Can Blockchain be Trusted in Industry 4.0? Study of a Novel Misleading Attack on Bitcoin

Ghader Ebrahimpour, Mohammad Sayad Haghighi, and Mamoun Alazab

***Abstract*— After Bitcoin's emergence, blockchain found its way to many industries, including Fintech, Energy, and Manufacturing. Blockchains consensus algorithms, like Nakamoto's, are mechanisms to probabilistically guarantee that a transaction is not undone after confirmation. This mechanism requires that no one's computational power exceeds 50% of the network power. However, recent attacks on blockchains have raised serious questions about their security and whether they can be trusted to be employed in critical infrastructure and Industry 4.0. In this paper, we introduce a new category of blockchain attacks which we call "Misleading-Attacks". In this type of attack, a fraction of network power is misled so that the attacker reaches her goal. The technique is most effective when miners are rational and algorithm-oriented, similar to machines/agents in future Industry 4.0 or Industrial Internet of Things (IIoT). Moreover, this technique has the potential to be used in inventing new attacks, or can be used in combination with other known attacks. We first analyze a case in which the attacker uses misleading techniques to prevent her newly mined block from becoming orphan. We show that the proposed technique can push the attack success probability up by 16.42%. In a case study, we demonstrate how the technique promotes the success rate from 39% to 55.42%. Initiating the attack will be profitable if the attacker's power is more than 24% of the network power. By combining this novel technique with Bribery Attack, we show how the cost of *Guaranteed Variable-Rate Bribing with Commitment (GVC)* strategy can be drastically reduced.**

***Index Terms*— Blockchain, Industry 4.0, Critical Infrastructure, Misleading-Attack, Bribery Attack, Markov Chain**

I. Introduction

BITCOIN [1] was the first in the family of blockchains and created a virtual currency that uses peer-to-peer technology to facilitate online transactions without relying on trusted third parties. It proved to be more successful than any prior electronic cash in Fintech [2], [3]. Bitcoin nodes use a consensus algorithm to agree on the transactions recorded in a distributed shared database called blockchain [4]. Bitcoin consensus mechanism probabilistically guarantees that a transaction will not be undone after confirmation, assuming that attackers' computational power does not exceed 50% of the network mining power.

Soon the idea was picked by researchers and taken to other industries such as energy (for electricity exchange in smart grid) and manufacturing (for supply chain management). But recent attacks has cast doubt on the security of blockchain especially in sensitive areas such as critical infrastructure or Industry 4.0 where agents are rational and non-human.

The rise of mining pools has made %51 Attack feasible. $Krypton$ and $Shift$, two Ethereum-based blockchains, suffered 51% attacks in August 2016 [5]. $BitcoinGold$ is the another cryptocurrency which suffered a 51% Attack in 2018. This resulted in the theft of 18$ million worth of Bitcoin Gold. In 2014, $GHash.IO$ owned 54% of the computational power in the Bitcoin network [6], which raised serious concerns about the cryptocurrency's vulnerabilities. Morever, some have introduced attacks such as Block Withholding Attack (BWH) [7], Selfish Attack [8], etc. to abuse other aspects of blockchain protocol and eventually, reduced the minimum computational power required to make a profit out of an attack to 25% [8]. Eligius was an example of BWH victim pool in 2014 and lost 300 BTC back then. Since the attacker had used only two accounts for the attack, Eligius managed to find them [9].

Many of current blockchains, including the ones used in critical infrastructure or Industrial IoT, share the same foundation with Bitcoin and inherit similar problems. In this paper, we will introduce a novel attack which we tend to call *Misleading-Attack,* and show how it is initiated alone, and also how it can be combined with other known attacks to help them be more practical and/or reach their goal at lower costs. We assume miners are rational and have some incentives. We will design a new attack for orphaned blocks in which the attacker uses the proposed technique to prevent her block from going orphan.

Misleading-Attack can slow down the mining process for some malicious purposes, which is exactly the idea that we have used alongside Bribery Attack to increase attacker's gain. In general, any attack that requires a slow down in the mining process over a competitor chain can adopt Misleading-Attack. This novel technique opens a new category in the set of blockchain attacks, especially in machine-populated ecosystems like Industry 4.0. In simple words, this technique helps

Paper received Sep. 28, 2021, revised Nov. 27, 2021, accepted Jan. 4, 2022.

This work is based upon a research funded by Iran National Science Foundation (INSF) under grant No. 99022710. This research was also in part supported by a grant from IPM (CS1400-4-223).

Ghader Ebrahimpour is with the School of Electrical and Computer Engineering, University of Tehran,Iran, e-mail: g.ebrahimpour@ut.ac.ir.

M. Sayad Haghighi (corresponding author) is with the School of Electrical and Computer Engineering, University of Tehran, Iran, and also, the School of Computer Science, Institute for Research in Fundamental Sciences (IPM), P.o.Box 19395-5746, Tehran, Iran, e-mail: sayad@ut.ac.ir.

M. Alazab is with College of Engineering, IT & Environment, Charles Darwin University, Australia, E-mail: Mamoun.Alazab@cdu.edu.au.

attacker to remove a fraction of competitor's computational power in order to increase her chance to succeed during an attack. Eclipse attack [10], in which an adversary takes control of a sufficient number of IP addresses and partitions the network between the public and a specific miner (victim), falls into this category too. It is also used to turn away the computational power of a fraction of network. However, in Eclipse attack, attacker must have a large number of entities in network to partition it, which is not very practical in the current Bitcoin network. Misleading attack can be used for the same purposes, but with much less complexity. We will analyze Misleading Attack through Markov modelling. Markov model is probabilistic abstraction in which the probability of transiting to a new state depends only on the current state.

The rest of paper is organized as follows. In Section II related studies are reviewed. In Section III and IV, method of the attack and its analysis will be presented. Simulation results will be given in V. Other strategies and the future of the attack will be presented in Section VI. In Section VII, the effect of the technique on Bribery Attack will be analyzed and discussed. The paper is finally concluded in Section VIII.

II. Related Work

Industrial blockchains, including the ones employed in smart grids, IIoT or critical infrastructure, share the same foundation with classic blockchains. Therfore, we review the issue of blockchain security from a more generic layer.

The Bitcoin protocol was introduced in a white paper published in 2008 by Satoshi Nakamoto [1]. Different attacks on on Bitcoin have been proposed ever since. Double-Spending [11] is the most known attack in almost any cryptocurrency. Adaptive Double-Spending Attack (ADSA) [12] is an extended version of Double Spending Attack (DSA) in which an attacker uses adaptive strategy to increase her success probability. In adaptive strategy, the attacker starts a branch right after submitting a transaction. When the transaction becomes visible, she compares the length of main chain and her branch, and she leaves her chain and starts new branch in case the main chain grows longer than her old branch. A mathematical analysis on ADSA was done by [13]. It examined the effect of this attack on network parameters such as the required number of confirmation blocks and came to the conclusion that this number should be set higher when ADSA is probable (compared to DSA). The so called 51% Attack refers to the threat in which adversaries take control of the majority of the network computational power. A combined attack model is presented in [14] which increases the success probability of DSA by making delay on block distribution with the help of Sybil Attack [15] and, as a result, decreases the required computational power (to succeed in the attack) from 51% to 32%. Bastiaan [6] proposed a Two phase Proof-of-Work (2P-PoW) to prevent the formation of large pools, and as a result, stop 51% Attack.

Eyal and Sirer [8] proposed Selfish Mining in which a miner delays broadcasting newly founded blocks in order to mine a larger proportion of the blocks. The attacker should have 25% of the network's computational power to gain more reward than she deserves. A generalized strategy of Selfish-Mining attack was proposed [16], next, this strategy was combined with Eclipse Attack [10] to obtain higher gains. Göbel et al. evaluated Selfish-Mining in the presence of delays [17].

In Block Withholding (BWH) attack, which was proposed by Rosenfeld in 2011 [7], miners of a mining pool attack the pool itself by submitting only Partial Proof of Work (PPoWs), and not Full Proof of Work (FPoWs). BWH harms the pool and its honest members, but increases the revenue of malicious miners. Eyal, by modeling a game between two BWH pools, studied the Miners' dilemma in [18]. The result says that ”launching the BWH attack between two pools against each other, results in a loss for both in the equilibrium”. Kwon et al. [19] proposed a new attack called Fork After Withholding (FAW). This attack combines the BWH attack with intentional forks. The study proved that the amount of extra reward in the FAW attack is at least as high as what BWH achieves.

Several works analyzed transaction propagation in Bitcoin from the incentive perspective. Krombholz and et al. [20] showed that some malicious nodes in Bitcoin network tend not to propagate information (transactions) in the network. A mechanism was then suggested in the paper to fix this problem via incentivizing nodes to participate in information propagation. Using Bitcoin to send covert/secret messages is another security challenge which is studied in [21]. The authors of this study presented a new way of covert communication through Bitcoin's transactions. Blockchain-based steganography is claimed to be more resistant to tampering than traditional methods of steganography.

Using Bitcoin to send covert/secret data is more powerful than old steganography methods in case of data tampering Bribery Attack has been explored in [22]. In this attack, an adversary obtains the majority of computational power by bribing other miners. In [23] a practical example of Bribery Attack was introduced, which was later implemented in [24].

The authors of [25] used out-band bribing to decrease the cost of attack. In this approach, the attacker loses money if the attack fails. Sun and et al. [26] tried to model bribery attack when there are also honest miners involved. They assumed that the bribed miners were members of a pool (for better coordination). Teutsch et al. [27] presented another form of Bribery Attack in which the attacker uses smart contracts on platforms like Ethereum to issue spurious puzzles with her own payoffs in order to increase her proportion of the network computational power. Velner and et al. [28] introduces an attack against mining pools in which an attacker pays pool members to withhold their solutions from their pool manager.

Another attack scenario through incentivizing the miners to change their behavior has been proposed in [29] under the name of Whale Attack. In Whale Attack, attacker creates a new fork and issues whale transactions with high transaction fees which are valid on the forked chain. This is done to motivate miners to work on the attacker's chain. The paper's initial results showed the impracticability and inefficiency of Bribery Attack. However, in our previous work [30], we re-analyzed the attack in a generic way and proved that by taking proper and optimized strategies, one can launch the attack in practice with a much lower budget. In this paper, we present

TABLE I
SYMBOLS & NOTATIONS

Symbol	Description
A	The attacker (Alice)
X	The attacker's chain
B	The honest miners (plus Bob) mining on main chain
Y	The main chain
M	The misled miners
Z	The name of the misled miners' chain
$\mathcal{C}_k$	The computational power of $k \in \{A, B, M\}$

another technique with which an attacker can reduce the cost of Bribery Attack. We will bribe miners to mislead them so that the computational power on the main chain is reduced. We will show how this technique can boost several attacks.

III. The Proposed Misleading Attack

Suppose that Alice and Bob, who control a minority of computational power in the Bitcoin/blockchain network, have mined new blocks and distributed them in the network at almost the same time. Miners in the network will select one of these blocks to mine on. If the next block is found by the miners who are mining on Alice's block, Alice will earn her block reward, otherwise, she will lose it. In the latter case, Alice can initiate a Misleading Attack to revive the lost reward. Alice adopts the misleading technique because:

- Bribing other miners to mine on her fork needs more than the orphaned block's reward (based on the result of [30]) and makes the attack unprofitable for her. If the cost of attracting miners becomes more than the attack profit, initiating it will be unreasonable.
- Publishing her fork and attracting miners to her chain makes Bob adopt mitigating strategy to make the attack unprofitable for Alice.

In the proposed attack, Alice misleads a fraction of miners (or equivalently, their computational power) who are mining on Bob's chain (which we call the main chain) to drag them to a third chain so that her success probability in extending her private chain is increased. By considering the following rules, Alice designs her attack:

- The misled chain branching point should be picked in a way that the prospective resultant chain looks profitable for a fraction of the network computational power.
- The success probability in the third chain should be low to prevent Alice from sustaining potential big losses.
- Alice shall mine privately on her chain until her chain exceeds Bob's in length. Then, she publishes her blocks. Otherwise, rational miners will not accept Alice's bribe.

A. Misleading Attack Strategies and Analysis

In our analysis, we use the notations in Table I. We have:

$$\mathcal{C}_A + \mathcal{C}_B + \mathcal{C}_M = 1 \tag{1}$$

We refer to finding a new block as an 'event'. After launching the attack, Alice will use one of the following strategies upon the occurrence of each event:

1) Join the other chain and stop bribing ($JoinY$, $JoinZ$).
2) Continue mining on the current chain ($Cont$).
3) Continue mining on the current chain and increase the length of the misleading chain by one to make the chain profitable for misled miners ($AddZ$).
4) Continue mining on the current chain and add incentives on the misleading chain to make the chain profitable for the misled miners ($IncZ$).

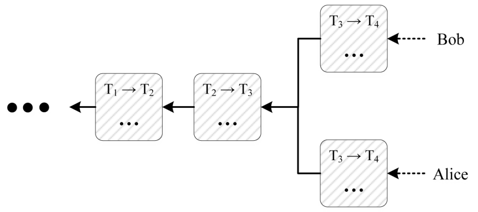


Fig. 1. Demonstration of the event that two blocks are found simultaneously.

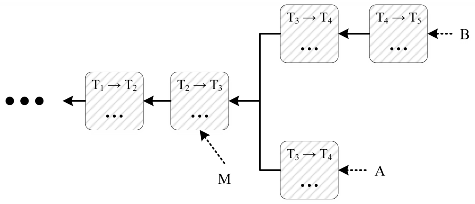


Fig. 2. "$\boldsymbol{A}$" misleads "$\boldsymbol{M}$" by publishing $\boldsymbol{T_3}$'s private key. $\boldsymbol{M}$ starts mining on the $\boldsymbol{T_2} \to \boldsymbol{T_3}$ block to gain the money withheld in $\boldsymbol{T_3}$.

When Alice sees that mining is not profitable on her chain anymore, she picks the first strategy. When an event does not make the attack unprofitable for Alice and it does not make mining on chain Z unprofitable for the misled miners either, she chooses the second strategy. She will use the third strategy when mining on chain Z is not profitable for misled miners. Finding any new block on chain Z needs new bribes to incentivize the misled miners to continue mining on Z (forth strategy). The attacker will join the main chain ($JoinY$) when one of the following conditions is met:

- Mining on chain X is not profitable anymore.
- The misled miners extended the chain Z in a way that can cause big losses for the attacker. The attacker will join the main chain to prevent the misled chain to succeed.

Fig. 1 shows the starting point of competition between Alice and Bob. In each block, there are transactions that are issued by Alice. Some or all of these (depending on Alice's calculations in the process of attack and by releasing the associated private keys) will be used as the bribe to mislead miners. We denote their source addresses by T_i and the corresponding destination addresses by T_{i+1}. The amount of each transaction can be obtained from a formula which will be explained in Section IV. In Fig. 2, the attack starts by Alice publishing the private key(s) of T_3. By mining on the block which has a transaction to T_3, miners can transfer the money to their accounts. Mining on this block is only profitable for a fraction of miners with a computational power of M.

At this point of attack, one of the following will happen:

- A finds a new block. She will continue to find other blocks to finish the attack successfully.

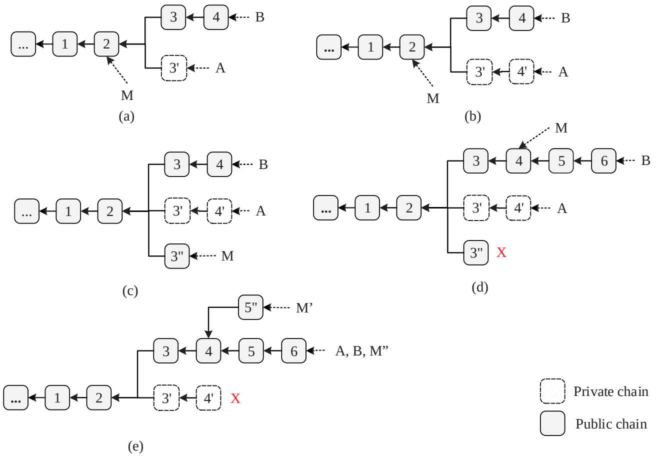


Fig. 3. a) A miner has found Block 4, and Alice started the attack by releasing the private key of a transaction in Block 2. b) Alice mined Block 4' but kept it private. c) The misled miners have found a new block and Alice defines new bribe in Block 3". d) Two blocks are found on main chain and Alice defines new a bribe in Block 4. e) Z exceeds X and Alice joins the main chain.

- B finds a new block. Mining on Z will not be profitable for M anymore. Alice will disclose the private key of T_4 to prevent M from joining Y. At this point, the length of Z will be equal to that of X. In next step, if B finds another block, the attack will fail, otherwise, if M finds new block, A will join Y because Z is exceeding X.
- M finds a new block. A will continue on her chain. If M finds another new block, A will join Y to prevent M from succeeding.

Fig. 3 shows an example with snapshots of consecutive events. In the figure, white and grey blocks represent private and public chains, respectively. In Fig. 3(a) a miner on Y has just found Block 4, so the attack is started by Alice to prevent Block 3' from going orphan. In Fig. 3(b) Alice has just found a new block but, keeps it private. In Fig. 3(c) the misled miners have found a new block. Since the money on T_3 is collected by the misled miners, based on [30], Alice should create new bribe to keep the misled chain profitable for them. In Fig. 3(d) two blocks have been found on Bob's chain which have made chain Z unprofitable for M. At this point, Alice can create new incentives in Block $3''$, or she can release the private key of a transaction in Block 4. She had better choose the latter option though, because adding incentives to $3''$ has two drawbacks: Firstly, it needs a high budget of BTC (maybe more than the block reward) to be released as the bribe, and secondly, if the misled miners finish the attack successfully, the amount of loss for Alice will be big. By revealing the private key of one or more transactions in Block 4 (whose value is enough for misleading the already-mislead miners), we assume that all misled miners will join the new chain. Hypothetically speaking, a miner who has found Block $3''$ may decide to keep mining on $3''$, but mining on $3''$ does not change the total computational power that is misled by the attacker and it remains at $\mathcal{C}_M$. This is enough for the attacked to proceed with her plan. In Fig. 3(e), the misled miners have found a block on $5''$ and the length of Z exceeds that of X (but it is till shorter than Y). In this case, Alice leaves her chain immediately and joins Y. Since Alice will not release new incentives for the misled miners to continue mining on Z, the probability of chain Z's profitability diminishes significantly (for the misled miners) and some (or all) will go back to Y. The ones who decide to remain on Z (with the power of $M' \leqslant M$), will mine on Z until they succeed or they find it unprofitable.

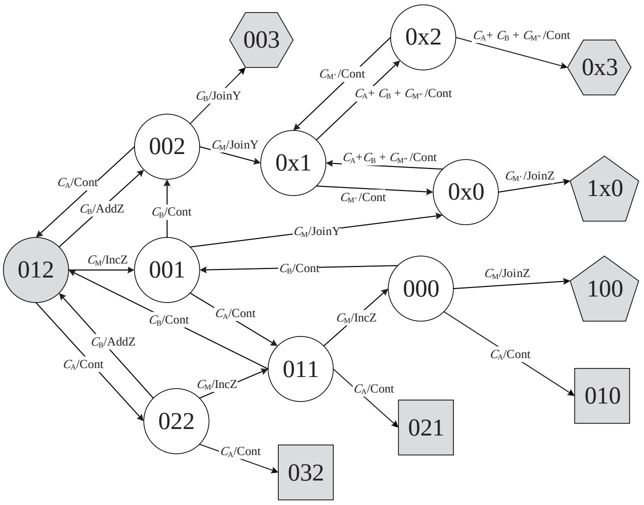


Fig. 4. Markov model of the events from the attacker's point of view.

Markov model of the attack is shown in Fig. 4. In this figure, each state is given a three-letter label like abc, where a is the difference of length between chain Z and the shortest chain, b represents the same quantity for chain X, and c measures this difference for chain Y. Notice that the shortest could be any of the three at different times. For example, in the starting state (012), Z is the shortest chain and X and Y have 1 and 2 blocks more than Z, respectively. Letter x in the label of states means that the corresponding chain length is not of importance to us, because there is no miner on the chain. We denote the (transition) probability matrix of Fig. 4 by Pr. In this figure, each absorbing state is shown by one of the following shapes:

- Square: The attack has finished successfully ($Success$).
- Hexagon: The attack has failed, but Bob has got his reward successfully ($Fail$).
- Pentagon: The attack has failed and Alice has lost her rewards as well as the money she paid as the bribe to misled miners. Bob has lost his reward too ($BigFail$).

In Fig. 4, we have:

$$\mathcal{C}_M = \mathcal{C}_{M'} + \mathcal{C}_{M"} \tag{2}$$

$\mathcal{C}_M = \mathcal{C}_{M'}$ results in an upper bound for $BigFail$. Fig. 5 shows the probabilities of attack success and failure. The power of misled miners (M) can be obtained using (1).

Fig. 6, which is a 2D cross-section of Fig. 5(a), shows the success probabilities in different situations. In the figure, $\mathcal{C}_M = 0$ means that Alice did not adopt the misleading technique, $\mathcal{C}_B = 0$ means Alice misled all miners (including Bob, even though this is an extreme case that does not happen in reality), and $Max\ Prob$ is the maximum value of success probability that Alice can reach. As shown in the figure, adopt-

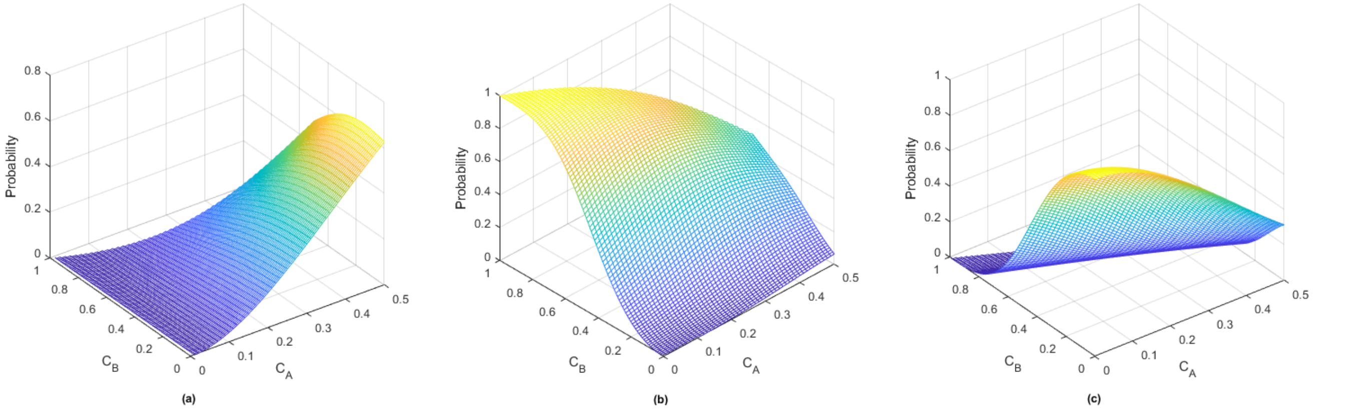


Fig. 5. The probability of (a) attack success, (b) attack failure, and (c) big attack failure. $\boldsymbol{A}$ is the computational power of the attacker (assuming that the power does not exceed 50% of the network's power) and $\boldsymbol{B}$ is the computational power of Bob and the other miners who are mining on his chain.

ing misleading techniques can push the success probability up by 16.42% (at $\mathcal{C}_A = 0.39$). Another interesting point in the figure is that the attacker would reach the maximum success probability if she could mislead all miners (i.e. $\mathcal{C}_B = 0$). But this would be true only if the attacker's computational power was 0.31. With a power that is more than 0.31, for maximum profitability, she does not need to mislead everyone and misleading only a fraction of miners suffices ($\mathcal{C}_B > 0$).

IV. Attack Analysis

Bribery attack has been analyzed by us in [30]. In order to calculate the required bribe for Alice to offer, we adopt the cited paper's analysis but customize it based on our variables. Note that Alice can be a malicious IoT device in an IIoT network or a module in critical infrastructure. By proposing a bribe, rational miners (i.e. M) select the more profitable chain. Similar to [26], we assume that (because of rationality) the miners form a mining pool/group whose computational power is denoted by $\mathcal{C}_M$. If M accepts the bribe, the associated miners receive the block reward (F) as well as the bribe (R_i) in state i, only if the attack ultimately succeeds. If they refuse to accept the bribe, they receive the block rewards (F) only if the attack fails. We denote M's gain as Ψ, when accepts the bribe, and as Ω, when refuses to accept the bribe. The success probability of Bribery Attack when M accepts the bribe is denoted by $P_{Z,S}$, and the failure probability when M refuses to accept the bribe and mines on the main chain Y is denoted by $P_{Y,F}$. In the latter case, and from M's point of view, the attack (which is to them, similar to a bribery attack) will certainly fail ($P_{Y,F} = 1$).

$$\Psi = P_{Z,S} \times (R_i + F) \tag{3}$$

$$\Omega = P_{Y,F} \times \frac{\mathcal{C}_M}{\mathcal{C}_M + \mathcal{C}_A + \mathcal{C}_B} \times F \tag{4}$$

In (4), the ratio shows M's share of block reward. When $\Psi > \Omega$, accepting the bribe will be profitable for M. By setting $P_{Y,F} = 1$ and $P_{Z,S} = (\mathcal{C}_M/(\mathcal{C}_A + \mathcal{C}_B))^{i+1}$, we can find minimum R_i as follows:

$$R_i > \frac{(1 - \mathcal{C}_M)^{i+1}}{\mathcal{C}_M^i} \times F - F \tag{5}$$

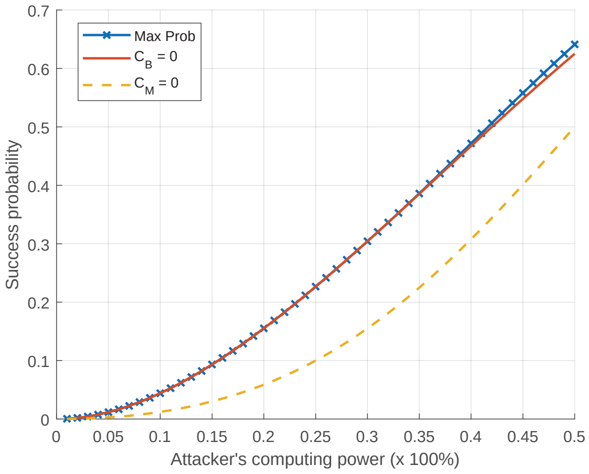


Fig. 6. Attack success probability with and without the misleading technique.

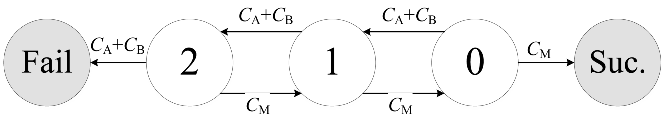


Fig. 7. The Markov model from miners' point of view.

A. Rational Miners

Network can have both rational and honest miners. In this attack, we assume that only some members of the network are rational, who form a pool by joining the chain Z. We assume that the attacker and miners know the cumulative computational power of rational miners ($\mathcal{C}_M$). Since we assumed that Alice keeps her chain private, from miners' point of view, she is an honest miner who mines on Y. In this view, the Markov model of the attack will be similar to Fig. 7, which is different from what the attacker considers. Each state in this model refers to the length difference of the two chains (Y and Z) which is represented by variable i in equations (3) and (5).

B. Bribe Amounts

According to (5), the attacker can find the amount of required bribe for each $i \in \{0,1,2\}$ using (6), (7), and (8).

$$R_0 = \epsilon > 0 \tag{6}$$

$$R_1 = \begin{cases} F \times (\mathcal{C}_M - 3 + \frac{1}{\mathcal{C}_M}) & \mathcal{C}_M \leqslant 0.3819 \\ \epsilon > 0 & \mathcal{C}_M > 0.3819 \end{cases} \tag{7}$$

$$R_2 = \begin{cases} \frac{(1-\mathcal{C}_M)^3}{\mathcal{C}_M^2} \times F - F & \mathcal{C}_M \leqslant 0.4301 \\ \epsilon > 0 & \mathcal{C}_M > 0.4301 \end{cases} \tag{8}$$

As the bribe should always be positive, for $i = 0$, the right hand side of (5) is always less than or equal 0, meaning that by setting any positive amount like ϵ for R_0, equation (5) will be satisfied. This is same for $i = 1$, and for $\mathcal{C}_M > 0.3819$, meaning that setting a positive amount like ϵ for R_1 is enough for the attacker.

C. Edge Categorization

We categorize the edges in the Markov model of Fig. 4 based on the amount of reward or bribe that the attacker gains or releases. We have the following categories:

$$\begin{aligned} AEdge &= \{(002,012),(012,022),(022,032),\\ &\quad (001,011),(011,021),(000,010)\}\\ R2Edge &= \{(012,001),(022,011),(002,0x1)\}\\ R1Edge &= \{(011,000),(001,0x0)\}\\ R0Edge &= \{(000,100)\}\\ ResetEdge &= \{(022,012),(012,002)\}\\ ZeroEdge &= \{(002,003),(001,002),(011,012),\\ &\quad (0x1,0x0),(0x0,0x1),(000,001),\\ &\quad (0x2,0x1),(0x0,1x0),(0x1,0x2),\\ &\quad (0x2,0x3)\} \end{aligned}$$

$AEdge$ refers to the events/edges in which the attacker has found a block in her chain. $RiEdge$ are the events in which a block is found in the misled chain and the bribed Ri is earned by M. The events in $ResetEdge$ suggest that the attacker has initiated a new chain for the misled miners with a new bribe, and that M leaves the old Z and starts mining on the new Z. All the bribes earned by M on the old Z will be lost. In $ZeroEdge$, no reward or bribe is gained or spent.

We also categorize the edges to the final states as below:

$$\begin{aligned} ASuccEdge &= \{(022,032),(011,021),(000,010)\}\\ ALoseEdge &= \{(002,003),(0x2,0x3)\}\\ ABigLoseEdge &= \{(000,100),(0x0,1x0)\} \end{aligned}$$

We use these to calculate the attacker's gain and cost.

V. Simulations

We take Bitcoin for our evaluations. Industrial POW blockchains (in IIoT or Industry 4.0) are similar. We assume the attacker starts the attack from the state "012" in Fig. 4. At the beginning of the attack, the graphical view of the Bitcoin's Blockchain will be similar to Fig. 3(a). If Alice launches an attack and succeeds, the reward obtained through state $3'$ will be her whole gain from the attack. However, if the attack fails, the reward of that state will not be deemed as an attack cost, because it was already lost before launching the attack.

We calculate the cost and revenue of Misleading Attack by approximation. To this end, all of the paths (Pa) from the starting state to the final states (Fig. 4) with path probabilities (K) greater than $10^{-8}\%$ (or 0.0000000001) are taken into account in the calculation of the Attacker's Average Gain ($AAGain$), the Misled Miners' Average Gain ($MAGain$), the Attacker's Average Loss ($AALoss$), and the Attacker's Average BigLoss ($AABLoss$) as in (15), (16), (17), and (18), respectively. The error rate for this assumption is negligible. For example for $10^{-9}\% \leqslant K \leqslant 10^{-7}\%$ the error lies in $-0.0161BTC < Er < 0.0376BTC$. The number of paths for the assumption ($10^{-8}\%$) varies for different values of $\mathcal{C}_A, \mathcal{C}_B$, and $\mathcal{C}_M$. For example for $\mathcal{C}_A = 0, \mathcal{C}_B = 1$, and $\mathcal{C}_M = 0$, only one path exists in the Markov chain, but for $\mathcal{C}_A = 0.35, \mathcal{C}_B = 0.39$, and $\mathcal{C}_M = 0.26$ more than 3.3 million paths are checked. Up to 4.3 million paths are included for each probability samples in the calculation of the results. In the following, $x \in Pa$ stands for a path (composed of links) and Pr is the transition probability matrix of Fig. 4.

$$K(x) = \prod_{\forall (u,v) \in x} Pr[u,v] \tag{9}$$

$$A(x) = \{(u,v) | (u,v) \in x, (u,v) \in AEdge\} \tag{10}$$

$$Ri(x)_{i \in \{0,1,2\}} = \{(u,v) | (u,v) \in x, (u,v) \in RiEdge\} \tag{11}$$

$$AS = \{x | x[Last] \in ASuccEdge\} \tag{12}$$

$$AL = \{x | x[Last] \in ALossEdge\} \tag{13}$$

$$ABL = \{x | x[Last] \in ABigLossEdge\} \tag{14}$$

$$AAGain = \sum_{\forall x \in AS} K(x) \times F \tag{15}$$

$$MAGain = \sum_{\forall x \in ABL} K(x) \sum_{i=0}^{2} (R_i \times |Ri(x_L)|) \tag{16}$$

$$AALoss = \sum_{\forall x \in AL} K(x) \times F \times |A(x)| \tag{17}$$

$$\begin{aligned} AABLoss = &\textstyle\sum_{\forall x \in ABL} (K(x) \times (\sum_{i \in \{0,1,2\}} (R_i \times |Ri(x_L)|) \\ &+ F\,|A(x_L)|)) \end{aligned} \tag{18}$$

x_L will be defined later. $AAGain$ and $MAGain$ refer to the average reward that Alice and misled miners gain, respectively. $AALoss$ is the average reward of the blocks that Alice found on her chain, but lost because of the success of the main chain (Y). $AABLoss$ is the average reward of the blocks that Alice found on her chain, as well as the bribes she published on the chain M, but lost due to misled miners' success. As explained before, the reward of Block $3'$ in Fig. 3(a) is not included in the calculation of $AALoss$ and $AABLoss$. All the rewards that Alice collects by mining blocks will not be included in the attacker's gain (from the attack) either. Calculating $AABLoss$ is complex compared to the others because of the existence of $ResetEdge$ in the path. Any event $\varepsilon \in ResetEdge$ on path x implies that misled miners left the old Z at the point ε was fired, and started mining on the new Z. Therefore, Alice

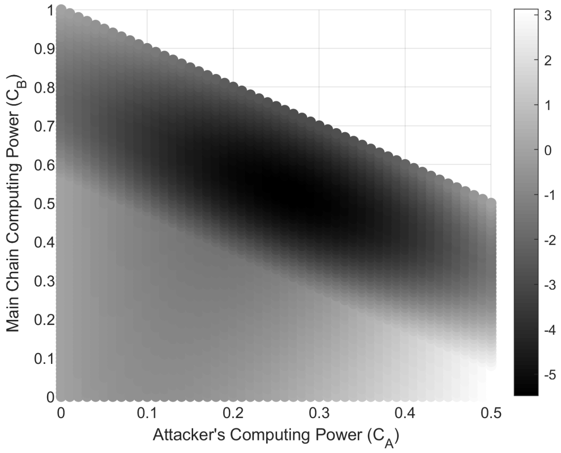


Fig. 8. Heatmap of $\boldsymbol{EG}$ at different computational powers for the players.

will not lose the money she bribed before this point. If misled miners succeed, the rewards they gain will be accounted for from the last point event ε was raised, to the end of path x. Accordingly, to calculate $AABLoss$, we divide path x based on the edges in $ResetEdge$ and select the last sub-path from it (x_L in (16) and (18) refers to the last sub-path).

The attacker's Expected Gain (EG) from the attack can be calculated using Bayes' expectation as in (19).

$$EG = \sum_{\upsilon \in \{Gain, Loss, BigLoss\}} (-1)^{\mathcal{K}(\upsilon)} \times \upsilon \times P(\upsilon) \tag{19}$$

$$\mathcal{K}(\upsilon) = \begin{cases} 0 & \upsilon \in \{Gain\} \\ 1 & oth. \end{cases} \tag{20}$$

Based on (15), (17), (18), and (19), we can obtain EG as,

$$EG = AAGain - AALoss - AABLoss \tag{21}$$

Fig. 8 shows a heat map for EG at different computational powers for the players. As seen, EG ranges from -5.48 BTC (at $\mathcal{C}_A = 27\%$ and $\mathcal{C}_B = 53\%$) to 3.12 BTC (at $\mathcal{C}_A = 50\%$ and $\mathcal{C}_B = 0\%$). For example, if the attacker's computational power is 37% of the network power, and the main chain's is 14%, the expected gain of the attack (in which the attacker tries to revive her orphaned block with 6.25 BTC reward inside) will be about 1 BTC. Starting from 24%, the attack turns profitable.

Fig. 9a shows the profitable working points for a potential attack by Alice through sweeping the values of $\mathcal{C}_A$ and $\mathcal{C}_B$. In this figure, the amount of profit has been obtained from (21). Except at $\mathcal{C}_A = 50\%$, there is no profitable point for $\mathcal{C}_B > 23\%$. This means the attacker should keep the $\mathcal{C}_B$ below this value.

VI. Other Attack Strategies for Orphan Blocks

To make Misleading Attack more profitable, one can create and use different strategies from the ones demonstrated in Fig. 4, or use a subset of the strategies. For example, removing the strategy $IncZ$ (the strategy which adds incentives on Z to make the chain profitable for M) will make chain Z unprofitable for misled miners in 000, 001, and 011 states and extremely reduces the success probability of Z, but increases the success probability of the main chain. Removing $IncZ$ prevents the attacker from suffering a $BigLoss$ and decreases the speed of block finding on the main chain for a small period of time (by removing the computational power of M from the main chain).

Furthermore, in our analysis, the attacker stops the attack when the difference between the length of Y and X exceeds 2 (i.e. 3 or more). One can continue the attack for one more step, and start giving M incentives from state 1 in Fig. 3(a).

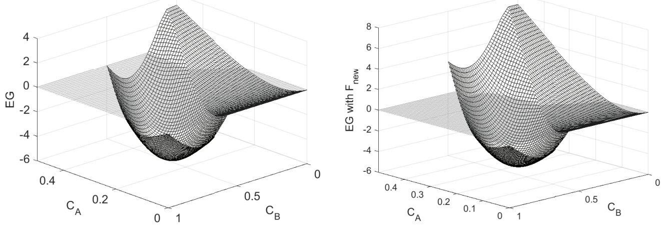
A. Sweet Point to Launch the Attack

Approximately, every four years, Bitcoin halving cuts the mining rewards in half. Halving time is the best point to launch a misleading attack. At this time, doubling $AAGain$ (in (22)) not only gives the attacker more gain, but also, as depicted in Fig. 9b, creates more profitable working points than the ones shown in Fig. 9a. After the halving, F_{new} (new block reward) will replace F (old block reward) in (17), (18), and (15), but at the halving point, the attacker tries to revive her lost block reward which is the reward of the last block before halving. In case it is revived, the attacker's gain will be $AAGain$ with F, but if the attack fails, the attacker loses the blocks which she has mined on her orphaned block after halving (all with F_{new} as their reward). Therefore, at the point, F_{new} replaces F in (17), and (18), but (15) turns into (22).

$$\begin{aligned} F_{new} &= \frac{1}{2} F \\ AAGain &= \sum_{\forall x \in AS} K(x) \times 2 \times F_{new} \end{aligned} \tag{22}$$

B. The Future of Misleading Attack for Orphan Blocks

Different attacks may be affected by halving of Bitcoin mining reward in aspects like profitability or practicability, which shows the importance of studying the future of attacks. In halvings, the value of F diminishes over time, which in turn, reduces the threat of Misleading Attack for orphan blocks. This is because the attacker tries to prevent the block from being orphaned to obtain the block reward. From a mathematical point of view, the $AAGain$ value greatly depends on F, and halving F hugely affects it.



(a) (b)

Fig. 9. Profitable working points to initiate a misleading attack. All the points above the $\boldsymbol{EG = 0}$ surface are profitable. (a) Without halving, there is no profitable point when $\boldsymbol{C_B > 0.23}$. (b) After halving and with $\boldsymbol{F_{new}}$, the number of profitable points increases dramatically.

VII. Using the Misleading Attack Technique in Other Attacks & Open Problems

In this paper, we introduced Misleading Attack and analyzed its effect on preventing a block from being orphaned. The Misleading Attack's technique can also be used in other known attacks to boost the attack profitability, reduce its costs or increase its success probability. In this section, we study the effect of adopting Misleading Attack in a Bribery Attack [30]. To this end, we redesign the attack process as follows.

As soon as a block is found and distributed in the network, the attacker issues a transaction to buy some goods from Bob (the victim) and starts mining on the newly generated block privately to undo the purchase transaction. We call this phase $Preprocess$. After confirming the transaction, the attacker may start the $Attack$ phase based on the number of the blocks she has found in the $Preprocess$ phase. Let us denote this number by z. If the network needs c blocks to confirm a transaction, the attacker will need two more blocks to succeed (one to revert the spent money on the block containing the transaction, and one to overcome the main chain). Accordingly, the $Attack$ phase will start if $z < c + 2$.

Using the Misleading Attack in the $Preprocess$ phase will give the attacker the opportunity to increase z. Misled miners will increase the time required to find $c+1$ blocks (one block containing the transaction and c blocks for confirmation) on the main chain, and the attacker will have more time to mine more blocks on her private chain.

We assume that the attacker's goal in adopting the Misleading Attack is to create only one (more) block on her private chain in the $Preprocess$ phase. In average, the comparative rate of block creation on chain X with respect to Y is $\mathcal{C}_A/\mathcal{C}_B$. Creating one more block on chain X needs the attacker to mislead M in a way that (23) is satisfied. The left side of the equation shows the rate of block creation after misleading M, and in the second term in right side the attacker wants to increase the rate by 1 during the $Preprocess$ phase ($c + 1$ blocks in the main chain).

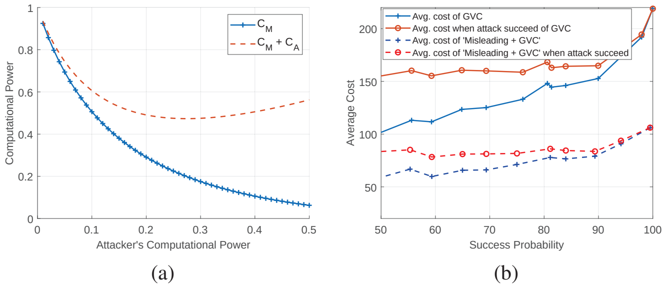
$$\frac{\mathcal{C}_A}{\mathcal{C}_B - \mathcal{C}_M} = \frac{\mathcal{C}_A}{\mathcal{C}_B} + \frac{1}{c+1} \tag{23}$$

After some simplifications, we have:

$$\mathcal{C}_M = \mathcal{C}_B - \frac{(c+1) \times \mathcal{C}_A \times \mathcal{C}_B}{(c+1) \times \mathcal{C}_A + \mathcal{C}_B} \tag{24}$$

Fig. 10a shows that for different values of $\mathcal{C}_A$, how much computational power the attacker should take away from the main chain (by misleading) to be able to create one more block (in average) on her chain. For example, when $\mathcal{C}_A = 30\%$, Alice should mislead almost 18% of the network. As depicted in the figure, when the attacker's computational power is in the range $0.18 < \mathcal{C}_A < 0.39$, the sum of $\mathcal{C}_M$ and $\mathcal{C}_A$ can be less than half of the network's computational power.

In this part, we want to study a case in which Misleading Attack helps the attacker to create one more block on her private chain. We show how this extra block can help the attacker to substantially reduce the cost of Bribery Attack. Fig. 10b shows the cost of GVC Bribery Attack of [30] with and without the help of Misleading Attack. GVC stands for "Guaranteed Variable-Rate Bribing with Commitment" and this strategy adopts Differential Evolution (DE) to optimize the bribe amounts. The results show that starting the attack from state 5 (rather than 6, as assumed in [30]), almost halves the attack cost without reducing the success probability. In the figure, we used the Bitcoin pools mining power distribution of [30]. The value of Average Cost is obtained by DE for a given target success probability. DE returns the Average Cost as well as the best strategy in each state of the Bribery Attack. As the mining power distribution is discrete, in some cases, DE increases the amount of commitment to attract more miners to the attacker's chain in the states close to the final state. This decreases the required bribe for earlier states and, as a result, decreases the attack cost (59% and 81% in Fig. 10b).



(a) (b)

Fig. 10. (a) The minimum amount of $\mathcal{C}_M$ in order to create one more block in the attacker's chain in the $\boldsymbol{Preprocess}$ phase. (b) Average cost of the GVC strategy when adopting the Misleading Attack technique to create one more block on the attacker's chain.

VIII. Conclusion

In this paper, a new technique is presented which can be used to launch attacks on the blockchains employed in critical infrastructure or Industry 4.0. It can be used alone or be combined with other known attacks to increase the attack profitably. This technique introduces a new category of attacks which we call Misleading Attacks. In these attacks, one prevents a fraction of network from mining on the main chain. Compared to Eclipse Attack, this new type of attack adopts a totally different approach which makes it more practical to implement. We introduced and analyzed Misleading Attack on orphan blocks with different strategies. An attacker can design other strategies to make the attack even more profitable. Furthermore, we studied the effect of this technique on the cost of Bribery Attack. An attacker can launch Misleading Attack for different purposes. If she tends to revive her orphan block, Misleading Attack can be used in most blockchain-based cryptocurrencies that have block reward. This attack may not be usable in Ethereum in which uncle-block concept is defined and accepted. However, one can still use the technique for purposes such as Double Spending. All the previous works which tried to combine Eclipse Attack with other types of attacks can use Misleading Attack for the same purposes. As a future work, we tend to apply this technique on Selfish Mining.

References

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," Manubot, Tech. Rep., 2019.

[2] D. Chaum, “Blind signatures for untraceable payments,” in *Advances in cryptology*, 1983, pp. 199–203.

[3] J. Camenisch, S. Hohenberger, and A. Lysyanskaya, “Compact e-cash,” in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2005, pp. 302–321.

[4] S. Haber and W. S. Stornetta, “How to time-stamp a digital document,” in *Conference on the Theory and Application of Cryptography*, 1990, pp. 437–455.

[5] T. Ahmed Teli, F. Masoodi *et al.*, “Security concerns and privacy preservation in blockchain based iot systems: Opportunities and challenges,” in *Int. Conf. on IoT Based Control Networks & Intelligent Systems*, 2021.

[6] M. Bastiaan, “Preventing the 51%-attack: a stochastic analysis of two phase proof of work in bitcoin,” in *Availab le at http://referaat. cs. utwente. nl/conference/22/paper/7473/preventingthe-51-attack-a-stochasticanalysis-oftwo-phase-proof-of-work-in-bitcoin. pdf*, 2015.

[7] M. Rosenfeld, “Analysis of bitcoin pooled mining reward systems,” *arXiv preprint arXiv:1112.4980*, 2011.

[8] I. Eyal and E. G. Sirer, “Majority is not enough: Bitcoin mining is vulnerable,” in *International conference on financial cryptography and data security*. Springer, 2014, pp. 436–454.

[9] S.-Y. Chang, Y. Park, S. Wuthier, and C.-W. Chen, “Uncle-block attack: Blockchain mining threat beyond block withholding for rational and uncooperative miners,” in *International Conference on Applied Cryptography and Network Security*, 2019, pp. 241–258.

[10] E. Heilman, A. Kendler, A. Zohar, and S. Goldberg, “Eclipse attacks on bitcoins peer-to-peer network,” in *USENIX Security Symposium*, 2015, pp. 129–144.

[11] G. O. Karame, E. Androulaki, and S. Capkun, “Double-spending fast payments in bitcoin,” in *ACM Conference on Computer and Communications Security*, 2012, pp. 906–917.

[12] G. Ramezan, C. Leung, and Z. J. Wang, “A strong adaptive, strategic double-spending attack on blockchains,” in *IEEE International Conference on Internet of Things (iThings) and GreenCom and CPSCom and SmartData*, 2018, pp. 1219–1227.

[13] G. Ramezan and C. Leung, “Analysis of proof-of-work-based blockchains under an adaptive double-spend attack,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 11, pp. 7035–7045, 2020.

[14] S. Zhang and J.-H. Lee, “Double-spending with a sybil attack in the bitcoin decentralized network,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 10, pp. 5715–5722, 2019.

[15] J. R. Douceur, “The sybil attack,” in *International workshop on peer-to-peer systems*. Springer, 2002, pp. 251–260.

[16] K. Nayak, S. Kumar, A. Miller, and E. Shi, “Stubborn mining: Generalizing selfish mining and combining with an eclipse attack,” in *IEEE European Symposium on Security and Privacy*, 2016, pp. 305–320.

[17] J. Göbel, H. P. Keeler, A. E. Krzesinski, and P. G. Taylor, “Bitcoin blockchain dynamics: The selfish-mine strategy in the presence of propagation delay,” *Performance Evaluation*, vol. 104, pp. 23–41, 2016.

[18] I. Eyal, “The miner’s dilemma,” in *IEEE Symposium on Security and Privacy*, 2015, pp. 89–103.

[19] Y. Kwon, D. Kim, Y. Son, E. Vasserman, and Y. Kim, “Be selfish and avoid dilemmas: Fork after withholding (faw) attacks on bitcoin,” in *ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 195–209.

[20] M. Babaioff, S. Dobzinski, S. Oren, and A. Zohar, “On bitcoin and red balloons,” in *ACM Conference on Electronic Commerce*, 2012.

[21] L. Xiangyang, Z. Pei, Z. Mingliang, L. Hao, and Q. Cheng, “A novel covert communication method based on bitcoin transaction,” *IEEE Transactions on Industrial Informatics*, 2021.

[22] J. Bonneau, “Why buy when you can rent?” in *International Conference on Financial Cryptography and Data Security*, 2016, pp. 19–26.

[23] J. Bonneau, “Hostile blockchain takeovers,” in *International Conference on Financial Cryptography and Data Security*, 2018, pp. 92–100.

[24] P. McCorry, A. Hicks, and S. Meiklejohn, “Smart contracts for bribing miners,” in *International Conference on Financial Cryptography and Data Security*, 2018, pp. 3–18.

[25] A. Judmayer, N. Stifter, A. Zamyatin, I. Tsabary, I. Eyal, P. Gazi, S. Meiklejohn, and E. Weippl, “Pay-to-win: Incentive attacks on proof-of-work cryptocurrencies,” Cryptology ePrint Archive No. 2019/775, Tech. Rep., 2019.

[26] H. Sun, N. Ruan, and C. Su, “How to model the bribery attack: A practical quantification method in blockchain,” in *European Symposium on Research in Computer Security*, 2020, pp. 569–589.

[27] J. Teutsch, S. Jain, and P. Saxena, “When cryptocurrencies mine their own business,” in *International Conference on Financial Cryptography and Data Security*, 2016, pp. 499–514.

[28] Y. Velner, J. Teutsch, and L. Luu, “Smart contracts make bitcoin mining pools vulnerable,” in *International Conference on Financial Cryptography and Data Security*, 2017, pp. 298–316.

[29] K. Liao and J. Katz, “Incentivizing blockchain forks via whale transactions,” in *International Conference on Financial Cryptography and Data Security*. Springer, 2017, pp. 264–279.

[30] G. Ebrahimpour and M. S. Haghighi, “Analysis of bitcoin vulnerability to bribery attacks launched through large transactions,” *arXiv preprint arXiv:2105.07501*, 2021.

Ghader Ebrahimpour received the M.Sc. degree in Information Security from Amirkabir University of Technology (AUT), Iran in 2015. Ghader has had several positions in industry before. He is currently a researcher in Advanced Networking and Security research Laboratory (ANSLab). His current research focuses on the analysis and design of cryptocurrency attacks on Blockchain-based systems.

Mohammad Sayad Haghighi (IEEE SM’18) is the Head of IT Department at the University of Tehran, Iran. He is also the director of Advanced Networking and Security research Laboratory (ANSLab). His research interests are wireless networks and cybersecurity. Dr. Sayad Haghighi has served as a PC member of many conferences such as IEEE WNS, IEEE SICK, IEEE HPCC, IEEE DASC, and IEEE LCN. He has won several national grants including two from Iran National Science Foundation (INSF).

Mamoun Alazab received the Ph.D. degree in computer science from Federation University, Australia. He is currently an Associate Professor with the College of Engineering, IT and Environment, Charles Darwin University, Australia. His research is multidisciplinary and focuses on cyber security and digital forensics of computer systems with a focus on cybercrime detection and prevention. He is the Founder and the Chair of the IEEE Northern Territory (NT) Subsection Detection and Prevention.