Effects of miners’ location on blocks selection in blockchain

Kosuke Toda\textsuperscript{a)}, Naomi Kuze\textsuperscript{b)}, Toshimitsu Ushio\textsuperscript{c)}

Graduate School of Engineering Science, Osaka University
1-3 Machikaneyama-cho, Toyonaka, Osaka 560-8531, Japan
a) toda@hopf.sys.es.osaka-u.ac.jp, b) kuze@sys.es.osaka-u.ac.jp, c) ushio@sys.es.osaka-u.ac.jp

Abstract: Blockchain is a distributed ledger technology for recording transactions. To guarantee the immutability of the blocks, miners need a lot of computational resources for creating blocks (mining). Since mining is conducted distributedly by many miners, chain forks can occur in the blockchain. In this paper, we investigate the effects of miners’ locations, the size of the MP, and the network structure on the selection rate of blocks generated by the MP when chain forks occur through simulations.

Keywords: Blockchain, fork, mining pool, regular graph, scale-free network

Classification: Network

References

[1] S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system,” http://bitcoin.org/bitcoin.pdf, 2008.
[2] Y. Xiao, et al., “Modeling the impact of network connectivity on consensus security of Proof-of-Work blockchain,” arXiv preprint arXiv: 2002.08912, 2020.
[3] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” Nature, vol. 393, no. 6684, pp. 440, 1998.
[4] A. L. Barabási and R. Albert, “Emergence of scaling in random networks,” Science, vol. 286, no. 5439, pp. 509–512, 1999.

1 Introduction

Blockchain is a distributed ledger technology for recording transactions and underlies the digital currency such as Bitcoin [1]. In a blockchain-based service, users called miners distributedly generate blocks consisting of several transactions. These blocks are stored and managed as a chain of blocks (a blockchain). Miners are connected as a network, and the generated blocks are propagated hop by hop. To guarantee the immutability of the blocks, Proof-of-Work (PoW) is used as a consensus algorithm. In this algorithm, miners need a lot of computational resources for creating blocks (mining), which results in the resistance to block tampers. A lot of miners try to generate blocks, and a miner that success to generate a block get a reward. The more resources the miner has, the easier the block generation is to succeed.
results in an incentive for miners to have many resources. In order to get rewards efficiently, several miners form a mining pool (hereinafter referred to as an MP) for cooperating in mining.

Since mining is conducted distributedly by many miners, chain forks can occur in the blockchain. When a chain fork occurs, each miner needs to select a highest block as a parent block for a new block, which complicates the consensus of the blockchain. However, properties of chain forks have not been clarified. In this paper, we investigate the effects of miners’ locations, the size of the MP, and the network structure on the selection rate of blocks generated by the MP when chain forks occur through simulations.

2 Blockchain forks

2.1 Problem statement
We focus on chain forks. Since new blocks are propagated hop by hop on the network, we clarify how the occurrence of chain forks and the blocks selection are related to the network structure and the location of miners belonging to the MP. Xiao, et al. [2] studied the effects of network connectivity on consensus security such as chain fork rate and mining reward. We investigate the effects of the location of miners on the occurrence of chain forks and the blocks selection.

2.2 A blockchain model
We model the network of miners (miners’ network) by the undirected graph $G = (V, E)$. $V$ is a finite set of nodes and each node corresponds to a miner. $E$ is a set of edges, indicating that there is a direct data communication between miners. Let $N$ be the number of miners, i.e. $|V| = N$. Moreover, we assume that there is only one MP and $\mathcal{P} \subseteq V$ be the set of miners belonging to the MP. Miner $m \notin \mathcal{P}$ performs mining independently. Miners belonging to the MP cooperate with each other by sharing the information of blocks generation of other miners belonging to the MP.

We define the information $\beta_k$ about the $k$th block $B_k$ generated in the miners’ network as $\beta_k = (n_k, p_k, d_k, w_k)$, where $n_k$ is the identifier of $B_k$, $p_k$ is the identifier of the parent block of $B_k$, $d_k$ is the block height of the block $B_k$, and $w_k \in V$ is the miner that generated block $B_k$. Block $B_0$ represents a genesis block. A chain consisting of a group of blocks held by miner $m$ is called a local chain in miner $m$. Let $t^m_k$ be the time when miner $m$ received block $B_k$. In this paper, $t^m_0 = 0$ for all $m$. The block $B_k$ held by the miner $m$ is represented by the pair of $\beta_k$ and $t^m_k$, that is, $(\beta_k, t^m_k)$. If the miner $m$ holds $\ell$ blocks in addition to the genesis block at time $t$, the set of blocks held by miner $m$ is represented as $C^\ell,m = \{(\beta_0, 0), \ldots, (\beta_\ell, t^m_\ell)\}$.

When there are multiple blocks with the same height, a chain fork occurs. That is, for different blocks $B_k$ and $B_{k'}$ ($k \neq k'$) held by miner $m$, a chain fork occurs when $d_k = d_{k'}$. Let $H^\ell,m$ be the set of $(\beta, \tau) \in C^\ell,m$ such that $\beta = (n, p, d, w)$ satisfies $d = \max\{d_k \mid 0 \leq k \leq \ell\}$.

We consider a discrete-time model of the block generation. Miners share
their local chains per unit time through local interactions. We assume that each miner $m \in V$ performs the following behavior at time $t \in T = \{0, 1, \ldots\}$:

1. If miner $m$ receives a block that it does not hold from adjacent nodes, add that block to its local chain.

2. Miner $m$ selects one block contained in $H^{t,m}$ as a parent block and conducts mining for generating a new block. The block selection is described in detail below.

   If $m \notin P$, $m$ selects a block $(\beta_p, t^m_p) \in H^{t,m}$ that satisfies $t^m_p = \min \{t^m_k | (\beta_k, t^m_k) \in H^{t,m}\}$. If $m \in P$, $m$ selects a block $(\beta_p, t^m_p) \in H^{t,m}$ that satisfies $w_p \in P$ if it exists. If there is no block $(\beta_p, t^m_p)$ that satisfies $w_p \in P$, $m$ selects a block in the same way as $m \notin P$.

In this paper, we assume that it takes 1 unit time for a miner $m$ to transmit a block to its adjacent miners. Let $f$ be the probability that a miner generates a block within a unit time. Assuming that all miners have the same computational resources, the probability that each miner generates a block within a unit time is given by $f/N$.

We consider the following two locations of miners belonging to the MP. One is the location in which a subgraph whose nodes belong to $P$ is connected (Location 1). The other is the location in which no edge in $V$ belongs to $P \times P$ (Location 2). Here, it is assumed that both Location 1 and 2 include a node with the maximum degree (hereinafter referred to as the hub node).

3 Simulation and Evaluation

In this section, we conduct simulation evaluations for investigating the effects of miners’ location in the MP on the blocks selection.

3.1 Simulation settings

3.1.1 Settings

We consider a network modeled by both a regular graph where every vertex has the same degree and a graph whose degree distribution follows a power-law (scale-free network). The former graph is generated by the Watts-Strogatz model (WS model) $WS(N, k, p)$ with $p = 0$ [3]. The latter graph is generated by the Barabási-Albert model (BA model) $BA(N, m_0)$ [4]. Note that $N$ is the number of nodes, $k$ is the mean degree, $p$ is randomness, and $m_0$ is the initial number of nodes. We consider the regular graph $N = 100$ and $k = 6, 8, 10$ and the BA model $N = 500$ and $m_0 = 1, 2, 3, 4, 5$. $|P|/N = 0.1$, $f = 0.02$, and $T = 10,000$ in this simulation.

3.1.2 Metrics

Among the blocks included in the longest chain of a certain miner $m$ obtained at the end of each simulation, the percentage of the blocks generated by the miners belonging to the MP is used as the index ($I$).
Let $X$ and $Y$ be the samples of $I$ obtained in Location 1 and 2, respectively. Assuming that $X$ and $Y$ follow the Gaussian distributions $N(\mu_1, \sigma^2)$, $N(\mu_2, \sigma^2)$ with the same variance, conduct the following $t$-test.

Here, when the variance $\sigma^2$ is unknown, we test the following two-sided hypothesis at the significance level $\alpha = 0.05$.

$$
\begin{align*}
H_0 : \mu_1 &= \mu_2 \quad \text{(Null hypothesis),} \\
H_1 : \mu_1 &\neq \mu_2 \quad \text{(Alternative hypothesis).}
\end{align*}
$$

The null hypothesis $H_0$ indicates that there is no significant difference in the block selection by the location, and the alternative hypothesis $H_1$ indicates that there is a significant difference in the blocks selection by the location. The simulation is performed $n$ times for each of the Location 1 and 2. Let the sample mean of $X$ and $Y$ be $\bar{X}$ and $\bar{Y}$, and the estimator of variance $\sigma^2$ be $\hat{\sigma}^2$. Using two-sided $\alpha$ points with $2n - 2$ degrees of freedom, the rejection range $W$ for the significance level $\alpha$ in the $t$-test

$$
T = \frac{|\bar{X} - \bar{Y}|}{\hat{\sigma} \sqrt{\frac{2}{n}}}
$$

is $W = \{T \geq 1.96\}$ for sufficiently large $n$. We set $n = 100$. In other words, the $t$-value obtained by Eq. (2) satisfies $T \geq 1.96$, it can be said that there is a significant difference in the blocks selection by the location.

### 3.2 Results and discussions

The results of the $t$-test are shown in Table I. With the regular graphs, there is no significant difference by the location regardless of the degree. With the BA models, there is a significant difference by the location when $m_0 = 1$, but the others do not.

First, we focus on the reason why there is no significant difference in the regular graph. In the regular graphs, all nodes have the same degree. The results show that the location of miners belonging to the MP has no effect

| Table I. The results of the simulation and $t$-test. |
|-----------------|----------------|----------------|----------------|
| $k$ | $\bar{X}$ | $\bar{Y}$ | $\hat{\sigma}^2$ | $T$ |
| 6 | 9.81 | 10.09 | 6.26 | $7.91 \times 10^{-1}$ |
| 8 | 10.12 | 9.98 | 5.01 | $4.56 \times 10^{-1}$ |
| 10 | 10.36 | 9.98 | 4.88 | 1.24 |

| $m_0$ | $\bar{X}$ | $\bar{Y}$ | $\hat{\sigma}^2$ | $T$ |
|---|---|---|---|---|
| 1 | 10.30 | 9.47 | 5.61 | 2.47 |
| 2 | 10.34 | 9.97 | 4.56 | 1.22 |
| 3 | 9.71 | 9.62 | 4.47 | $3.04 \times 10^{-1}$ |
| 4 | 10.11 | 9.96 | 4.97 | $4.67 \times 10^{-1}$ |
| 5 | 9.74 | 9.53 | 4.55 | $6.71 \times 10^{-1}$ |
on the selected block when chain forks occur if all nodes in the network have the same degree.

Next, we focus on the relationship between the average length in the BA model and the ratio of the occurrence of chain forks. Let $\langle l \rangle$ be the average length of the network and $\langle l_h \rangle$ be the average length from the hub node to other nodes. For $m_0 = 1, 2, 3, 4, 5$, we obtain a pair $(\langle l \rangle, \langle l_h \rangle) = (6.45, 4.49), (3.73, 2.21), (3.27, 2.17), (2.93, 1.89), (2.77, 1.95)$, respectively. We evaluate the ratio of the occurrence of chain forks over all generated blocks, denoted by $F$ [%]. Shown in Fig. 1 is the relationship between the average length $\langle l \rangle$ and the ratio $F$. The horizontal axis indicates the average length and the vertical axis indicates the ratio of the occurrence of chain forks. The plotted points are the average value of $F$ in 100 simulations, and the error bar indicates the $2\sigma$ confidence bound. The results show that $\langle l \rangle$ and $F$ are in a linear relationship.

Finally, we focus on the relationship among the average length in the BA model, the location of miners belonging to the MP, and the ratio at which blocks generated by miners belonging to the MP are selected when chain forks occur. We classify the chain forks into four types. The first one is
Table II. t-test about the fork.

| (N, m₀) | X  | Y  | σ²  | T  |
|---------|----|----|-----|----|
| (500, 1)| 0.570 | 0.406 | 0.090 | 3.87 |
| (500, 2)| 0.535 | 0.550 | 0.144 | 2.82 \times 10^{-1} |
| (500, 3)| 0.453 | 0.428 | 0.157 | 4.40 \times 10^{-1} |
| (500, 4)| 0.498 | 0.400 | 0.170 | 1.68  |
| (500, 5)| 0.403 | 0.401 | 0.175 | 5.07 \times 10^{-2} |

chain forks “PP” such that the highest block of each branch of the fork is generated by miners belonging to the MP. The second one is chain forks “II” such that the highest block of each branch of the fork is generated by miners who do not belong to the MP. The third one is chain forks “PIp” where there exists the highest block of a branch generated by a miner who does not belong to the MP while the selected highest block is generated by a miner who belongs to the MP. The fourth one is chain forks “PII” where there exists the highest block of a branch generated by a miner who belongs to the MP while the selected highest block is generated by a miner who does not belong to the MP. Let n(A) be the number of the occurrence A. We then define the following index I and perform the t-test as in Section 3.1.2.

\[ I = \frac{n(PP) + n(PIP)}{n(PP) + n(PIP) + n(PII)}. \]

The results of the t-test are shown in Table II. There is a significant difference by the location when \( m₀ = 1 \), but the others do not. When \( m₀ = 1 \), \( \langle l \rangle \) and \( \langle l_h \rangle \) are large, and the ratio of selected blocks generated by miners belonging to MP in Location 2 is significantly lower than that in Location 1. When \( m₀ = 2, 3, 4, 5 \), \( \langle l \rangle \) and \( \langle l_h \rangle \) are small, and the distance between the miners belonging to the MP in Location 2 is also small. Furthermore, Location 1 and 2 make little difference in the ratio of selection of the blocks generated by the miners belonging to the MP when the chain forks occur.

We summarize that in networks with large \( \langle l \rangle \) and \( \langle l_h \rangle \), the ratio of the occurrence of chain forks is large, the ratio of selected blocks depends on the location of miners belonging to MP, and a block generated by a miner with a small distance from the hub node belonging to MP is likely to be selected.

4 Conclusion

We have clarified the relationship among the average length of the network, the location of miners belonging to the MP, the ratio of the occurrence of chain forks, and the ratio of selected blocks when chain forks occur.

Our future work is to investigate the relationship between the miners’ locations and attacks such as selfish mining.

Acknowledgments

This research was supported by Grant-in-Aid for Early-Career Scientists No. 18034267 from the Japan Society for the Promotion of Science (JSPS).