) power.

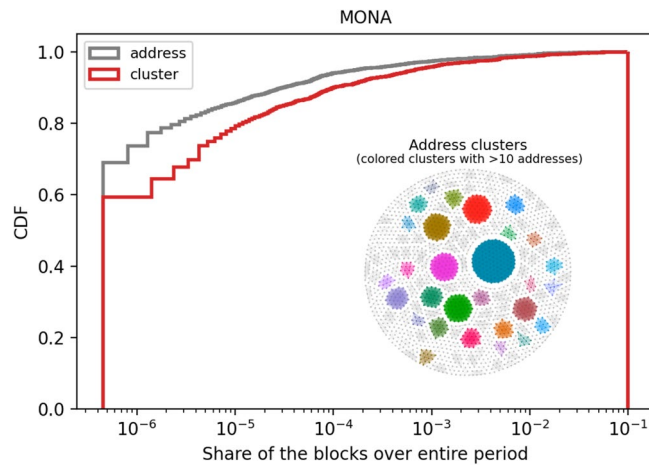


Figure 4. Address Clustering in Monacoin.

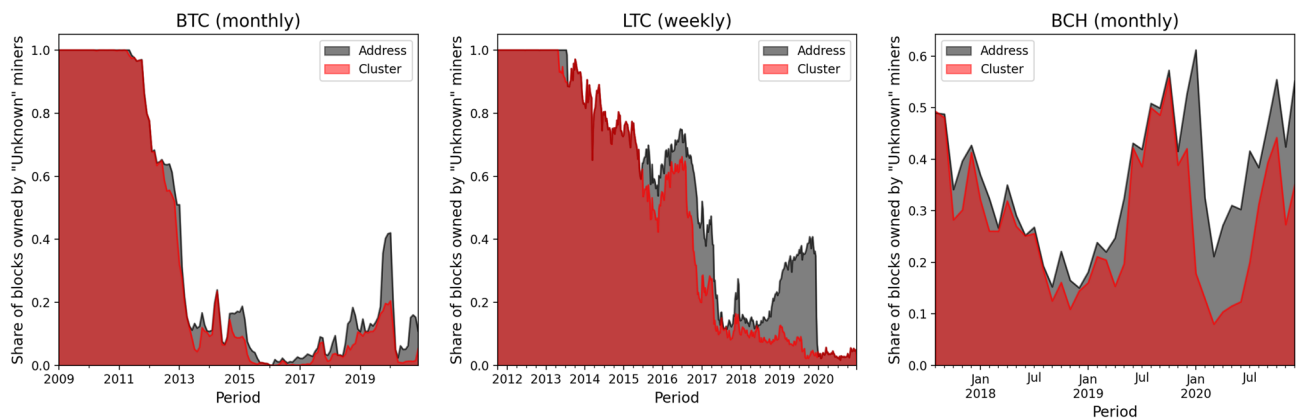


Figure 5. Periodic share of blocks owned by “Unknown” miners before and after clustering in BTC, LTC, BCH.

Abnormal miners in real-world cryptocurrencies

To detect abnormal selfish mining behaviour, we devise a statistical test that we apply on each miner’s sequence of mined blocks. The null hypothesis is that miners are “honest”, i.e. they act without selfish behaviour. As we show in the Methods section, under the null hypothesis the event of whether a miner mines a block or not is a Bernoulli random variable, with the success probability equal to the miner’s hashing power share. However a successful selfish mining attack could lead to anomalies in a miner’s outcome of discovering blocks *in sequence*. Therefore, we design our test statistic to identify suspicious miners by the amount of times in which they mine successive blocks, i.e. the number of success *runs* of length 2, whose probability distribution under the null is given by a *type II binomial distribution of order 2*²⁴. To account for multiple hypothesis testing errors we apply the Benjamini-Hochberg correction²⁹ for the *p*-values to control for excess false positives, setting the target False Discovery Rate (FDR) to 5%.

The results of our tests (before address clustering) are shown in aggregate in Fig. 6. In Fig. 6a, each bar shows the proportion of abnormal miners (with the corrected *p*-values, $\hat{p} < 0.05$) in the five cryptocurrencies. Bars in different colours represent results under different classification criteria: the blue, orange, green, yellow and the grey bar respectively show the fraction of abnormal miners for whom at least 25%, 50%, 75% or all (max) tests during the considered time periods reject the null hypothesis at 5% FDR. For example, in Monacoin, the result expressed by the grey bar shows that about half of the miners behaved selfishly in all the periods they were active. We then compare the detection results before and after address clustering, shown in Fig. 6b where the red dashed lines display the results after address clustering in Bitcoin (circle), Litecoin (square), Monacoin (inverted triangle) and Bitcoin Cash (star). Following address clustering the ratio of abnormal miners in each coin decreases; however, even after clustering, there were more than 46% miners who always engaged in strategic mining behaviour in Monacoin. In addition, the reduction in abnormal ratio when changing the criterion from lower quartile (25%) to maximum (max) is larger in Bitcoin Cash than in Monacoin, which shows the abnormal miners in Monacoin might be more likely to continuously behave with the selfish strategy, or alternatively that many malicious miners might enter Monacoin only to run the SM attack, leaving the system right after.

In addition, we show the number of abnormal miners for each period in Monacoin, Ethereum and Bitcoin Cash in Fig. 7. The result of each period includes all miners whose corrected *p*-value, \hat{p} , is smaller than 0.05

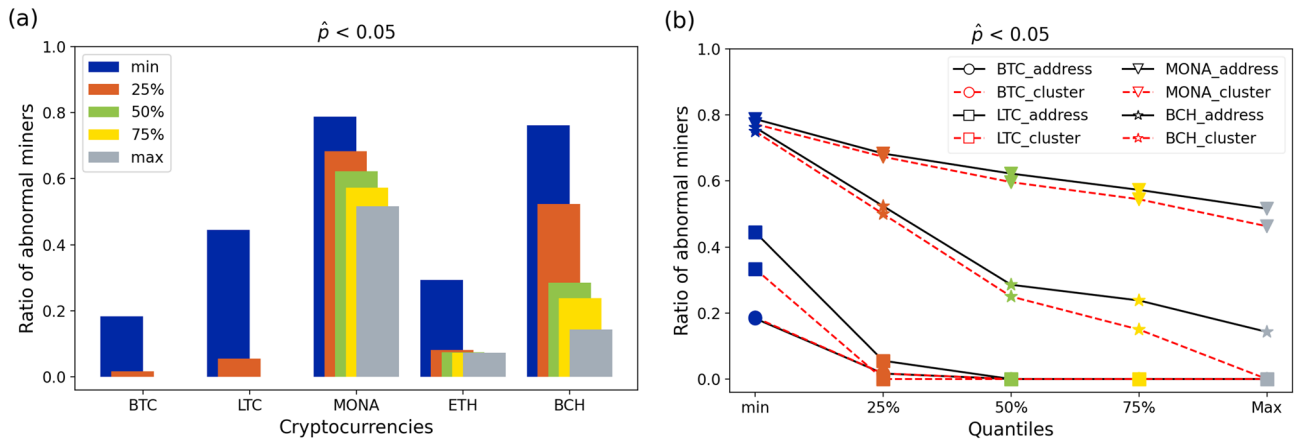


Figure 6. Ratio of abnormal miners in BTC, LTC, MONA, ETH and BCH.

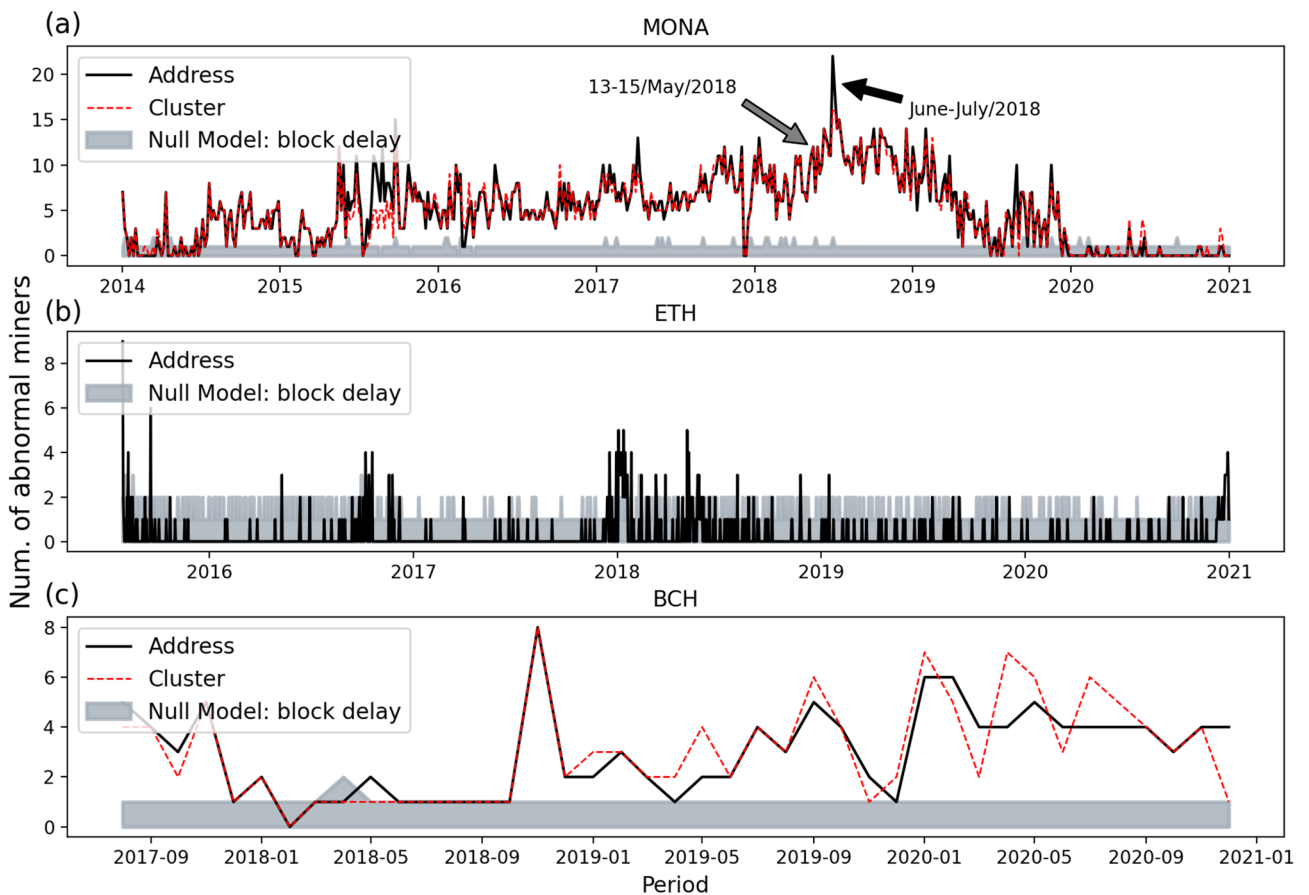


Figure 7. Number of abnormal miners during each time period in MONA, ETH and BCH. The grey shaded area highlights the 5% significance critical level under the null model of no selfish mining with network delay, obtained by bootstrap.

in that period. The empirical results of Monacoin (7a) show that the period with the most abnormal miners is around June-July 2018, which is near but much longer than the period 13-15 May 2018 when Monacoin announced they had suffered from a selfish mining attack. Besides, a part of miners might have been trying the selfish mining attack throughout time, not only during the mentioned periods. It seems that the selfish mining attack on Monacoin was contained after 2019 as we see a downward trend in the amount of abnormal miners. The result in Fig. 7c shows that several miners in Bitcoin Cash might try to conduct the selfish mining attack much more erratically, and a large number of abnormal miners appeared in Nov. 2018 in Bitcoin Cash, with still a few abnormal miners persisting into more recent years. Similarly in Ethereum (Fig. 7b), there were more abnormal miners at ETH’s launch, with SM attacks being more frequent in 2018 and occasionally occurring during the run time. Besides the selfish mining behaviours, network inefficiencies in relaying information could also cause

statistical abnormalities³⁰. Thus, to further verify the robustness of our detection method, we compare the detection results with a null model where there is no selfish mining but only a spurious autocorrelation generated by network delay. As shown by the grey area, which shows the 5% significance critical level under the network delay null obtained by bootstrap, in each subperiod of Fig. 7, only few abnormal miners should be expected, even in systems with low block intervals such as Ethereum and Monacoin. The details about the null model, including the parameters of network delay in different systems obtained by Fadda et al.³¹ and the influence of the network delay on the autocorrelation in the mining sequence, can be found in the Supplementary Information.

To further research the effect of increasing mining power on the potential of doing SM attack, we group active miners in each period by their corresponding hashing power in that period, and calculate the proportion of abnormal miners in each hashing power interval. In Fig. 8, one can find that in Monacoin the incidence of SM behaviour increases with miners' power when below 50% hashing power, and this increasing incidence also exists in Ethereum when below 30% hashing power, as well as in Bitcoin Cash below 25% power.

Network of mining cartels

In order to detect the existence of a mining cartel, where different miners share the information in advance among themselves and perform a coordinated selfish mining attack, we have extended our methods from testing single miners to pairs of miners. Considering pairs of miners i and j as a group ij , we conduct the similar hypothesis tests as above for each pair of miners and also calculate their corrected p -values, \hat{p}_{ij} in each period. Then we consider the pairs with $\hat{p}_{ij} < 0.05$ (but such that both \hat{p}_i and \hat{p}_j are greater than 0.05 in the given period) as potential cartels composed by miners i and j . After testing each pair of miners in the five cryptocurrencies, we show the network of identified mining cartels for each cryptocurrency in Fig. 9, where each node represents a pool (in BTC, LTC and BCH) or an address (in MONA and ETH) and each link represents an identified cartel between two miners. The weight (width) of each link is the number of periods this pair of miners has been detected as a cartel in all periods. The size of the node reflects the miner's average hashing power over all its active periods.

As shown in Fig. 9a,b, we find two abnormal cartels in Bitcoin and only one in Litecoin, and each of these cartels only includes two members. In Bitcoin Cash as shown in Fig. 9c, there are a few cartels, most of which are in a connected subgraph, and the four most powerful mining pools AntPool, BTC.com, ViaBTC and Bitcoin.com (the four biggest blue nodes) are fully connected with each other. We identified many abnormal cartels in both Monacoin (in Fig. 9d) and Ethereum (in Fig. 9e), but the connectivity of cartel networks in Monacoin and Ethereum is totally different. In Monacoin, a bit like in Bitcoin Cash, we find large cartels containing two or three powerful miners and many small miners, with a very high connectivity. In addition, there are also some separated small cartels with a few miners. However, the whole cartel network of Ethereum has a low connectivity. There are two separated large cartels, each of which contains several powerful miners, as well as a few cartels of varying size and with a generally low connectivity.

Discussion

The ledger of cryptocurrencies is always maintained through distributed consensus. Proof-of-work (PoW) is the most widely used consensus mechanism, maintaining the consistency of the system's ledger by requiring validators to solve an arbitrary mathematical puzzle to earn the right to verify transactions. Following the tremendous increase in market capitalisation of cryptocurrencies these years, developing defences against potential attacks on blockchain system has become an important topic. We considered the problem of selfish mining, one of the attacks which breaks information symmetry in blockchain systems, proposed by Eyal and Sirer in 2014. When employing the selfish mining (SM) strategy, malicious miners selectively keep their newly mined blocks temporarily private instead of publishing them immediately. To our knowledge, most of the previous studies on detection of selfish mining attacks are analytical models without empirical tests on real blockchain systems.

In this study, we proposed a statistical method to conduct empirical research on detection of selfish miners in five "PoW"-based cryptocurrencies, namely Bitcoin (BTC), Litecoin (LTC), Monacoin (MONA), Ethereum (ETH) and Bitcoin Cash (BCH). Regardless of whether SM actually leads to monetary gains or not, we emphasise that the strategy could lead to anomalies in the frequency with which a miner discovers successive blocks.

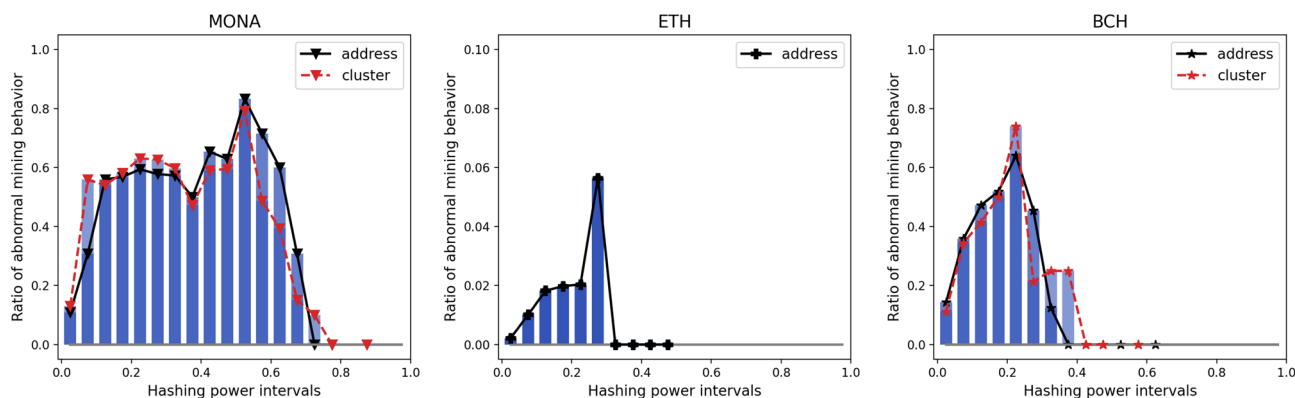


Figure 8. Fraction of abnormally behaving miners sorted by hashing power ranges in MONA, ETH and BCH.

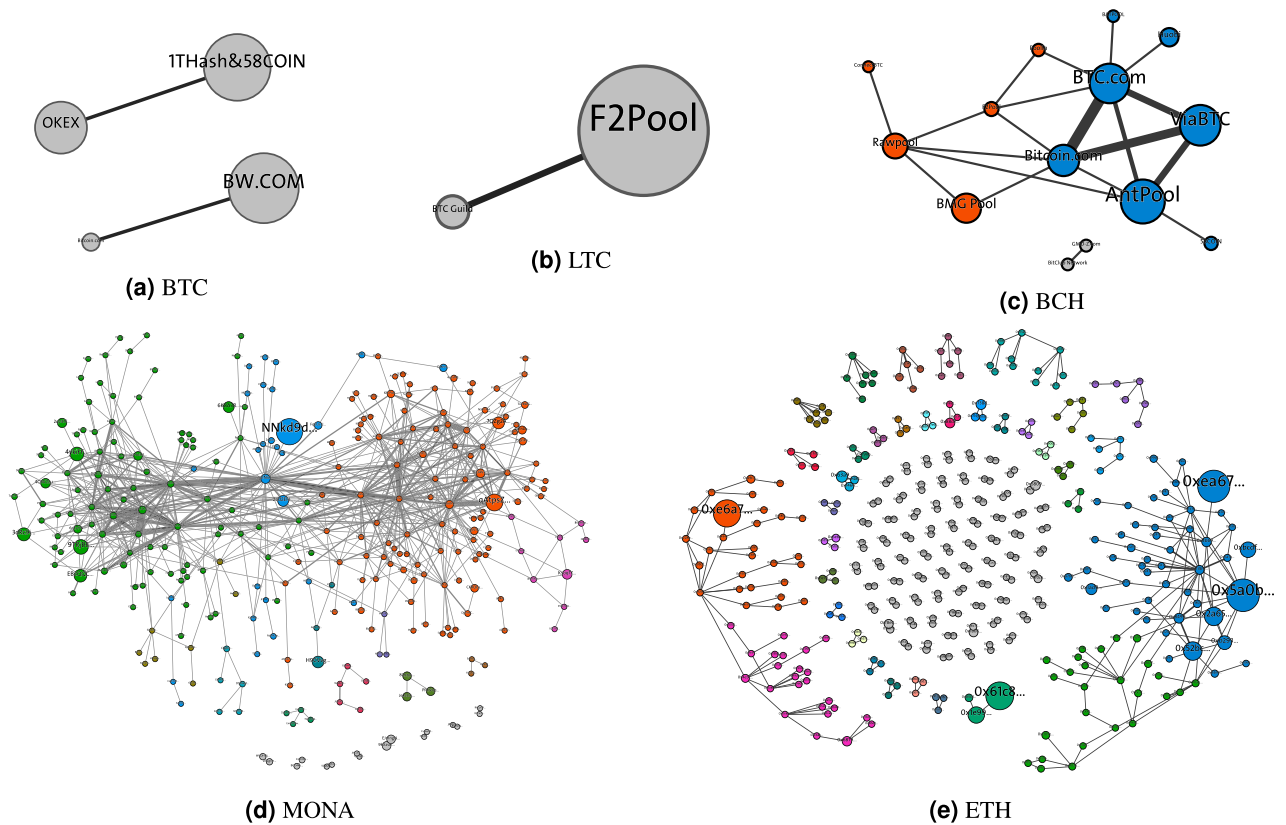


Figure 9. Networks of Mining Cartel in BTC, LTC, BCH, MONA and ETH.

We also investigated mining cartels, where miners secretly share information about new blocks among partners to pursue a collective SM strategy. Since the existence of mining cartels may generate threats to the security of blockchain-based systems, but has been ignored in previous studies, we also adapted our methods to test miners pairwise in order to detect at least some potential cartels. Our results suggest that although the SM strategy was proposed as an attack to the Bitcoin system, it was employed by more miners in Monacoin and Bitcoin Cash. In particular in Monacoin, out of all the different miners that have operated on the blockchain historically, about 50% were detected as potential selfish miners. Our detection results are consistent with Monacoin's own report about having suffered selfish mining attacks. We also detect more mining cartels in Monacoin, Ethereum and Bitcoin Cash compared with the two in Bitcoin and only one in Litecoin. The cartel network in Monacoin has a very high connectivity, but the cartels in Ethereum are more separated and most of them are in a tree structure. In addition to that, our results also show the importance of address clustering when conducting empirical studies in real blockchain systems.

There are some limitations to our work. First of all, selfish mining attacks and forming mining cartels are only two of the possible reasons for the anomalies we detect in miner's rates of successive block discoveries; alternatives include for example finite block diffusion times³². Secondly, we relied on the empirical frequency of mined blocks to estimate constant miners' hash rates within each time window: while this is the best estimator we can obtain from blockchain data, it is not necessarily accurate for several reasons, including that miners may vary their hash rate within a time window. This is expected to affect the test results, but we believe this can be overcome with further methodological research.

Future steps in this research include adapting the methodology to the more complex situation where miners have a variable hashrate within the time periods, relaxing the constant hashing power hypothesis, as well as including mining difficulty and cost as parameters. Another future development is the analysis of miners' (validators') selfish behaviour in cryptocurrencies that apply other consensus mechanisms^{33,34} using various detection methods (e.g. machine learning^{35,36} and edge computing-based method³⁷). In addition, our research sets the stage to investigate whether the "uncle block reward" could cause Ethereum to be more vulnerable to selfish mining attacks^{38,39}. Finally, our methods can also be applied as a forensics tool to characterise strategic mining behaviours, contributing to monitoring the security in current cryptocurrency ecosystems.

Methods

Anomalies in selfish mining attack

Following the "PoW" protocol, a miner's discovery of each block should be random and independent without any influence from the previous blocks, if the information diffuses through the network instantaneously³². Thus, during a certain time period where each miner's hashing power h_i is assumed constant, the event whether miner i

mined block t or not follows a Bernoulli distribution with probability h_i . However, when doing a strategic mining attack (e.g. selfish mining), the miners selectively publish their mined blocks to keep their leading height in block competition. This could lead to identifiable anomalies in statistics of successive blocks discovery.

How many consecutive blocks an attacker could mine is important to blockchain security and also to attack detection. As we can imagine when selfish miners keep mining on their private chain, they also take risks of losing the expected revenue. In the competition of solving hash-puzzle, in order to ensure a fair revenue, most of the attacks in PoW systems won't have a long private chain. According to the strategies of selfish mining⁸, when the private chain falls behind the public chain or the lead drops to 1, the attacker will immediately publish their private block. Furthermore, in the research on the alternative "stubborn mining"⁹, the authors did not observe any case where a selfish miner could earn more revenue if they don't merge with public when they fall behind by more than 1 block. In addition, there is a heuristic using the NS3 bitcoin simulator to detect malicious miners by observing the fork height⁴⁰. The results show that if the mean height of the fork is higher than 2, the blockchain system can be considered under selfish mining attack. All the arguments above then indicate that the length of a private chain would not be very long, usually no more than 2 blocks.

Therefore, in this paper the statistical analysis of selfish mining behaviour focuses on the case of the same miner mining *two* consecutive blocks. Although a selfish mining strategy may not significantly increase the proportion of blocks mined by strategic miners⁴¹, using our methodology we can detect the abnormal miners by testing the probability of miners' successive block discovery.

Probability of successive block discovery

Assume there are N miners in the system, identified by index $i = 1, \dots, N$. Define a random variable $X(t)$, where $t = 1, \dots, T$ is the block index and $X(t) = i$ if miner i mined block t . Assuming over the given time period the miner's hashing power h_i is constant, $X(t)$ is characterised by a multinomial probability distribution with unit size, $X(t) \sim \mathcal{M}(h, 1)$ where $P[X(t) = i] = h_i$ is proportional to miner i 's hashing power and, of course, $\sum_i h_i = 1$. The auxiliary random variable $Y_i(t)$ which is 1 if $X(t) = i$ and 0 otherwise then follows a Bernoulli distribution with probability h_i . We can then define our test statistic as the number of times c_i that the event $Y_i(t) = Y_i(t + 1) = 1$ has occurred, i.e. that miner i has mined c_i consecutive pairs of blocks among the T total blocks.

The probability distribution that characterises c_i is given by Ling²⁴. Indeed the random variable c_i follows what is called a *type II binomial distribution of order 2*, and the expression below (Equation 1) for the probability mass function is also given in the original paper. Calling $c_i^{(T)}$ the random variable c_i for a sequence of length T ,

$$P(c_i^{(T)} = x) = \begin{cases} h_i^T & \text{if } x = T - 1 \\ 2h_i^{T-1}(1 - h_i) & \text{if } x = T - 2 (> 0) \\ \sum_{j=x+2}^T h_i^{j-1}(1 - h_i)P(c_i^{(T-j)} = x - \max\{0, j - 2\}) & \text{if } 0 \leq x < T - 2 \end{cases} \quad (1)$$

In applying the above formula, Ling also put $P(c_i^{(T)} = 0) = 1$ if $T < 2$. In Fig. 10a, we show an example that probability distribution of variable c under different hashing powers h where the amount of blocks T is 100. The most probable value of miner's runs c increases with their hashing power, where a run is defined as two consecutively mined blocks.

Detection of abnormal miners

One of the main purposes of this paper is to obtain empirical evidence about whether selfish mining behaviours occur in practice or not. To achieve this purpose, we conducted hypothesis tests for every miner in various cryptocurrencies under the null hypothesis that the miner is honest, such that rejections of the null would identify a potential selfish miner in a certain period. Our null hypothesis means that miner i acts non-selfishly in compliance with the protocol, i.e. all blocks are mined randomly and independently. Under this null, the p -value is going to be the probability that miner i has at least c_i runs of two consecutively mined blocks occurring in a

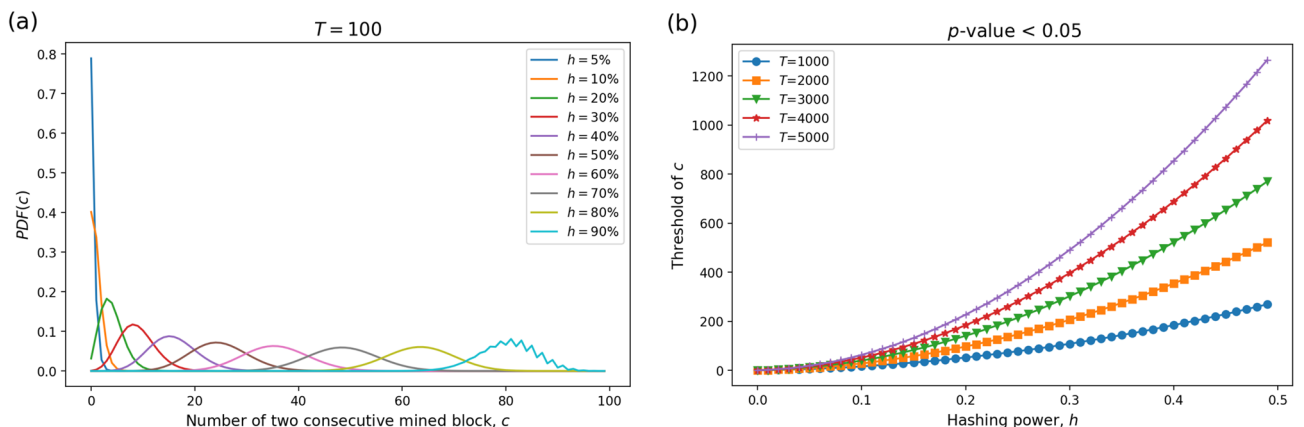


Figure 10. Illustrative diagrams of the statistical method used to detect abnormal mining behaviour in Proof-of-Work protocols.

sample of T blocks. Thus, the p -value corresponding to the observation of c_i consecutively mined block pairs is $p_i = P(x \geq c_i)$, or

$$p_i = \sum_{x=c_i}^{T-1} P(c_i^{(T)} = x) = 1 - \sum_{x=0}^{c_i-1} P(c_i^{(T)} = x) \quad (2)$$

To give better intuition, in Fig. 10b, we report the critical values c^* of the number of consecutively mined blocks at significance level $\alpha = 0.05$, for different values of mining power h and amount of blocks T . That is to say, for example in the purple line, when the amount of the block is 5000, the miner with less than 30% hash power but more than 491 runs of two consecutive blocks might conduct strategic mining behaviours within the 95% confidence interval.

When running multiple hypothesis tests the probability of obtaining one or more false positives (in this case identifying honest miners as abnormal) quickly becomes very high. For this reason it is important to adjust the p -values of each test to control for the False Discovery Rate (FDR), i.e. the expected fraction of false rejections among all rejected null hypotheses. We then adjust the p -values according to the procedure by Benjamini and Hochberg (BH)²⁹, where the corrected p -value reads

$$\hat{p}_k = \frac{p_k * T}{k} \quad (3)$$

with p_k being the k -th smallest p -value out of T total p -values in the test. After getting the \hat{p} for all the miners, we can reject the null with a 5% FDR for miners $k < k^*$, where k^* is the maximum k such that $\hat{p}_{k^*} < 0.05$, implying our results are expected to return a 5% rate of false positives.

Detection of mining cartels

Either an individual miner or a mining pool doing strategic mining could be considered as a single attacker. However, if some attackers share information earlier or only among themselves and collaborate to achieve the attack, they can be seen as a cartel. A mining cartel is then a secretly coordinating group where miners get together and privately share information about their mined blocks. In previous papers^{22,23}, Li et al. already pointed out that the existence of mining cartels has always been ignored so far. Considering a mining protocol secure as long as the pool's mining power is limited below a certain threshold always relies on the assumption that the miners (or pools) are operating independently. However, strategic miners may have incentives to associate in cartels, such as to benefit from the increased mining power and having information in advance about the blocks mined by the other members. On the other hand, detection of mining cartels contributes to revealing the potential relationships between attackers.

Based on the assumption that the collaboration in a cartel will cause the same anomalies of successive blocks discovery by cartel members, we run our test method on pairs of miners to detect potential cartels in five cryptocurrency systems. Therefore, we verify whether a cartel has formed between two miners i and j by measuring the anomalies in their consecutive blocks' statistics. Specifically, we use c_{ij} which is the number of times that two consecutive blocks is mined by the pair of miners i and j (regardless of the order), to replace c_i in Eq. 1. Likewise, we replace h_i by $h_{ij} = h_i + h_j$ which is the estimated aggregated mining power. As a result we can calculate the p -value of a pair of miners i and j , p_{ij} to which we also apply the usual FDR correction, and then use the corrected \hat{p}_{ij} -value to classify miners i and j as a cartel.

Of course one may identify a cartel because each miner is independently selfish. For this reason we only consider a cartel if neither of them is individually selfish, i.e. i and j form a cartel if $\hat{p}_{ij} < 0.05$, while $\hat{p}_i \geq 0.05$ and $\hat{p}_j \geq 0.05$.

Address clustering

The blockchain protocol adopted by all the analysed cryptocurrencies except Ethereum allows users to have more than one address linked to their wallet, which might be used to hide the track of their transactions and balances. Thus, accurately clustering together the different addresses of a miner is very important to estimate the mining powers and detect miner behaviour. We applied three known methodologies to the different cryptocurrencies in our dataset⁴²⁻⁴⁴ which are available from the blockchain analytics library BlockSci²⁵:

- **Heuristic 1 (H_1): Multi-input Addresses**
If two (or more) addresses are inputs to the same transaction, they are controlled by the same user.
- **Heuristic 2 (H_2): Optimal Change Address**
If the amount of an output address is lower than any of the inputs used in the transaction, then it is reasonable to assume that the output is used for the transaction change and is controlled by the same user as the inputs.
- **Heuristic p (H_p): Peeling chain**
A transaction is considered to be in a peeling chain if it includes one input and two outputs, and both the previous and the following transaction follow this structure. It is reasonable to state that the outputs linking the peeling chain are change addresses.

Specifically, in the implementation, when using the first heuristic method (H_1), different addresses used as inputs to one transaction are treated as being controlled by the same user. Then in H_2 , the identified so-called change addresses are treated as being controlled by the same user as the inputs. Finally in H_p , the “peeling chain” structure is used as a different definition to identify change addresses.

Dataset

Our empirical analysis focuses on five popular PoW-based cryptocurrencies which are Bitcoin (BTC), Litecoin (LTC), Monacoin (MONA), Ethereum (ETH) and Bitcoin Cash (BCH). Our dataset contains information of blocks from their launch to the end of 2020, including blocks’ height, mined time, the tag of corresponding miners (miner address/ pool name). In detail, datasets of Monacoin and Ethereum querying by BlockSci only have the miner address of each block, while the datasets (download from Blockchair: <https://gz.blockchair.com/>) of other three cryptocurrencies already have the labeled name of mining pools. In this research, we define time windows for each cryptocurrency depending on the block time to ensure a similar amount of blocks in each detection interval. We present a summary description of the five datasets in Table 1, and further enumerate each cryptocurrency for detailed introduction.

- **Bitcoin** was started on 3 January 2009 when the internet persona Satoshi Nakamoto mined the first (so-called *genesis*) block of the chain, known as the genesis block. Nowadays, this most famous digital asset has a rich and extensive ecosystem with a total market capitalisation of about 800 billions US dollars. About every 10 minutes, a new block is created and quickly published to all nodes, without requiring central oversight.
- **Litecoin** was a fork of the Bitcoin Core client released by Charlie Lee, a Google employee and former Engineering Director at Coinbase. The Litecoin network went live on 13 October 2011 differing primarily by having a decreased block generation time (2.5 minutes), increased maximum number of coins, different hashing algorithm, and a slightly modified GUI.
- **Monacoin** was a fork of Litecoin launched on 31 December 2013 by an anonymous person under the moniker of Mr. Watanabe. It bills itself as the first Japanese cryptocurrency and is predominantly used in Japan. Monacoin has an average block creation time of 1.5 minutes. Most notably, Monacoin was reported to have suffered from selfish mining attacks between May 13th and 15th in 2018 that caused roughly 90,000 dollars in damages²⁰.
- **Ethereum** is the second largest cryptocurrency after Bitcoin, with currently over 300 billions US Dollars market capitalisation. Ethereum is the blockchain that issues Ether and was proposed in late 2013 by Vitalik Buterin, a cryptocurrency researcher and programmer, and the system went live on 30 July 2015 featuring smart contract functionality. The block time of Ethereum is around 14 seconds. In 2022, Ethereum will be moving from PoW to proof of stake (PoS) as part of its “Ethereum 2.0” upgrade.
- **Bitcoin Cash** was a hard fork of Bitcoin that seeks to add more transaction capacity to the network. On 1 August 2017, Amaury Séchet released the first Bitcoin Cash software implementation. As in Bitcoin, new blocks are on average generated every 10 minutes by using a difficulty adjustment algorithm (DAA). Bitcoin Cash also uses an Emergency Difficulty Adjustment (EDA)⁴⁵, algorithm which has caused an instability in mining difficulty of the Bitcoin Cash system, resulting in Bitcoin Cash being thousands of blocks ahead of Bitcoin.

Coin	Launch time	Block time	Height of blocks	Interval	Miners
Bitcoin	2009-01-03	10 minutes	663913	Monthly	Named pool
Litecoin	2011-10-07	2.5 minutes	1974760	Weekly	Named pool
Monacoin	2013-12-31	1.5 minutes	2206670	5 days	Address
Ethereum	2015-07-30	14 seconds	11564743	Daily	Address
Bitcoin Cash	2017-08-01	10 minutes	189735	Monthly	Named pool

Table 1. Summary description of datasets.

Data availability

The datasets generated and analysed during the current study are completely publicly available in the Blockchair repository (<https://gz.blockchair.com/>). And the datasets are also available from the corresponding author on reasonable request.

Received: 19 August 2022; Accepted: 11 January 2024

Published online: 15 March 2024

References

1. Tasca, P. & Tessone, C.J. A taxonomy of blockchain technologies: Principles of identification and classification. *10.5195/ledger.2019.140* (2019).
2. Spychiger, F., Tasca, P. & Tessone, C. J. Unveiling the importance and evolution of design components through the tree of blockchain. *Front. Blockchain* <https://doi.org/10.3389/fbloc.2020.613476> (2021).
3. Nakamoto, S. Bitcoin: A peer-to-peer electronic cash system. *Decentralized Business Review* 21260 (2008).
4. Alangot, B., Reijnsbergen, D., Venugopalan, S. & Szalachowski, P. Decentralized lightweight detection of eclipse attacks on bitcoin clients. In *2020 IEEE International Conference on Blockchain (Blockchain)*, 337–342 (IEEE, 2020).
5. Apostolaki, M., Zohar, A. & Vanbever, L. Hijacking bitcoin: Routing attacks on cryptocurrencies. In *2017 IEEE symposium on security and privacy (SP)*, 375–392 (IEEE, 2017).
6. Karame, G.O., Androutaki, E. & Capkun, S. Double-spending fast payments in bitcoin. In *Proceedings of the 2012 ACM conference on Computer and communications security*, 906–917 (2012).
7. Bag, S., Ruj, S. & Sakurai, K. Bitcoin block withholding attack: Analysis and mitigation. *IEEE Trans. Inf. Forensics Secur.* **12**, 1967–1978 (2016).
8. Eyal, I. & Sirer, E.G. Majority is not enough: Bitcoin mining is vulnerable. In *International conference on financial cryptography and data security*, 436–454 (Springer, 2014).
9. Nayak, K., Kumar, S., Miller, A. & Shi, E. Stubborn mining: Generalizing selfish mining and combining with an eclipse attack. In *2016 IEEE European Symposium on Security and Privacy (EuroS & P)*, 305–320 (IEEE, 2016).
10. Liu, H., Ruan, N., Du, R. & Jia, W. On the strategy and behavior of bitcoin mining with n-attackers. In *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*, 357–368 (2018).
11. Sapirshstein, A., Sompolinsky, Y. & Zohar, A. Optimal selfish mining strategies in bitcoin. In *International Conference on Financial Cryptography and Data Security*, 515–532 (Springer, 2016).
12. Bai, Q. *et al.* A deep dive into blockchain selfish mining. In *ICC 2019–2019 IEEE International Conference on Communications (ICC)*, 1–6 (IEEE, 2019).
13. Zhang, R. & Preneel, B. Publish or perish: A backward-compatible defense against selfish mining in bitcoin. In *Cryptographers' Track at the RSA Conference*, 277–292 (Springer, 2017).
14. Solat, S. & Potop-Butucaru, M. Brief announcement: Zeroblock: Timestamp-free prevention of block-withholding attack in bitcoin. In *International Symposium on Stabilization, Safety, and Security of Distributed Systems*, 356–360 (Springer, 2017).
15. Heilman, E. One weird trick to stop selfish miners: Fresh bitcoins, a solution for the honest miner. In *International Conference on Financial Cryptography and Data Security*, 161–162 (Springer, 2014).
16. Nicolas, K., Wang, Y., Giakos, G. C., Wei, B. & Shen, H. Blockchain system defensive overview for double-spend and selfish mining attacks: A systematic approach. *IEEE Access* **9**, 3838–3857 (2020).
17. Hou, C. *et al.* Squirrl: Automating attack analysis on blockchain incentive mechanisms with deep reinforcement learning. arXiv preprint [arXiv:1912.01798](https://arxiv.org/abs/1912.01798) (2019).
18. Schwarz-Schilling, C., Li, S.N. & Tessone, C.J. Agent-based modelling of strategic behavior in pow protocols. In *2021 Third International Conference on Blockchain Computing and Applications (BCCA)*, 111–118, <https://doi.org/10.1109/BCCA53669.2021.9657011> (2021).
19. Schwarz-Schilling, C., Li, S.-N. & Tessone, C. J. Stochastic modelling of selfish mining in proof-of-work protocols. *J. Cybersec. Privacy* **2**, 292–310. <https://doi.org/10.3390/jcp2020016> (2022).
20. Saad, M., Njilla, L., Kamhoua, C. & Mohaisen, A. Countering selfish mining in blockchains. In *2019 International Conference on Computing, Networking and Communications (ICNC)*, 360–364 (IEEE, 2019).
21. Bonneau, J. *et al.* Sok: Research perspectives and challenges for bitcoin and cryptocurrencies. In *2015 IEEE symposium on security and privacy*, 104–121 (IEEE, 2015).
22. Li, S.N., Yang, Z. & Tessone, C.J. Mining blocks in a row: A statistical study of fairness in bitcoin mining. In *2020 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, 1–4 (IEEE, 2020).
23. Li, S.N., Yang, Z. & Tessone, C.J. Proof-of-work cryptocurrency mining: A statistical approach to fairness. In *2020 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*, 156–161 (IEEE, 2020).
24. Ling, K. On binomial distributions of order k . *Statist. Probab. Lett.* **6**, 247–250 (1988).
25. Kalodner, H. *et al.* Blocksci: Design and applications of a blockchain analysis platform. In *29th {USENIX} Security Symposium ({USENIX} Security 20)*, 2721–2738 (2020).
26. Vallarano, N., Tessone, C. J. & Squartini, T. Bitcoin transaction networks: An overview of recent results. *Front. Phys.* **8**, 286 (2020).
27. Campajola, C. *et al.* The evolution of centralisation on cryptocurrency platforms. arXiv preprint [arXiv:2206.05081](https://arxiv.org/abs/2206.05081) (2022).
28. Campajola, C., D'Errico, M. & Tessone, C.J. Microvelocity: rethinking the velocity of money for digital currencies. arXiv preprint [arXiv:2201.13416](https://arxiv.org/abs/2201.13416) (2022).
29. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
30. Gervais, A. *On the security, performance and privacy of proof of work blockchains*. Ph.D. thesis, ETH Zurich (2016).
31. Fadda, E., He, J., Tessone, C. J. & Barucca, P. Consensus formation on heterogeneous networks. *EPJ Data Sci.* **11**, 34 (2022).
32. Decker, C. & Wattenhofer, R. Information propagation in the bitcoin network. In *IEEE P2P 2013 Proceedings*, 1–10 (IEEE, 2013).
33. Neuder, M., Moroz, D.J., Rao, R. & Parkes, D.C. Selfish behavior in the tezos proof-of-stake protocol. arXiv preprint [arXiv:1912.02954](https://arxiv.org/abs/1912.02954) (2019).
34. Li, S.-N., Spychiger, F. & Tessone, C. J. Reward distribution in proof-of-stake protocols: A trade-off between inclusion and fairness. *IEEE Access* **11**, 134136–134145 (2023).
35. Shafiq, M., Tian, Z., Sun, Y., Du, X. & Guizani, M. Selection of effective machine learning algorithm and bot-iot attacks traffic identification for internet of things in smart city. *Futur. Gener. Comput. Syst.* **107**, 433–442 (2020).
36. Shafiq, M., Tian, Z., Bashir, A. K., Du, X. & Guizani, M. Corrauc: A malicious bot-iot traffic detection method in iot network using machine-learning techniques. *IEEE Internet Things J.* **8**, 3242–3254 (2020).
37. Yu, X., Yang, X., Tan, Q., Shan, C. & Lv, Z. An edge computing based anomaly detection method in iot industrial sustainability. *Appl. Soft Comput.* **128**, 109486 (2022).
38. Feng, C. & Niu, J. Selfish mining in ethereum. In *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, 1306–1316 (IEEE, 2019).

39. Ritz, F. & Zugenmaier, A. The impact of uncle rewards on selfish mining in ethereum. In *2018 IEEE European Symposium on Security and Privacy Workshops (EuroS &PW)*, 50–57 (IEEE, 2018).
40. Chicarino, V., Albuquerque, C., Jesus, E. & Rocha, A. On the detection of selfish mining and stalker attacks in blockchain networks. *Annal. Telecommun.* **75**, 143–52 (2020).
41. Wright, C.S. & Savannah, S. The fallacy of the selfish miner in bitcoin: An economic critique. Available at SSRN 3009466 (2017).
42. Meiklejohn, S. *et al.* A fistful of bitcoins: characterizing payments among men with no names. In *Proceedings of the 2013 conference on Internet measurement conference*, 127–140 (2013).
43. Androulaki, E., Karame, G.O., Roeschlin, M., Scherer, T. & Capkun, S. Evaluating user privacy in bitcoin. In *International conference on financial cryptography and data security*, 34–51 (Springer, 2013).
44. Ermilov, D., Panov, M. & Yanovich, Y. Automatic bitcoin address clustering. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 461–466 (IEEE, 2017).
45. Kwon, Y., Kim, H., Shin, J. & Kim, Y. Bitcoin vs. bitcoin cash: Coexistence or downfall of bitcoin cash? In *2019 IEEE Symposium on Security and Privacy (SP)*, 935–951 (IEEE, 2019).

Acknowledgements

S.N.L. acknowledges funding from the China Scholarship Council (CSC) (No.201808310212). S.N.L. acknowledges funding from the Cardano Foundation. C.C. acknowledges support from the Swiss National Science Foundation grant #200021_182659.

Author contributions

C.J.T and S.N.L. conceived the experiment(s). C.C and S.N.L. developed the methodology, S.-N.L. and C.J.T retrieved the data. S.N.L. performed the data analysis. S.N.L. and C.C wrote a first draft. All authors completed and reviewed the final version of this paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-55348-3>.

Correspondence and requests for materials should be addressed to C.J.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024