

Equilibrium analysis of block withholding attack: an evolutionary game perspective

Yukun Cheng*, Zhanghao Yao, Zhiqi Xu and Xiaojing Ye

Suzhou University of Science and Technology, Suzhou, China

* Correspondence author; E-mail: ykcheng@amss.ac.cn.

Abstract: With the development of blockchain technology, blockchain-based digital cryptocurrencies, such as Bitcoin, have received broad interest. Due to the permissionless environment, the blockchain is vulnerable to different kinds of attacks, including the block withholding (BWH) attack. BWH attack is one of the common selfish mining attacks, by which the attacking pool infiltrates the attacked pool by sending some infiltrating miners. These infiltrating miners try to find (partial) proof-of-work solutions and share the reward of the attacked pool, but discard the solution that can actually create blocks. Therefore, the attacking pool benefits by withholding blocks from the infiltrating miners, while damaging the benefits of victim pools. In this paper, we introduce the reward reallocation mechanism by paying additional rewards to the miners who successfully mine blocks, and propose an evolutionary game model for BWH attack among pools to study the strategy selection of pools. By constructing the replicator dynamic equations, the evolutionary stable strategies of pools are explored based on different levels of additional rewards. Our results provide enlightening significance to mitigate the negative influence from BWH attacks in practice.

Keywords: block withholding attack; evolutionary game; mitigation measure; blockchain

1. Introduction

Bitcoin, being one of the most popular blockchain applications, has risen in popularity. Satoshi Nakamoto [1] first proposed Bitcoin as a digital crypto-currency. Bitcoin system applies Proof of Work (PoW), one consensus protocol, to secure transactions [2]. PoW protocol requires that the nodes in the network (also known as miners) employ their computing power to solve a complex SHA 256 mathematical problem that is difficult to solve but simple to verify. The first miner to solve this problem has the right to extend the blockchain by a new block and get rewards when this block has been validated by the majority of miners in the network. Generally, the process of generating a block is called mining.

In the bitcoin system, one block is typically generated every ten minutes. With the growth of computing power in the whole network, the bitcoin system automatically increases the difficulty of block generation to preserve this time interval. It means that a single miner must consume more computing power or spend more time creating a block. Therefore, more and more miners decide to join the open mining pools to cooperate in mining by aggregating their computing power [3], and then share corresponding revenue once someone successfully creates a block. Usually, an open mining pool consists of a



Copyright©2022 by the authors. Published by ELS Publishing. This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium provided the original work is properly cited

manager and a group of miners. The manager assigns the mining work to the miners. All miners in this mining pool need to complete the work and submit a partial proof of work (PPoW) to the manager. Solving a PPoW is easier than the work that meets Bitcoin's difficulty goal. Miners also have the probability of directly finding a hash that meets the bitcoin network's difficulty goal, which is named full proof of work (FPoW). When a miner generates a FPoW, he submits it to the manager, who then sends it to the bitcoin system. As a result, the pool receives the entire revenue of a newly discovered block, which is distributed by the manager depending on each miner's contribution, including PPoW miners.

When a miner in a mining pool only submits PPoW and withholds FPoW, it means this miner is launching a BWH attack. Specifically, if BWH attack is launched among mining pools, the attacking pool shall send some infiltrators to the victim pool. The attacking pool also acts as an attacker in addition to receiving blocks and releasing rewards [4]. The ratio of the infiltrator's computing power to the computing power of the attacking mining pool is defined as its *infiltration rate*. The infiltrators receive tasks from the victim pool. When the attacking pool's infiltrators find the PPoW, they submit it to the victim pool. Once finding the FPoW, the infiltrators discard it since the pool protocol requires all blocks discovered by miners to be submitted to the pool manager. The victim pool is not aware of this discarding behavior and still distributes the revenue to all miners, which also allows the attackers to benefit from it. Therefore, the existence of the attackers does not bring any contributions to the victim pool, but reduces the revenue of other honest miners in the victim pool.

Therefore, BWH attack influences the mining pools' revenue from the following three perspectives.

- The effective computing power of the attacking pool is reduced because the infiltrators take away their computing power to the victim pool to launch the BWH attack.
- The victim pool's effective computing power remains the same, but the number of miners sharing revenue grows, resulting in the lower revenue for honest miners.
- The difficulty of block generation is increased because the total effective computing power of the system is reduced.

In this paper, we study the BWH attack among mining pools by establishing an evolutionary game model. Similar to the previous works in [4, 5], we study the symmetric case containing two mining pools to simplifying our discussion. In our discussion, each pool has the same computing power, and may play BWH attack or not, and thus has two strategies. One is honest mining, and the other is launching BWH attack. The whole of mining process contains multiple rounds, in each of which one block is mined and verified and is attached on the blockchain. In addition, in one mining round, each of the two pools shall decide and adjust its strategy by imitating the opponent's strategy. To mitigate the negative effect from BWH attack, we introduce a reward mechanism to encourage the miners who submit the FPoW. To be specific, when a mining pool obtains a revenue from successful mining, then it shall firstly assign a part of revenue to the miner who submits the FPoW, and then allocate the rest of revenue to all miners, including the infiltrators from the opponent pool, proportional to their computing power. Our numerical experiments show that such an additional reward mechanism can effectively mitigate the BWH attack by properly setting the portion of reward to the honest miner.

1.1. Related work

As Bitcoin has become more popular, the number of attacks on the Bitcoin system has increased. Such attacks can be divided into two categories: one is the attack on the

Bitcoin network, like DDoS attacks. And the other is the attack on the Bitcoin protocol itself. The damage of attacks on the protocol itself is often more serious, and typical attacks include selfish mining attack [6], block withholding (BWH) attack [4], fork after withholding (FAW) attack [7]. Selfish mining attacks allow attackers to release their blocks at particular times to increase profits. BWH attack, on the other hand, reduce the total revenue of the attacker's mining pool by discarding FPoW. FAW attack is when the attacker forks the blockchain to gain revenue. This article mainly focuses on BWH attack. BWH attack was first proposed by Rosenfeld [8]. The attacker participates open mining pool, submits only PPoW to the manager and withholds FPoW, damaging both his own and the pool's revenue. According to the different purpose of launching BWH attacks, BWH attack can be categorized into two types. One is to seek financial gain when a double-spend occurs, which is called *Finney Attack* [9]. The goal of the other type is studied in this paper, which causes financial damage to mining pools. Courtois and Bahack expanded the strategy selection of attackers who launch BWH attacks in [10], arguing that attackers can freely allocate their computation power to mine and launch BWH attack at the same time and demonstrating that higher rewards are achieved in this situation. Luu *et al.* [11] developed a computational power-splitting game model to analyze revenue under various computation power distributions. They demonstrated that attackers are encouraged to launch a BWH attack in an equilibrium state. Eyal [4] investigated a game between two mining pools to evaluate the Nash equilibrium of BWH attack, and discovered that miners would confront the miner's dilemma. Alkalay-Houlihan and Shah proposed [12] a detailed analysis based for this game. They proved that this game always admits a pure Nash equilibrium, and the pure price of anarchy (PPoA) is at most 3. Done *et al.* [13] proposed SelfHolding attack as a combined attack model based on selfish mining and BWH attack, and demonstrate that it is possible for an attacker to obtain higher returns than the classical selfish mining attack using the SelfHolding attack strategy by analyzing the expected returns of the attacker under different computing power settings. Wang and Chen [14] gave a tight bound of $(1, 2]$ for PPoA, and showed the tight bound holds under a more general setting, in which infiltrators may betray. Chen *et al.* [15] proposed a revised approach to reallocate the reward to the miners. Instead of proportionally allocating the reward to all miners, a pool manager deducts a fraction from the reward to award the miner who actually mined the block. Under this setting, the authors proved that for any number of mining pools, no-pool-attacks is always a Nash equilibrium. Recently, many research focus on how to mitigate BWH attack. Bag *et al.* [16] improved the existing Bitcoin mining protocol to make it difficult for miners to distinguish between PPoW and FPoW so as to resist BWH attack. Schrijver *et al.* [17] proposed a game-theoretic model, in which a novel reward mechanism to encourage miners to submit blocks as soon as possible is introduced. Wu *et al.* [18] constructed a generalized model between the mining pool game to examine the dynamic payoffs of the participants. In the extension, the author proved that increasing information asymmetry can help avoid BWH attacks.

Evolutionary game theory is a common tool for constructing models and analyzing strategy decisions. Now, many scholars use evolutionary game theory to study BWH attack. Liu *et al.* [19] firstly applied evolutionary game to analyze the dynamics of mining pool selection. The authors studied the strategy evolution of miners in a case study of two mining pools, and the results show that computing power and propagation delay are the main factors affecting miners' election. Kim *et al.* [20] explored the evolutionary game model of the miner's dilemma and the change of the miner numbers in the pool over time under the BWH attack and formulate a revenue model for mining pools, but Kim's model ignored revenue from attacks. Then, Fujita *et al.* [5] studied the pool dynamic selection from the perspective of miners revenue, especially the revenue

from attacks between two mining pools. Their results showed that mining pools can attract miners to join by launching BWH attacks. On this basis, Tuo *et al.* [21] proposed a revised evolutionary game model and proved that only launching BWH attack may not help the pool to attract miners, and the attacking time is also important. These studies well reflect mining pools strategy selections under BWH attack from the perspective of evolutionary game. But what we need more is how to reduce or prevent this kind of attack. Cheng *et al.* [22] proposed a supervision and punishment measures and modeled the BWH attack between two miners in a mining pool as an symmetric evolutionary game and analyzes miners' decisions under different supervision and punishment. The results in [22] showed that mining pools can effectively mitigate BWH attacks by formulating proper penalties. Cheng *et al.* [23] later established an asymmetrical evolutionary game of two miners to study the decision-making of miners. The authors proved that the miner with more computational power tends to behave honestly, and a proper additional reward measure can help the pool manager to encourage miners to mine honestly.

Comparing with the previous research, we mainly study the BWH attacks from the perspective of mining pools in this paper. We introduce an approach to mitigate the negative effect from BWH attack by giving additional rewards to honest miners, and propose an symmetric evolutionary game model for BWH attack between two mining pools to study the strategy selection of pools.

1.2. Paper organization

The rest of this paper is organized as follows. Section 2 introduces the evolutionary game model for BWH attack among pools, the concept of evolutionary stable strategy, and so on. Some necessary assumptions are proposed firstly, and then the evolutionary stable strategies are explored under different situations. In Section 3, a series of numerical experiments are conducted to verify our theoretical results. The sensitivities of the additional reward mechanism and the infiltration ratio are also analyzed. The last section concludes this paper.

2. The evolutionary game model for BWH attack

2.1. Basic evolutionary game model

To study the changing process of the strategy selection, we model BWH attack among mining pools as an evolutionary game. Similar to the previous works in [4]-[5], a symmetric case of two mining pools are discussed in this paper. Therefore, each mining pool has the same computing power and has two strategies: *Attack* (denoted by A) or *Not Attack* (denoted by N). Thus the strategy profiles of two pools are: (N, N) , (N, A) , (A, N) and (A, A) . Before analyzing the payoff of different strategy profiles, the following assumptions are necessary for the model, which are similar to those in [23].

- (1) Consider two pools in the PoW-Based blockchain system, denoted by pool 1 and pool 2. Each pool initially has m units computing power, $m \geq 0$.
- (2) Each pool may conduct BWH attack by sending some miners with an amount of computing power, called *infiltration computing power*, to the other pool. Because of the symmetric model, we assume that a pool sends some miner to another pool at the same infiltration rate $\alpha \in (0, 1)$, and thus the infiltration computing power of the pool is αm .
- (3) The size of the mining power is determined by dedicated hardware facilities such as high-performance GPUs and electricity. Therefore, a greater computing power per unit of time means that miners have invested a high cost in hardware and

electricity. When a pool behaves honestly, the mining cost per unit time is denoted by C_1 . If a pool launches BWH attack, then its free-riding behavior could decrease its mining cost, which is denoted by C_2 , $0 \leq C_2 < C_1$.

(4) When a block is successfully appended on the blockchain by a pool, this pool would obtain an amount of revenue. Here let R , $R \geq 0$, be the mining revenue per unit of computing power and per unit time.

(5) Other pools in the blockchain system are assumed to mine honestly. If neither of two pools launches BWH attack, then the favorable mining environment will bring more revenue to the pools. Thus we assume that the revenue per unit computing power and per unit time increases to be γR , $\gamma > 1$. Note that in [4], the authors have shown that if pools can cooperate, then both revenues increase in the long run, but this is not an equilibrium. Because one pool can increase its own revenue by attacking and destroying this equilibrium for the case of both pools. Therefore, as long as one of the pools launches a BWH attack in a two-pool situation, then both of their revenues will go down to R .

(6) To encourage the miners to submit the FPoW, the pool first assigns a portion of revenue kR , $k \in (0, 1)$, as additional award to the miner who submits the FPoW for the current block mining. Then the rest of revenue shall be shared by all miners proportional to their computing power. Here k is called the the additional rewards ratio.

In the following, we go to analyze the payoffs per unit of time of the two pools for each strategy profile. Here the payoff of one pool is defined as the difference of the total revenue of its all miners, including the infiltrators, and its mining cost.

- Strategy Profile (N, N) . Under the strategy profile of (N, N) , neither of two pools conducts BWH attack. By assumption, each pool has more revenue under this case, which shall be $m\gamma R$ per unit of time. At the same time, as each pool mines honestly, the mining cost is equal to C_1 . Therefore, pool 1 and pool 2 both have payoff of $m\gamma R - C_1$.
- Strategy Profile (A, N) or (N, A) . Under the strategy profile (A, N) , pool 1 launches BWH attack by infiltrating αm units of computing power to pool 2, while pool 2 does not. At this time, the revenue of pool 1 comprises of two parts. One is the direct mining revenue of $(1 - \alpha)mR$. The other is the infiltrating income by attacking pool 2. For pool 2, although it has $(1 + \alpha)m$ units of computing power as αm units of computing power coming from pool 1, actually only m units of computing power are useful for pool 2. So pool 2 obtains mR revenue per unit of time, and has to assign a part of revenue to the infiltrators. To be specific, pool 2 firstly assigns kmR revenue to the miner who submits the FPoW, and then allocates the rest of revenue to all miners in pool 2, including the infiltrators from pool 1. Thus the infiltrators obtain the revenue from pool 2 is $(1 - k)mR \frac{\alpha m}{m + \alpha m}$. Above all, pool 1 gets its payoff of $(m - \alpha m)R + (1 - k)mR \frac{\alpha m}{m + \alpha m} - C_2$, and the payoff of pool 2 is $kmR + (1 - k)mR \frac{m}{m + \alpha m} - C_1$. The case of (N, A) is symmetric and then we have that the payoffs of pool 1 and pool 2 are $kmR + (1 - k)mR \frac{m}{m + \alpha m} - C_1$ and $(m - \alpha m)R + (1 - k)mR \frac{\alpha m}{m + \alpha m} - C_2$, respectively.
- Strategy profile (A, A) . Under the strategy profile (A, A) , both pools play the BWH attack by infiltrating αm units of computing power to each other, and only $(1 - \alpha)m$ units of computing power are useful for mining. After assigning $k(1 - \alpha)mR$ revenue to the miner who submits the FPoW, pool 1 shall allocate $(1 - k)(1 - \alpha)mR \frac{\alpha m}{m + \alpha m}$ revenue to the infiltrators from pool 2, and also obtain the same infiltrating revenue from pool 2. Thus the total revenue of pool 1 is $(1 - \alpha)mR$ and its payoff is $(1 - \alpha)mR - C_2$ as its mining cost is C_2 per unit time. Since the pay-

Table 1. Payoff matrix of the model.

Pool 1	Pool 2	
	Not attack(N)	Attack(A)
Not attack(N)	$m\gamma R - C_1$ $m\gamma R - C_1$	$kmR + (1 - k)mR\frac{m}{m+\alpha m} - C_1$ $(m - \alpha m)R + (1 - k)mR\frac{\alpha m}{m+\alpha m} - C_2$
Attack(A)	$(m - \alpha m)R + (1 - k)mR\frac{\alpha m}{m+\alpha m} - C_2$ $kmR + (1 - k)mR\frac{m}{m+\alpha m} - C_1$	$(m - \alpha m)R - C_2$ $(m - \alpha m)R - C_2$

off computation is the same for each pool, the payoff of pool 2 is also $(1 - \alpha)mR - C_2$.

Based on the above analysis and the aforementioned assumptions, the payoff matrix of two mining pools is shown in Table 1.

2.2. The stable solutions of the evolutionary game

In our game model, the two pools have different strategies and payoffs. Because each pool has the same amount of computing power, the evolutionary game is symmetric. Thus we assume the probability that a pool does not attack by x ($0 \leq x \leq 1$), and thus the probability to launch BWH attack is $1 - x$. As each pool constantly adjusts its probability of strategy selection by learning and imitating the behavior of the opponent with higher revenue, we are very interested in the convergence of the probability, which is thus the evolutionary stable solution (ESS), denoted by x^* .

Since the evolutionary game is symmetric, we denote the payoff of a pool to be U_1 , when it decides not to attack. And U_2 is used to denote the payoff of a pool, if it plays BWH attack. Based on the payoff matrix 1, we derive the expected payoff of a pool, when it does not conduct BWH attack.

$$U_1 = x(m\gamma R - C_1) + (1 - x)[kmR + (1 - kmR)\frac{m}{m + \alpha m} - C_1]. \quad (1)$$

When a pool chooses to attack, its expected payoff can be expressed as:

$$U_2 = x[(m - \alpha m)R + (1 - k)mR\frac{\alpha m}{m + \alpha m} - C_2] + (1 - x)[(m - \alpha m)R - C_2]. \quad (2)$$

Hence, the average expected payoff U_3 of this pool is:

$$U_3 = xU_1 + (1 - x)U_2. \quad (3)$$

By [24], the difference between the payoff by adopting this strategy and the player's average expected payoff determines the strategy's growth rate. Combining (1) and (3), we have the following growth rate.

$$F(x) = \frac{dx}{dt} = x(U_1 - U_3) = x(1 - x)[(\gamma - 1)mRx + \frac{(k\alpha + \alpha^2)mR}{1 + \alpha} - C_1 + C_2]. \quad (4)$$

Clearly, when the growth rate of a strategy is equal to 0, the participant will stably select this strategy at last, and thus this strategy is stable. So, by setting $F(x) = 0$, we have three fixed points:

$$x_1 = 0, \quad x_2 = 1, \quad x_3 = \frac{(C_1 - C_2)(1 + \alpha) - (k + \alpha)\alpha mR}{(1 + \alpha)(\gamma - 1)mR}. \quad (5)$$

Clearly, if x_3 is a probability, then it shall be in $[0, 1]$.

2.3. Evolutionary equilibrium analysis on the decision of BWH attack

In our symmetric discussion for BWH attack between mining pools, each pool needs to decide the infiltration ratio α , besides whether to attack. Thus the ratio of α is an important parameter decided by pools. In addition, we would like to mitigate the influence from BWH attack by imposing the additional reward mechanism to the miners who submit the FPoW. So the additional rewards ratio k is also a variable we need to determine.

In this subsection, we will explore the relationship of the infiltration ratio α and the reward ratio k by distinguishing three cases: $x_3 \leq 0$, $0 < x_3 < 1$ and $x_3 \geq 1$, according to (5).

- Case 1. $x_3 \leq 0$. Under this case, we have $\frac{(C_1-C_2)(1+\alpha)-(k+\alpha)\alpha mR}{(1+\alpha)(\gamma-1)mR} \leq 0$, as $x_3 \leq 0$. Because the probability of x shall be in $[0, 1]$, then there are two fixed points $x_1 = 0$ and $x_2 = 1$. By the condition $\frac{(C_1-C_2)(1+\alpha)-(k+\alpha)\alpha mR}{(1+\alpha)(\gamma-1)mR} \leq 0$, it is not hard to obtain that

$$k > -\alpha + \frac{C_1 - C_2}{\alpha m R} + \frac{C_1 - C_2}{m R}.$$

Based on (4), we have the phase diagram for this case, shown in Figure 1. When the slope of the tangent line at the intersection of the curve and the horizontal coordinate is negative in the dynamic evolution diagram, the point is an evolutionary stable strategy [25]. From the Figure 1, we can discover the evolutionary stable strategy is $x^* = 1$ in this case. This indicates that when the additional rewards ratio k is higher, the pool gains fewer when it attacks, which motivates the pool to mine honestly.

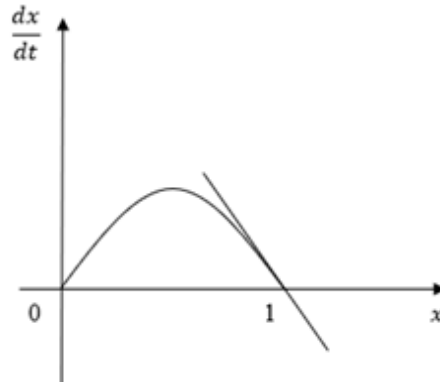


Figure 1. Replicator dynamic diagram for Case of $x_3 \leq 0$.

- Case 2. $0 < x_3 < 1$. According to the condition in this situation, we have $0 < \frac{(C_1-C_2)(1+\alpha)-(k+\alpha)\alpha mR}{(1+\alpha)(\gamma-1)mR} < 1$, i.e. $0 < x_3 < 1$. So there are three fixed points $x_1 = 0$, $x_2 = 1$ and $x_3 = \frac{(C_1-C_2)(1+\alpha)-(k+\alpha)\alpha mR}{(1+\alpha)(\gamma-1)mR}$. By $0 < x_3 < 1$, we have

$$-\alpha + \frac{[C_1 - C_2 - (\gamma - 1)mR]}{\alpha m R} + \frac{[C_1 - C_2 - (\gamma - 1)mR]}{m R} < k < -\alpha + \frac{C_1 - C_2}{\alpha m R} + \frac{C_1 - C_2}{m R}.$$

Based on (4), we have the phase diagram for this case, shown in Figure 2. Two points $x^* = 0$ or $x^* = 1$ may be the evolutionary stable strategies, which depends on the initial probability selected by the pool. To be specific, if the initial probability $x \in (0, x_3)$, then $F(x) \leq 0$, and the probability converges to $x^* = 0$ at last, meaning that the pool selects to attack eventually. When $x \in (x_3, 1)$, we have $F(x) \geq 0$, the pool finally decides not to attack.

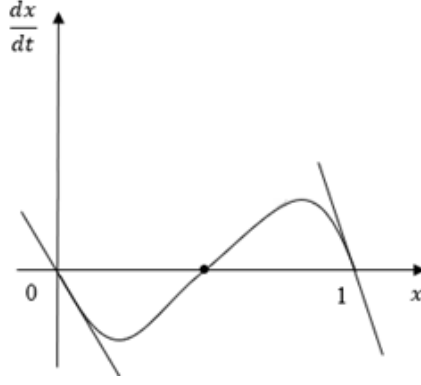


Figure 2. Replicator dynamic diagram for Case of $0 < x_3 < 1$.

- Case 3. $x_3 \geq 1$. As $x_3 \geq 1$, we have $\frac{(C_1 - C_2)(1 + \alpha) - (k + \alpha)\alpha m R}{(1 + \alpha)(\gamma - 1)m R} \geq 1$. There are two fixed points $x_1 = 0$ and $x_2 = 1$. Because $\frac{(C_1 - C_2)(1 + \alpha) - (k + \alpha)\alpha m R}{(1 + \alpha)(\gamma - 1)m R} \geq 1$, we can deduce that

$$0 < k < -\alpha + \frac{[C_1 - C_2 - (\gamma - 1)mR]}{\alpha m R} + \frac{[C_1 - C_2 - (\gamma - 1)mR]}{m R}.$$

Based on (4), we have the phase diagram for this case, shown in Figure 3, and the evolutionary stable strategy is $x^* = 0$. It demonstrates that if reward ratio k is relatively small, then the infiltrators get more from the victim pool, increasing the revenue of the attacker. So under this case, launching BWH attack is the evolutionary stable strategy of each pool.

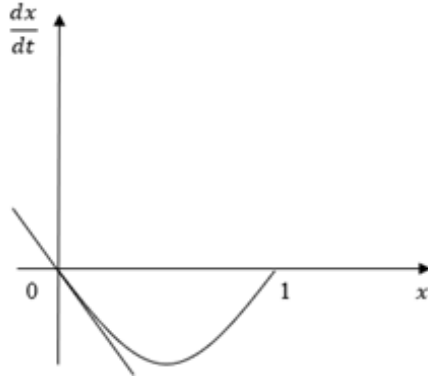


Figure 3. Replicator dynamic diagram for Case of $x_3 \geq 1$.

3. Numerical experiments

3.1. Simulations under different cases

In order to intuitively observe the evolution process of the pool strategy selection and to verify our results from above theoretical analysis, we conduct several experiments by using MATLAB.

- Case 1. $x_3 \leq 0$.
For the case of $x_3 \leq 0$, i.e.,

$$\frac{(C_1 - C_2)(1 + \alpha) - (k + \alpha)\alpha m R}{(1 + \alpha)(\gamma - 1)m R} \leq 0,$$

let us set $C_1 = 30$, $C_2 = 20$, $\gamma = 1.05$, $m = 2$, $R = 70$, $\alpha = 0.1$ and $k = 0.8$. Figure 4 depicts the dynamic evolution of a pool's strategy. In this case, $x^* = 1$ is

the pool's evolutionary stable strategy, and the pool eventually decides to mine honestly rather than launch a BWH attack.

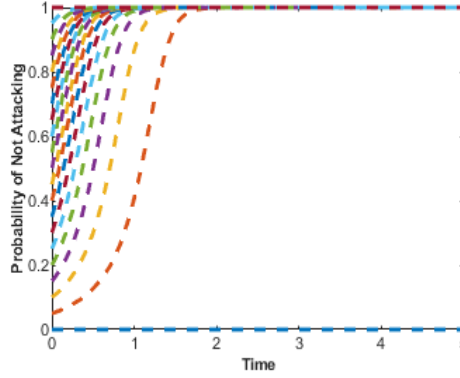


Figure 4. Dynamic evolution process of two pools when $x_3 < 0$

- Case 2. $0 < x_3 < 1$.

Let us set $C_1 = 30$, $C_2 = 20$, $\gamma = 1.05$, $m = 2$, $R = 70$, $\alpha = 0.1$, $k = 0.4$ to satisfy $0 < x_3 < 1$, i.e.,

$$0 < \frac{(C_1 - C_2)(1 + \alpha) - (k + \alpha)\alpha m R}{(1 + \alpha)(\gamma - 1)m R} < 1.$$

The dynamic evolution process is shown in Figure 5. We can see that $x^* = 0$ is the evolutionary stable strategy when the initial probability $x \in (0, x_3)$, and $x^* = 1$ is the evolutionary stable strategy when the initial probability $x \in (x_3, 1)$.

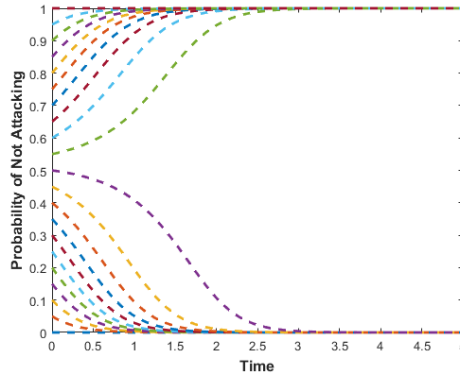


Figure 5. Dynamic evolution process of two pools when $0 < x_3 < 1$

- Case 3. $x_3 \geq 1$.

Let $C_1 = 30$, $C_2 = 20$, $\gamma = 1.05$, $m = 2$, $R = 70$, $\alpha = 0.1$, $k = 0.1$ to satisfy $x_3 \geq 1$, i.e.,

$$\frac{(C_1 - C_2)(1 + \alpha) - (k + \alpha)\alpha m R}{(1 + \alpha)(\gamma - 1)m R} \geq 1.$$

The dynamic evolution process is shown in Figure 6 and $x^* = 0$ is the unique evolutionary stable strategy of the pool. This implies that pool eventually chooses to launch the BWH attack.

3.2. Simulations for the influence of parameters

In this section, we analyze the sensitivity of the infiltration ratio α and the additional reward ratio k on the evolutionary process of pools. Under each different settings, the pool will converge to a specific strategy (*Attack* or *Not Attack*), which provides a insightful reference for the pool to set reward parameters. We also explore the effective reward mechanism to mitigate the negative effect from the BWH attack.

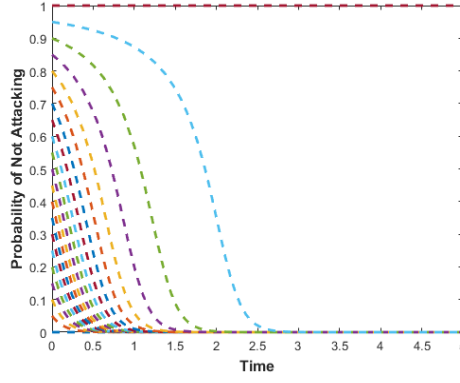


Figure 6. Dynamic evolution process of two pools when $x_3 > 1$

3.2.1. The influence of infiltration rate α on the evolutionary behavior of pools.

To see how α affects the pool's evolutionary behavior, we set $C_1 = 30$, $C_2 = 20$, $\gamma = 1.05$, $m = 2$, $R = 70$, $x = 0.5$, and α to be 0.1, 0.15, and 0.2, respectively. Figure 7a, 7b, 7c and 7d depicts the pool's strategy when α is set to 0.06, 0.1, 0.15 and 0.2, while k is set to 0.1, 0.2, 0.4, 0.6, 0.7, 0.8, respectively. When the infiltration rate is low and the additional reward ratio is high (e.g., $k = 0.8$), the pool finally decide not to attack, as seen in Figure 7a. And we can deduce from the Figure 7b, 7c and 7d that when the attacking pool's infiltration rate rises, a tiny additional incentive (e.g., $k = 0.1$) is capable of preventing the pool from attacking. Furthermore, we can see that when $k = 0.7$ or 0.8 , the evolutionary speed of pools towards no attack strategy rises. When we compare the curves for $k = 0.1$ (or $k = 0.2$) in the four graphs, we can observe that as the infiltration rate increases, the pool begins to abandon its attack strategy. This is because, when the marginal effect of attacks decreases, rational pools would abandon the Attack strategy in order to maximize revenue.

It is obvious from the above research that changes of α always have an influence on the pool's strategy choice. The greater the infiltration rate of the pools, the sooner the pools' strategy converges to the *Not Attack* strategy.

3.2.2. The influence of additional reward ratio k on the evolutionary behavior of pools.

We keep the other parameters the same as before, but change k to 0.1, 0.4 and 0.8, respectively. Figure 8a, 8b, 8c and 8d depict the pool's strategy options when k is set to 0.1, 0.4, 0.6 and 0.8, while α is set to 0.06, 0.08, 0.1, 0.12, 0.15, 0.2, respectively. As shown in Figure 8a, when the increased reward ratio is tiny, the pool with the higher infiltration rate abandons the attack strategy. Conversely, the pool with the lower infiltration rate continues to attack. This means that the pool's strategy with a high infiltration rate is vulnerable to the reward mechanism. By comparing Figure 8a, 8b and 8c, we can see that when $\alpha = 0.06$, the evolutionary speed of pools toward attack strategy begins to slow down. Figure 8c and Figure 8d show that when the extra reward ratio rises, the pool finally decide not to attack, even if some of them have a low infiltration rate. This is because the attack's income isn't enough to entice the pool to conduct the BWH attack.

As a result, we may conclude that the reward mechanism has a significant impact on pool strategy selection, particularly for pools with a high infiltration rate.

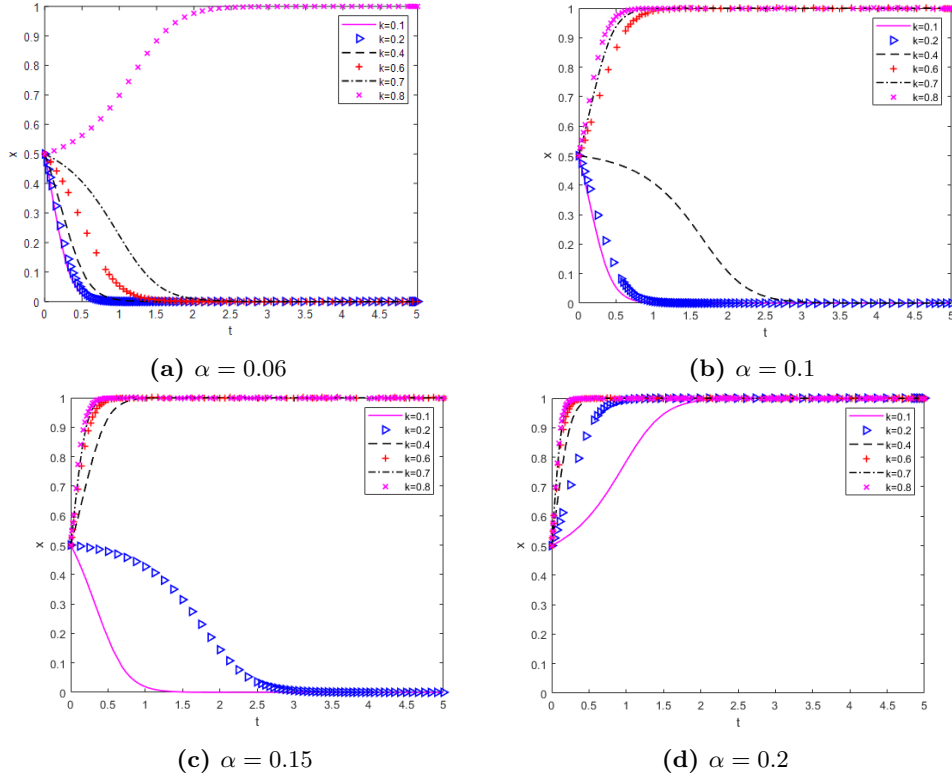


Figure 7. Impact of the infiltration rate α on the system evolution

4. Conclusions

In this paper, we construct a symmetric evolutionary game to study the evolution process of mining pools' behaviors under different levels of mitigation measures. By analyzing the evolutionary stable strategy, we discover that the higher additional reward to honest miners has a positive influence on pool's strategy selection. When the infiltration ratio is large, even a small reward also has a significant impact on the pool's revenue. In order to avoid loss, the pool will choose honest mining eventually. But this reward will hardly influence the pool's choice which has small infiltration ratio. We need to increase the reward to resist this kind of attack. Therefore, a mining pool can formulate an proper additional reward mechanism to prevent the BWH attack. There are also some limitations of the model considered in this article. Firstly, we study the symmetric evolutionary game in which each mining pool has the same computation power and the same infiltration ratio. Therefore exploring the asymmetric evolutionary game model is our future work. Especially, more general study shall be discussed in which a mining pool first decides to attack or not, and then decide its optimal infiltration ratio if it chooses to play BWH attack. Secondly, we only consider the case that two mining pools may launch BWH attack in this paper. Thus how to extend the study to the scenarios with more mining pools is very interesting and challenging.

Acknowledgments

This work is supported by the National Nature Science Foundation of China (No. 11871366), Qing Lan Project for Young Academic Leaders, Qing Lan Project for Key Teachers, and the Research Innovation Program for College Graduate Students of Jiangsu Province (No. KYCX22_3249, KYCX21_2998).

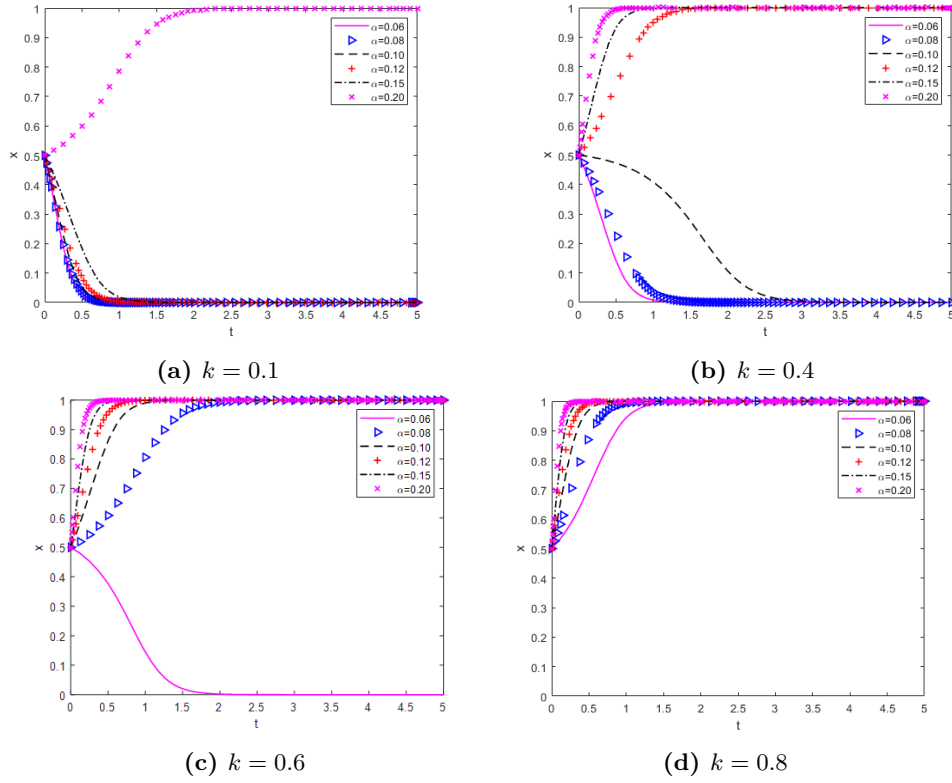


Figure 8. Impact of the additional reward ratio k on the system evolution

References

- [1] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. *Decentralized Business Review*, page 21260, 2008.
- [2] Florian Tschorsch and Björn Scheuermann. Bitcoin and beyond: A technical survey on decentralized digital currencies. *IEEE Communications Surveys & Tutorials*, 18(3):2084–2123, 2016.
- [3] Yi Liu, Xiayang Chen, Lei Zhang, Chaojing Tang, and Hongyan Kang. An intelligent strategy to gain profit for bitcoin mining pools. In *2017 10th International Symposium on Computational Intelligence and Design (ISCID)*, volume 2, pages 427–430. IEEE, 2017.
- [4] Ittay Eyal. The miner’s dilemma. In *2015 IEEE Symposium on Security and Privacy*, pages 89–103. IEEE, 2015.
- [5] Kentaro Fujita, Yuanyu Zhang, Masahiro Sasabe, and Shoji Kasahara. Mining pool selection under block withholding attack. *Applied Sciences*, 11(4), 2021.
- [6] Ittay Eyal and Emin Gün Sirer. Majority is not enough: Bitcoin mining is vulnerable. *Communications of the ACM*, 61(7):95–102, 2018.
- [7] Yujin Kwon, Dohyun Kim, Yunmok Son, Eugene Vasserman, and Yongdae Kim. Be selfish and avoid dilemmas: Fork after withholding (faw) attacks on bitcoin. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 195–209, 2017.
- [8] Meni Rosenfeld. Analysis of bitcoin pooled mining reward systems. *arXiv preprint arXiv:1112.4980*, 2011.
- [9] Janhvi Joshi and Rejo Mathew. A survey on attacks of bitcoin. In *International conference on Computer Networks, Big data and IoT*, pages 953–959. Springer, 2018.
- [10] Nicolas T Courtois and Lear Bahack. On subversive miner strategies and block

- withholding attack in bitcoin digital currency. *arXiv preprint arXiv:1402.1718*, 2014.
- [11] Loi Luu, Ratul Saha, Inian Parameshwaran, Prateek Saxena, and Aquinas Hobor. On power splitting games in distributed computation: The case of bitcoin pooled mining. In *2015 IEEE 28th Computer Security Foundations Symposium*, pages 397–411. IEEE, 2015.
 - [12] Colleen Alkalay-Houlihan and Nisarg Shah. The pure price of anarchy of pool block withholding attacks in bitcoin mining. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1724–1731, 2019.
 - [13] Xuewen Dong, Feng Wu, Anter Faree, Deke Guo, Yulong Shen, and Jianfeng Ma. Selfholding: A combined attack model using selfish mining with block withholding attack. *Computers & Security*, 87:101584, 2019.
 - [14] Qian Wang and Yurong Chen. The tight bound for pure price of anarchy in an extended miner’s dilemma game. *arXiv preprint arXiv:2101.11855*, 2021.
 - [15] Zhihuai Chen, Bo Li, Xiaohan Shan, Xiaoming Sun, and Jialin Zhang. Discouraging pool block withholding attacks in bitcoin. *Journal of Combinatorial Optimization*, 43(2):444–459, 2022.
 - [16] Samiran Bag, Sushmita Ruj, and Kouichi Sakurai. Bitcoin block withholding attack: Analysis and mitigation. *IEEE Transactions on Information Forensics and Security*, 12(8):1967–1978, 2016.
 - [17] Okke Schrijvers, Joseph Bonneau, Dan Boneh, and Tim Roughgarden. Incentive compatibility of bitcoin mining pool reward functions. In *International Conference on Financial Cryptography and Data Security*, pages 477–498. Springer, 2016.
 - [18] Di Wu, Xiang-dong Liu, Xiang-bin Yan, Rui Peng, and Gang Li. Equilibrium analysis of bitcoin block withholding attack: A generalized model. *Reliability Engineering & System Safety*, 185:318–328, 2019.
 - [19] Xiaojun Liu, Wenbo Wang, Dusit Niyato, Narisa Zhao, and Ping Wang. Evolutionary game for mining pool selection in blockchain networks. *IEEE Wireless Communications Letters*, 7(5):760–763, 2018.
 - [20] Seonggeun Kim and Sang-Geun Hahn. Mining pool manipulation in blockchain network over evolutionary block withholding attack. *IEEE Access*, 7:144230–144244, 2019.
 - [21] Jianyong Tuo, Guangyu Qin, Kailun Shang, and Youqing Wang. The mining selection in multi-pools under block withholding attack. In *2022 41st Chinese Control Conference (CCC)*, pages 7454–7459, 2022.
 - [22] YK Cheng and ZQ Xu. Study on the block withholding attack based on the evolutionary game. *Journal of Xidian University*, 47(5):1–11, 2020.
 - [23] Yukun Cheng, Zhiqi Xu, and Shuangliang Yao. The evolutionary equilibrium of block withholding attack. *Journal of Systems Science and Information*, 9(3):266–279, 2021.
 - [24] Daniel Friedman. On economic applications of evolutionary game theory. *Journal of evolutionary economics*, 8(1):15–43, 1998.
 - [25] Huawei Gong and Wenzhou Jin. Analysis of urban car owners commute mode choice based on evolutionary game model. *Journal of Control Science and Engineering*, 2015, 2015.