Countering Selfish Mining in Blockchains

Muhammad Saad  
University of Central Florida  saad.ucf@knights.ucf.edu

Laurent Njilla  
Air Force Research Laboratory  laurent.njilla@us.af.mil

Charles Kamhoua  
Army Research Laboratory  charles.a.kamhoua.civ@mail.mil

Aziz Mohaisen  
University of Central Florida  mohaisen@ucf.edu

Abstract—Selfish mining is a well known vulnerability in blockchains exploited by miners to steal block rewards. In this paper, we explore a new form of selfish mining attack that guarantees high rewards with low cost. We show the feasibility of this attack facilitated by recent developments in blockchain technology opening new attack avenues. By outlining the limitations of existing countermeasures, we highlight a need for new defense strategies to counter this attack, and leverage key system parameters in blockchain applications to propose an algorithm that enforces fair mining. We use the expected transaction confirmation height and block publishing height to detect selfish mining behavior and develop a network-wide defense mechanism to disincentivize selfish miners. Our design involves a simple modifications to transactions' data structure in order to obtain a “truth state” used to catch the selfish miners and prevent honest miners from losing block rewards.

I. INTRODUCTION

Blockchain technology has many applications, such as cryptocurrencies \[5,13,19\], smart contracts \[3,17\], Internet of things \[10,18\], health care \[8,15\], and supply chain management \[6\]. Blockchain applications use a constantly evolving distributed ledger that is capable of developing consensus in a decentralized environment. As the name suggests, a blockchain is a sequence of data blocks that are cryptographically chained to one another through one-way hash function. With the help of these mathematical constructs, blockchains employ an append-only model to prevent data tempering and preserve uniform consensus among peers in the network.

Despite these promising capabilities, blockchains are vulnerable to a series of attacks that evolve from its design constructs, its underlying peer-to-peer network, and the applications that make use of this technology. Some of the well known attacks on blockchains include the selfish mining attacks \[7\], block withholding attacks \[12\], the majority attack \[2\], distributed denial-of-service (DDoS) attacks \[12\], among others. The feasibility of each attack in blockchain applications varies depending on the network topology, adversarial requirements, peer behavior, and incentives. For example, in cryptocurrencies, it is more feasible to launch DDoS attack to exploit block size than to launch the 51% attack \[1\]. Per Baqer \et al. \[1\], the block size can be exploited by choking the network by excessively generating low-cost transactions. On the other hand, to launch the 51% attack, an adversary needs to acquire more than 50% of network’s hash rate to permanently gain control over the system.

Selfish mining is one such an attack that is considered to be infeasible in practice due to centralization of the mining pools, and the potential diminishing returns. However, in this paper, we argue that selfish mining may be more viable than commonly believed, and so the existing countermeasures are insufficient to prevent selfish mining attacks. We draw attention towards recent developments in the blockchain systems that have opened new attack avenues for selfish miners. Attackers may lease sufficient hash rate from online services to combine 51% attack with selfish mining and compromise the blockchain applications without being detected. In this paper, we also describe a threat model that partakes these new attack opportunities with a baseline attack procedure. We supplement our analysis by outlining the existing countermeasures and highlighting their limitations. We then propose a new scheme that utilizes a design combining various blocks of the prior solutions and provide more effective defense against selfish mining. We evaluate the workings of our model in light of our threat model and varying attack conditions. Our proposed scheme is effective in detecting the behavior of a selfish miner and encourages the network to discard his efforts.

Contributions. In this paper, we make the following contributions. 1) We describe a new form of selfish mining attack by outlining developments of new attack avenues in blockchain community with guaranteed rewards. 2) We empirically establish the feasibility of this attack by comparing high revenue guarantees with low attack cost. 3) We outline the limitations of simple and existing countermeasures and propose a new scheme for deterring selfish mining and promoting honest mining practices. 4) We validate our design effectiveness under varying attack conditions and against adaptive attackers.

Organization. The rest of the paper is organized as follows. In section \[II\] we review the prior work done to explore selfish mining and its countermeasures. In section \[III\] we outline the problem statement and preliminaries of our work. In section \[IV\] we describe the threat model and the attack procedure. In section \[V\] we present our scheme of countering selfish mining, followed by concluding remarks in section \[VI\].

II. RELATED WORK

Selfish mining in blockchains was identified by Eyal and Sirer \[7\], who demonstrated that mining protocols are not incentive-compatible and selfish miners may compromise the system and obtain higher rewards than their due shares. They used a state machine to outline the benefits of selfish mining to malicious miners. As a countermeasure, they proposed a random selection scheme at fork instance to disincentivize the selfish miner. However, in random block selection, the honest miner is equally likely to lose its block during fork and there is no guarantee that the honest miner will win under race conditions. Helima \[9\] proposed a “Freshness Preferred” (FP) technique in which blocks with recent timestamp are
preferred over old blocks. In FP, when a node is presented with blocks of an honest miner and a selfish miner, it selects the most recent blocks as identified by their timestamps. Although this is an effective technique to spot selfish mining behavior, the information flow in a peer-to-peer network is not always fair and unexpected delays in the propagation of honest blocks may favor the selfish miner.

Solat et al. [19] introduced Zeroblock, where miners are forced to release their blocks within an expected time. If the miners withhold their blocks for selfish mining and do not broadcast them within the expected time, the peers in the network create their own dummy blocks and append them to their blockchains. However, Zeroblock is not sustainable in varying hash rate of the network when the difficulty parameter is constant. Expected time of blocks may incur high variance due to hash rate fluctuations which, under Zeroblock, may invalidate valid blocks. In Figure 1 we plot the normalized values of expected block time and the actual block time in Ethereum. Due to varying hash rate or network latency, there is an expected delay in the actual block time which will result in invalidation of valid blocks in Zeroblock. Moreover, appending dummy blocks in blockchain creates an additional overhead on the blockchain size, which has reached beyond 163 GB and 450 GB in Bitcoin and Ethereum, respectively. Another method to combat selfish mining involves comparing timestamps of transactions in blocks of honest and selfish miners to discover selfish mining behavior. However, miners, as part of the standard practice, prioritize transactions based on fee. Aware of such countermeasures, a selfish miner may include recent transaction of low fee in his block and win the race against an honest miner who mines old transactions with high fee. Also, an adaptive attacker can include fewer or no transactions in his block to avoid the timestamp checking and still succeed in the attack. Therefore, there is a need for an effective deterrence mechanism to accurately distinguish between adaptive selfish and honest mining.

III. PROBLEM STATEMENT AND PRELIMINARIES

Over the years, only one selfish mining attack has been reported on a Japanese cryptocurrency called Monacoin. Such low prevalence of selfish mining in blockchains can be ascribed to the low probability of success, high attack cost, and low returns. Once a selfish miner finds a competing block in his private chain, he is forced into a race condition in which he competes with the hash rate of the rest of the network to extend his private chain before anyone else appends a block on the main blockchain. As such, the probability of finding a block before the rest of the network becomes a function of selfish miner’s hash rate and the aggregate hash rate of the network. To illustrate that, let the probability of success for the selfish miner be \( P(s) \). Let \( z \) be the number of blocks that the selfish miner wants to append to his private chain, \( \alpha \) be the fraction of selfish miner’s hashing power, and \( \gamma \) be the fraction of remaining hashing power, where \( \alpha + \gamma = 1 \). The success probability of selfish miner can then be defined as:

\[
P(s) = \begin{cases} 
1, & \alpha > \gamma \\
\left(\frac{\alpha}{\gamma}\right)^z, & \alpha < \gamma 
\end{cases}
\]

With low hash rate, a selfish miner may not succeed in launching an attack and is likely to lose the block rewards to the honest miner. To be able to win the race condition with guaranteed returns, the selfish miner needs at least 51% hash rate to succeed before the rest of the network finds a block. A combination of the 51% attack and the selfish mining attack will ensure selfish miner’s monopoly over the blockchain.

However, this strategy has two caveats. First, purchasing hardware to acquire a majority of hash rate in a major Bitcoin is expensive; the mining industry has moved from inexpensive CPU and GPU mining to sophisticated ASIC mining chips that are expensive. Second, as the 51% hash rate gives the attacker complete control over the network, it can easily be noticed by the network entities and they may discard all blocks published by the selfish miner. Due to these limitations, selfish mining attacks have not been observed frequently in blockchains.

However, we have noticed recent developments in the Bitcoin market that might facilitate selfish mining in disguise, and prevent peers in the network from discovering such fraudulent activity. Online hashing services, such as NiceHash, have emerged to outsource hashing power to the miners on hourly basis [14]. A selfish miner may rent up to 50% hashing power of a target Bitcoin for a short period of time and carry out successful selfish mining attack. In that case, the attack cost will be the money paid to NiceHash, and rewards will be the block rewards once the private chain is accepted. To calculate the profit of launching such an attack, let \( b \) be the block time of a Bitcoin in minutes, \( r \) be the block reward for publishing a block, and \( c \) be the cost of renting 50% hash rate of the Bitcoin for one hour. Then, the profit \( p \) for launching a successful attack of \( z \) blocks on that Bitcoin can be computed as:

\[
p = (z \times r) - \left(\frac{z \times b \times c}{60}\right)
\]

Using [2] we calculate the cost to launch a successful attack of two blocks in the top six cryptocurrencies, reported in Table 1. Notice that for each cryptocurrency, the block reward is always greater than the attack cost. To launch this
**TABLE I**

<table>
<thead>
<tr>
<th>SYSTEM</th>
<th>CAP</th>
<th>HASH RATE</th>
<th>COST</th>
<th>PROFIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bitcoin</td>
<td>$112.7B$</td>
<td>35,604 PH/s</td>
<td>$81K$</td>
<td>$69K$</td>
</tr>
<tr>
<td>Ethereum</td>
<td>$49.5B$</td>
<td>222 TH/s</td>
<td>$1.50K$</td>
<td>$1.6K$</td>
</tr>
<tr>
<td>B.Cash</td>
<td>$14.9B$</td>
<td>5,023 PH/s</td>
<td>$11.30K$</td>
<td>$5.4K$</td>
</tr>
<tr>
<td>Litecoin</td>
<td>$5.7B$</td>
<td>327 TH/s</td>
<td>$0.13K$</td>
<td>$3.6K$</td>
</tr>
<tr>
<td>Dash</td>
<td>$2.1B$</td>
<td>2 PH/s</td>
<td>$0.13K$</td>
<td>$1.4K$</td>
</tr>
<tr>
<td>Monero</td>
<td>$2.3B$</td>
<td>365 MH/s</td>
<td>$0.10K$</td>
<td>$0.9K$</td>
</tr>
</tbody>
</table>

**VIP. THREAT MODEL AND ATTACK PROCEDURE**

For our threat model, we assume a selfish miner capable of mining two or more blocks in race conditions. The aim of the selfish miner is to compute valid blocks and withhold them in a private chain to generate a fork against an honest miner. As a result, the attacker would want the network to switch to its longer private chain and discard the block of the honest miner. For that to happen, the attacker would want its private chain to be at least one block longer than the main blockchain to be able to convince the network for a longer proof-of-work and convince them to switch. On the other hand, we assume that the honest miner will follow the conventional mining practices, and will prioritize transactions based on their mining fee. He will try to include as many transactions in the block as possible to gain both block and fee rewards. Furthermore, the honest miner will not withhold his block and will timely broadcast it to the network upon computation.

**A. Baseline Attack**

The baseline attack procedure involves a selfish miner producing two blocks, $B_{S_1}$ and $B_{S_2}$, and forking the main blockchain to invalidate honest miner’s block $B_H$. The attacker rents 50% hash power of Bitcoin from NiceHash for 10 minutes. The attack sequence follows two rounds. In the first round, the attacker computes a first block, $B_{S_1}$, using his own hashing power. It then withholds the block and observes the honest miner’s block $B_H$ being accepted by the network.

In the second round, the attacker uses the rented hash power to compute the next block $B_{S_2}$ before anyone else in the network. Once the block is computed, the attacker forks the main blockchain with its private chain as illustrated in [Figure 2]. As a result, the network switches to the forked private chain of the selfish miner and discards the block of the honest miner. The selfish miner succeeds in the attack and wins more rewards than the cost incurred in the attack.

**V. COUNTERMEASURES**

To counter this attack, we introduce the notion of "truth state" for blocks at the fork instant to identify selfish mining behaviors. We append a parameter of “expected confirmation height” in the data structure of a transaction. In blockchains, the height of the block is the index number that denotes its position in the chain. A new block adds the height of the chain by factor of 1. Expected confirmation height is the index...
number of a future block in which the transaction is likely to be mined, and it depends upon the transaction size, the mining fee, and the size of the memory pool. The mining fee and the transaction size assign a priority factor to the transaction. The priority factor shows the incentive for a miner to select the transaction for his block. If the mining fee is high and the transaction size is small, miners are more inclined to prioritize that transaction for their block. Memory pool in blockchains is a repository that caches unconfirmed transactions. If the memory pool size is large, it creates a transaction backlog and pending transactions have to wait to be mined.

In Bitcoin, for example, an online service called “Earn” uses Monte Carlo simulation techniques to predict the expected confirmation height of a transaction with 90% confidence interval [4]. Their simulation parameters take the backlog of transactions, the fee priority of the miners over the last three hours, and the rate of the incoming transactions as inputs. Based on these parameters, Earn predicts the expected confirmation height and the expected delay for a given transaction. This prediction algorithm can also be applied on the software client of the users so that once a user generates a transaction, its software client can calculate the expected confirmation height of the block and append it to the transaction before broadcasting it to the network. Under standard mining, the transaction will likely be mined in the target block with 90% confidence. Thus, the average expected confirmation height of all transactions in the target block will be equal to the actual block height. This can be used to assign a “truth state” to the block and further be used to catch selfish miners who deviate from the standard mining. In the following, we elaborate on our design. We list notations in Table I and provide the description of our design in Algorithm 1.

### Table I: List of notations used in this paper

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_M$</td>
<td>Selfish miner</td>
</tr>
<tr>
<td>$H_M$</td>
<td>Honest miner</td>
</tr>
<tr>
<td>$F_S$</td>
<td>State of Fork</td>
</tr>
<tr>
<td>$N_S$</td>
<td>Normal State</td>
</tr>
<tr>
<td>$S_{state}$</td>
<td>Truth state for selfish miner</td>
</tr>
<tr>
<td>$F_{state}$</td>
<td>Future state for selfish miner</td>
</tr>
<tr>
<td>$H_{state}$</td>
<td>Truth state for honest miner</td>
</tr>
<tr>
<td>$n$</td>
<td>Total blocks computed by selfish miner</td>
</tr>
<tr>
<td>$B_{S_i}$</td>
<td>Selfish miner’s blocks, where $i = {1,2,\ldots,n}$</td>
</tr>
<tr>
<td>$X_{S_i}$</td>
<td>Height of each block $B_{S_i}$, where $i = {1,2,\ldots,n}$</td>
</tr>
<tr>
<td>$B_H$</td>
<td>Honest miner’s block</td>
</tr>
<tr>
<td>$Y_H$</td>
<td>Height of $B_H$</td>
</tr>
<tr>
<td>$T_{S_j}$</td>
<td>Transaction in $B_{S_j}$, where $j = {1,2,\ldots,p}$</td>
</tr>
<tr>
<td>$E(T_{S_j})$</td>
<td>Expected confirmation height of $T_{S_j}$</td>
</tr>
<tr>
<td>$T_{Y_k}$</td>
<td>Transaction in $B_H$, where $k = {1,2,\ldots,q}$</td>
</tr>
<tr>
<td>$E(T_{Y_k})$</td>
<td>Expected confirmation height of $T_{Y_k}$</td>
</tr>
</tbody>
</table>

### Algorithm 1: Detecting Selfish Mining Behavior

**State:** Fork on blockchain $F_S$

**Inputs:** $B_{S_1}, B_H$

1. $S_{state} = X_{S_n} - \left(\sum_{j=1}^{p} E(T_{S_j})\right)/p$; // Truth state for selfish miner
2. $F_{state} = X_{S_1} - \left(\sum_{j=1}^{p} E(T_{S_j})\right)/p$; // Future state for selfish miner
3. $H_{state} = Y_H - \left(\sum_{k=1}^{q} E(T_{Y_k})\right)/q$; // Truth state for honest miner
4. if ($H_{state} < S_{state}$ or $F_{state} < 0$)
   5. Reject $B_{S_i}$; // Reject selfish miner
   6. else
      7. ($H_{state} > S_{state}$); // Circumvention
      8. $A_{size} = 0$
      9. foreach $p \in B_{S_i}$ do
         10. $A_{size} = A_{size} + p$; // Compare number of transactions
         11. if ($q > A_{size}/n$ or $A_{size} = 0$)
            12. Reject $B_{S_i}$; // Reject if transactions size is small
         else
            14. Accept $B_{S_i}$;

**State:** Normal State $N_S$
An adaptive attacker may still circumvent detection as follows. 1) Include transactions with future expected block time in the first block to reduce the difference in the height of the latest block and the mean expected confirmation height of transactions in the first block. 2) Include fewer or no transactions in each block to achieve a higher truth state than the honest miner ($S_{\text{state}} < X_{\text{s}}$, and $S_{\text{state}} < H_{\text{state}}$).

To counter the first technique, we add a future state parameter $F_{\text{state}}$ in our algorithm (line 2), that verifies if the selfish miner has attempted to include transactions belonging to a future block in the current block. If the selfish miner does that, $F_{\text{state}}$ value will be less than zero, exposing the nature of transactions in each block. In algorithm 1 if the $F_{\text{state}}$ value is less than zero, then the private chain is rejected. To counter the second technique, we compare the number of transactions in the blocks of honest and selfish miner. If the average number of transactions in the selfish miner’s blocks are less than the honest miner’s block, it will expose that the selfish miner has tried to artificially achieve a higher truth state by publishing empty blocks. In our algorithm we reject the private chain if such fraudulent behavior is detected.

C. Exceptional Cases

Since mining is a lottery-based system, there might be instances where an honest miner finds two blocks within ten minutes and forks the network against another honest miner. In such a situation, the honest miner with the longer chain deserves to win the race condition and should not be accounted for selfish behavior. Our algorithm is flexible for such cases as long as standard mining practices are followed. The honest miner with the longer chain must have sizable transactions in each block and all of them need to have expected confirmation height close to the height of their respective block. If those conditions are met, the honest miner will win the race condition and his private chain will be accepted. Therefore, our algorithm ensures fairness even under such circumstances as long as the standard protocols are followed.

VI. CONCLUSION AND FUTURE WORK

In this paper, we introduce a form of selfish mining attack on blockchains, that guarantees high rewards with low cost. We outline the nature of this attack and show its profit margins in top six cryptocurrencies. We survey the prior work and discuss their approach and limitations. To counter this attack, we leverage honest mining practices to devise a notion of “truth state” for blocks during a selfish mining fork. We assign an expected confirmation height to each transaction to detect selfish mining behavior in the network. Our proposed algorithm effectively deters selfish mining and encourages fair mining practices. In future, we aim to estimate the fee overhead of appending the estimated confirmation height in each transaction as well as the processing overhead of applying our algorithm at the software client.

Acknowledgement. This work is supported by Air Force Material Command award FA8750-16-0301.