Fork Rate-based Analysis of the Longest Chain Growth Time Interval of a PoW Blockchain

Hirotsugu Seike [1], Yasukazu Aoki [2], Noboru Koshizuka [1, 3]
[1] Graduate School of Interdisciplinary Information Studies, The University of Tokyo, Tokyo, Japan
[2] Picolab Co., LTD, Tokyo, Japan
[3] YRP Ubiquitous Networking Laboratory, Tokyo, Japan
Email: hirotsugu.seike@koshizuka-lab.org, aoki@picolab.jp, noboru@koshizuka-lab.org

***Abstract*—Nakamoto's consensus protocol, which is well known for its resistance to sybil attacks by using PoW (Proof of Work), enables us to build public blockchains, such as Bitcoin. In this protocol, miners seek to extend the longest chain by solving blockhash-based cryptographic puzzles and the required time is probabilistically determined. Therefore, the distribution of the time interval affects security, performance and applications which utilize the block height information. Some researchers assumed that the time follows an exponential distribution but this assumption requires that the blockchain network is fully synchronized. To overcome this unreal scenario, the bounded delay model, in which there is an upper bound for block propagation delay on the network, was proposed. However, it is difficult to calculate the upper bound without observing delay and bandwidth on real-world network links.**

To solve this problem, we proposed another method to analyze the distribution of the longest chain growth time interval by using the observed fork rate. We derived a closed-form lower bound for the CDF (Cumulative Distribution Function) of the time to update the global block height. We also obtained the Pearson distance which can be used as the metric to judge whether the network is approximately synchronous or not. Finally, we conducted network simulations for comparing our lower bound with the lower bound that is based on the bounded delay model. In numerical examples, we show how the block size affects these lower bounds.

***Index Terms*—Nakamoto's consensus, PoW (Proof of Work), the longest chain growth time interval, block propagation delay**

I. Introduction

Public blockchain technology was first introduced by S. Nakamoto for Bitcoin [1], which is one of the decentralized digital currencies, and currently is applied to various applications beyond exchanging money. Public blockchains, such as Bitcoin [1], Namecoin [2] and Ethereum [3], have been maintained by Nakamoto's consensus protocol, which is well known for its resistance to sybil attacks by using PoW (Proof of Work). In this protocol, all miners seek to extend the longest chain[1] by solving blockhash-based cryptographic puzzles. Therefore, the time interval to extend the longest chain is also determined probabilistically and the distribution affects security, performance and applications that use the block height information. In Bitcoin, the faster the honest chain is generated, the more difficult it is for an attacker's chain to catch up with it [1]. In blockchains, TPS (the number of Transactions Per Second) is proportional to the longest chain growth rate [4, 5]. In blockchain-based naming systems, such as Namecoin [2] and Blockstack [6], domain names are registered by a two-phase commit process, in which it is required to send a name registration transaction after the pre-order transaction is confirmed by the pre-determined number of sequential blocks (The details are described in [6, 7, 8]).

To obtain the distribution of the longest chain growth time interval, the authors assumed that the time interval follows an exponential distribution[2]. Using this assumption, Grunspan et al. [9] corrected the double spend race analysis given by S. Nakamoto [1]. Kawase et al. [10] analyzed the mean transaction-confirmation time for Bitcoin. However, this assumption requires that there is no block propagation delay on the network. In other words, the blockchain network is fully synchronized. To overcome this unreal scenario, the bounded delay model [4, 11, 12], in which there is an upper bound for block propagation delay, was proposed, but it is difficult to calculate the upper bound without observing receiving times since the block creation by deploying many nodes on the network or considering real propagation delay on the links. The former attempts were done in [13, 14].

In this paper, we propose another method to analyze the distribution of the longest chain growth time interval by only using the observed fork rate and the network block generation rate. The low fork rate guarantees that there is a lower bound for the CDF (Cumulative Distribution Function) of the network block creation rate for extending the longest chain. From this point of view, we derive a closed-form lower bound for the CDF of the longest chain growth time interval. This lower bound can be used for calculating an upper bound for the mean time to update the global block height and numerically computing a lower bound for the CDF of the time it takes for the longest chain to grow by n (> 1) blocks. We also give the Pearson distance, which can be used as the metric to judge whether the network is approximately synchronous or not. Finally, we conducted network simulations for comparing our lower bound with the lower bound that is based on the

[1]Ethereum does not adopt the GHOST [4] rule. It can be said that Ethereum currently adopts Longest Chain rule in Bitcoin since Ethereum only uses uncle block rewards as incentives for miners who try to extend the longest chain.

[2]This derivation is reviewed in [9].

bounded delay model. In numerical examples, we show how the block size affects these lower bounds.

II. Related Work

To analyze the longest chain growth time interval of a PoW blockchain, there are two types of models: Synchronous Model and Asynchronous Model. In this section, we explain conventional approaches for evaluating how the longest chain grows with respect to time.

A. Synchronous Model

In the synchronous model, no blockchain forks occur on the honest network since there is no block propagation delay. Therefore, the distribution of the time to mine a new block strictly matches the distribution of the longest chain growth time because there are no orphaned blocks. In [9], the authors assumed that the distribution of the block interval follows an exponential distribution, and using this assumption, they gave a closed-form formula for the probability that a double spend attack succeeds. The formula was also derived in the discrete model for the longest chain growth [15]. The CDF of the time t (≥ 0) for the longest chain to grow by n blocks $G_{\mathrm{nd}}(n,t)$ can be calculated by the following equation in [9].

$$G_{\mathrm{nd}}(n,t) = 1 - \exp(-\lambda t)\sum_{k=0}^{n-1}\frac{(\lambda t)^k}{k!}. \tag{1}$$

The PDF (Probability Density Function) $g_{\mathrm{nd}}(n,t)$ can be derived from the CDF with the following equation in [9].

$$\begin{aligned} g_{\mathrm{nd}}(n,t) &= \frac{d[G_{\mathrm{nd}}(n,t)]}{dt}, \\ &= \frac{\lambda^n}{(n-1)!}\cdot t^{n-1}\exp(-\lambda t). \end{aligned} \tag{2}$$

λ is the block generation rate or the mining speed[3], and $1/\lambda$ is the average time to mine a new block. When $n = 1$, Equation (2) is re-written as follows.

$$g_{\mathrm{nd}}(t) = \lambda\exp(-\lambda t). \tag{3}$$

The right side of Equation (3) represents an exponential distribution with parameter λ, and to simplify our notations, we define $G_{\mathrm{nd}}(1,t) \equiv G_{\mathrm{nd}}(t)$ and $g_{\mathrm{nd}}(1,t) \equiv g_{\mathrm{nd}}(t)$. Equation (3) was used for deriving the mean transaction-confirmation time for Bitcoin in [10].

To evaluate security, performance and some applications on a PoW blockchain with respect to the time it takes for the longest chain to grow in the synchronous network, we can simply use Equations (1, 2, 3). However, there is block propagation delay in the actual network. Hence, to adopt Equations (1, 2, 3), we need a method to judge whether the blockchain network can be approximately synchronous.

[3]The block creation rate λ is given by the total computing power of a blockchain network and the algorithmically determined mining difficulty of each block. In this paper, we deal with this rate as a constant value (We ignore difficulty adjustments).

B. Asynchronous Model

To overcome the unreal scenario that the blockchain network is fully synchronized, the bounded delay model was proposed [4, 11, 12]. In this model, there is an upper bound D (≥ 0) for block propagation delay in the asynchronous network. Even in the worst case, all miners receive a block D seconds after the block is mined. Thus, in the bounded delay model, we can calculate the lower bound $G_{\mathrm{cd}}(t)$ for the CDF of the time it takes for the main chain to grow by n blcoks from the following equation, considering the network that block propagation delay is always a constant value D.

$$G_{\mathrm{cd}}(t) = \begin{cases} G_{\mathrm{nd}}(n, t-nD) & (t \geq nD), \\ 0 & (0 \leq t < nD). \end{cases} \tag{4}$$

$G_{\mathrm{cd}}(t)$ is regarded as the CDF of the chain growth time when the mining process of each block is always delayed by a constant time D. The PDF $g_{\mathrm{cd}}(n,t)$ is given by the following equation.

$$g_{\mathrm{cd}}(n,t) = \begin{cases} g_{\mathrm{nd}}(n, t-nD) & (t \geq nD), \\ 0 & (0 \leq t < nD). \end{cases} \tag{5}$$

In this bounded delay model, it is required to calculate the delay diameter of the network D. In [4], the author insisted that the upper bound can be calculated by aggregate propagation delay and an aggregate measure of bandwidth because there is a clear linear relation between the block size and its block propagation delay from the observed results [13] in the case of Bitcoin. However, in this method, it is necessary to deploy many nodes to observe receiving times since the block generation. In [12], the upper bound D was estimated by an assumption that the smallest bandwidth of the network links and the diameter of the network are given in advance. It is required to consider real propagation delay and bandwidth on each link.

III. Fork Model for PoW Blockchains

In this section, we briefly explain the fork model, which is explained in [13]. In addition, we newly introduce another formula to calculate the probability of a blockchain fork. In Section IV, the formula is used for deriving a lower bound for the CDF of the network block generation rate for extending the main chain. Using the lower bound, we obtain a lower bound for the CDF of the longest chain growth time interval.

A. Blockchain fork definition

In a blockchain, blocks are organized in a directed tree. Each block b, except the root block g, has a reference to its parent block[4]. If the root block, which is also known as the genesis block, is the block's ancestor, the distance between the block b and the root block g is defined as its block height h_b (This means that $h_g = 0$). We adopt the notation β_h to reference the set of blocks at height h. A blockchain fork occurs at block height h when $|\beta_h| > 1$.

[4]Each block uniquely determines its parent by its hash pointer.

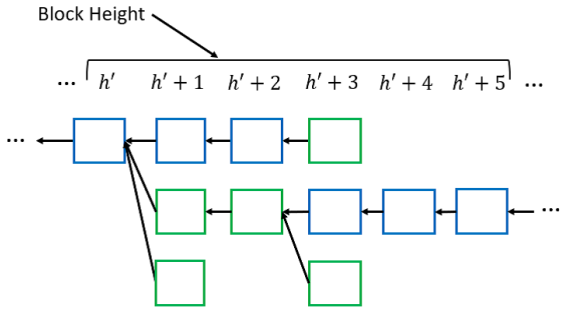


Fig. 1: This figure shows an example of all found blocks between the block height h' to $h'+5$. Each blue block is the earliest generated block at one height, and green blocks are the other blocks.

B. The probability of a blockchain fork

Figure 1 shows an example of all validated blocks between the block height h' to $h'+5$. In this example, there are three forks since $|\beta_h| > 1$ when the block height h is equal to $h'+1$, $h'+2$ or $h'+3$. Thus, the observed fork rate is 0.5. By assuming that the network topology is symmetric and the network computational resource is fully distributed, the theoretical probability of a blockchain fork P_F is derived in the following equation [13, eq. (2)].

$$P_F = 1 - (1 - P_b(\lambda))^{\int_0^\infty (1-F(t))dt}. \tag{6}$$

In [13], $F(t)$ is the CDF of the rate at which network miners know a block and the time t is the elapsed time since the first mined block at one height has been mined ($F(t)$ can also be treated as a lower bound for the ratio of the block generation rate for extending the longest chain to the network block creation rate[5]). $P_b(\lambda)$ is the probability of a block being found by all miners within a second. In [13], since they analyze Bitcoin, the block generation rate λ is small enough to approximate $P_b(\lambda)$ as λ. To calculate the theoretical fork rate, even if λ is high, we derive another formula, as follows.

$$P_F = 1 - \exp\Big(-\lambda \int_0^\infty (1 - F(t))dt\Big). \tag{7}$$

This equation holds whether λ is low or high. We give a derivation of Equation (7) in Appendix A. In the next section, we obtain a lower bound for the block creation rate for extending the main chain at time t by deriving a lower bound for $F(t)$.

IV. Proposed Analysis of the Longest Chain Growth Time based on the Blockchain Fork Rate

In this section, we derive a closed-form lower bound for the CDF of the longest chain growth time interval, give the Pearson distance which can be used for judging whether the network can be approximately synchronous or not, and show network simulation results to reveal how the block size affects our lower bound and the lower bound which is based on the bounded delay model.

A. Proposed lower bound based on the observed fork rate

In this subsection, we explain how to obtain our lower bound for the CDF of the longest chain growth time in the following procedure. First, we derive a lower bound for $F(t)$, which is also a lower bound for the ratio of the block generation rate for extending the main chain to the network block creation rate, by using the observed fork rate. Next, instead of $F(t)$, we substitute its lower bound into the differential equation which is used for calculating a lower bound for the CDF of the longest chain growth time. The solution of this differential equation is the lower bound for the CDF of the time to update the global block height.

1) A lower bound for F(t) by using the observed fork rate: Using the observed fork rate P_F, we derive a lower bound for $F(t)$. At first, we calculate T_w from Equation (7), as follows.

$$\begin{aligned} T_w &\equiv \int_0^\infty (1 - F(t))dt, \\ &= \frac{-\log(1 - P_F)}{\lambda}. \end{aligned} \tag{8}$$

T_w is defined as the "weighted average" delay [12] or the mean computational time wastefully consumed for propagating each block [13]. Equation (8) implies that the smaller P_F makes T_w smaller because $0 \le 1 - P_F < 1$. This also means that $F(t)$ must be faster converged to 1 with respect to time, and gives us a lower bound for $F(t)$ at any time $t \ge 0$. Suppose that α and ω_α such that $F(\omega_\alpha) = \alpha$, the following inequality is obtained from Equation (8).

$$\begin{aligned} T_w &= \int_0^\infty (1 - F(t))dt, \\ &\ge \int_0^{\omega_\alpha} (1 - F(t))dt, \\ &\ge \int_0^{\omega_\alpha} (1 - \alpha)dt = (1 - \alpha)\omega_\alpha. \end{aligned} \tag{9}$$

Inequality (9) can be simplified into Inequality (10).

$$\omega_\alpha \le \frac{T_w}{1 - \alpha}. \tag{10}$$

Here, we define $F_{\rm lb}(t)$ such that $F_{\rm lb}\left(T_w/(1-\alpha)\right) = \alpha$ at any time $t \ge T_w$. From this definition, $F(t) \ge F_{\rm lb}(t)$ is satisfied at any time ($t \ge T_w$) since $T_w/(1-\alpha)$ is the upper bound for ω_α. We also define $F_{\rm lb}(t) = 0$ when $0 \le t < T_w$, and then we get a closed-form lower bound for $F(t)$, as follows.

$$F_{\rm lb}(t) = \begin{cases} 1 - \dfrac{T_w}{t} & \big(t \ge T_w\big), \\ 0 & \big(0 \le t < T_w\big). \end{cases} \tag{11}$$

Figure 2 illustrates the relationship between Inequality (10) and Equation (11). If there is a time t' such that $F_{\rm lb}(t') \le F(t')$, the theoretical probability of a blockchain fork must be higher than the observed fork rate.

[5] In [13], $1 - F(t)$ is the ratio of computing resource which is not used for extending the first block mined at one height, if forks don't occur at the height. That is to say, the block generation rate for updating the global block height at time t is at least $\lambda F(t)$. This is because, even in the worst case, the first mined block can be extended with the block generation rate $\lambda F(t)$.

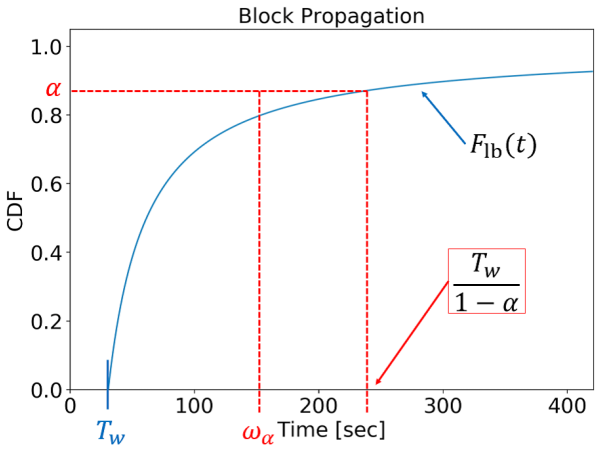


Fig. 2: An example of a lower bound for $F(t)$ ($\lambda = 1/600, P_F = 0.05$)

2) A lower bound for the CDF of the time for extending the longest chain by 1 block: We obtain a lower bound $F_{\mathrm{lb}}(t)$ for $F(t)$ from Equation (11). Considering the case that miners only extend the first found block at one height (In this scenario, miners stop mining if they don't know the target block), the block creation rate for updating the global block height strictly equals $\lambda F(t)$. In addition, we assume that the block generation rate for extending the main chain is $\lambda F_{\mathrm{lb}}(t)$ instead of $\lambda F(t)$. These assumptions only make the chain growth rate worse. Hence, we consider the CDF of the longest chain growth time interval in this scenario, and define the CDF as our closed-form lower bound $G_{\mathrm{lb}}(t)$. With this notation, Equation (12) holds since the right side of Equation (12) implies the probability that miners succeed to extend the first mined block at one height during the short period $[t, t+\Delta t)$[6]. Equation (13) can be easily derived from Equation (12).

$$G_{\mathrm{lb}}(t+\Delta t) - G_{\mathrm{lb}}(t) = \Big(1 - G_{\mathrm{lb}}(t)\Big)\lambda F_{\mathrm{lb}}(t)\Delta t. \tag{12}$$

$$\therefore \quad \frac{\partial G_{\mathrm{lb}}(t)}{\partial t} = \Big(1 - G_{\mathrm{lb}}(t)\Big)\lambda F_{\mathrm{lb}}(t). \tag{13}$$

The solution of Differential Equation (13) is given in Equation (14).

$$G_{\mathrm{lb}}(t) = \begin{cases} 1 - C\exp\Big(-\lambda\big(t - T_w \log t\big)\Big) \\ \qquad\qquad \big(t \geq T_w\big), \\ 0 \qquad\qquad \big(0 \leq t \leq T_w\big). \end{cases} \tag{14}$$
$$\Big(C = \exp\big(\lambda \cdot T_w(1 - \log(T_w))\big)\Big)$$

The constant value C is obtained by $G_{\mathrm{lb}}\big(T_w\big) = 0$. Lower bounds calculated from Equation (14) are shown in Figure 3. When the given fork rate is small, the lower bound approaches the CDF which is based on the synchronous model.

[6]In the short period $[t, t+\Delta t)$, there is a chance $\lambda F_{\mathrm{lb}}(t)\Delta t$ that the first found block will be extended if the chain growth event doesn't occur until time t.

We also define $g_{\mathrm{lb}}(t)$ as the function obtained by differentiating $G_{\mathrm{lb}}(t)$ with respect to time. $g_{\mathrm{lb}}(t)$ can be calculated in the following equation.

$$g_{\mathrm{lb}}(t) = \begin{cases} C\exp\Big(-\lambda\big(t - T_w \cdot \log(t)\big)\Big) \\ \quad \times \lambda\Big(1 - \dfrac{T_w}{t}\Big) \\ \qquad\qquad \big(t \geq T_w\big), \\ 0 \qquad\qquad \big(0 \leq t \leq T_w\big). \end{cases} \tag{15}$$
$$\Big(C = \exp\big(\lambda \cdot T_w(1 - \log(T_w))\big)\Big)$$

We plot $g_{\mathrm{lb}}(t)$ in Figure 4 for the case of $\lambda = 1/600, 1/15$. In Section IV-A3, using $g_{\mathrm{lb}}(t)$, we compute the Pearson distance between $g_{\mathrm{lb}}(t)$ and the exponential distribution based on the synchronous model. In Section IV-A4, we obtain the upper bound for the average chain growth time by using $g_{\mathrm{lb}}(t)$.

3) Estimating the degree of approximation between $g_{\mathrm{lb}}(t)$ and $g_{\mathrm{nd}}(t)$ which is based on the synchronous model : To evaluate the degree of approximation $G_{\mathrm{nd}}(t)$, which is based on the synchronous model, and $G_{\mathrm{lb}}(t)$, which is a lower bound for $G_{\mathrm{nd}}(t)$, we use the Pearson distance [16], is a measure used for the goodness of fit test which determines whether the observed frequency distribution is the same as the theoretical distribution. We don't adopt the Kullback-Leibler distance [17], which is one of the most frequently used distance measure, because we can't calculate $\log\left(g_{\mathrm{lb}}(t)/g_{\mathrm{nd}}(t)\right)$ with $t < T_w$.

The Pearson distance between $g_{\mathrm{nd}}(t)$ and $g_{\mathrm{lb}}(t)$ can be derived from Equation (16) by using the upper incomplete gamma function $\Gamma(a,x) = \int_x^\infty t^{a-1}\exp(-t)dt$ (A derivation of Equation (16) appears in Appendix B). In Table I, the distances are shown. This distance can be used as an indicator to judge whether the blockchain network is approximately synchronous or not by using two-sample test or by change-point detection [18]. In addition, this distance has a good characteristic that it only depends on the observed fork rate ($\lambda T_w = -\log(1 - P_F)$).

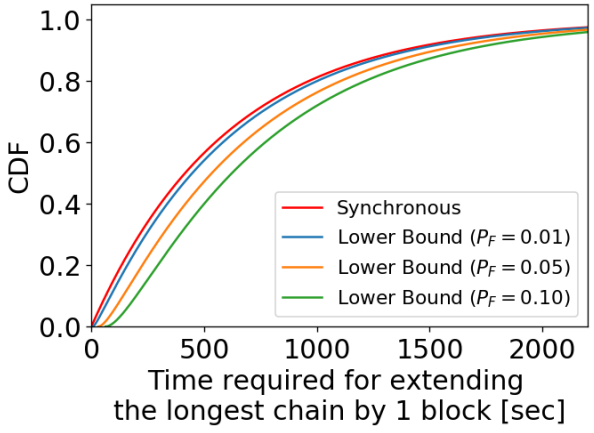
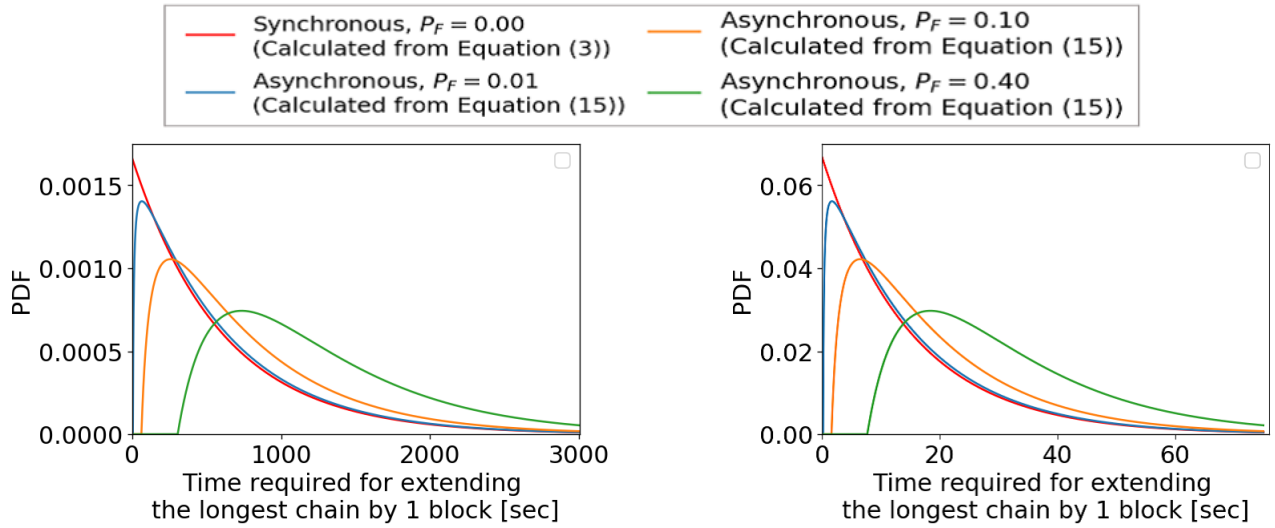


Fig. 3: Lower bounds for CDF of the longest chain growth time which are calculated from Equation (14) ($\lambda = 1/600$)



(a) $\lambda = 1/600$ (b) $\lambda = 1/15$

Fig. 4: Comparison of $g_{\mathrm{nd}}(t)$ which is based on the synchronous model and $g_{\mathrm{lb}}(t)$ which is calculated from our scenario to derive a closed-form lower bound

$$\int_0^\infty g_{\mathrm{nd}}(t)\left(\frac{g_{\mathrm{lb}}(t)}{g_{\mathrm{nd}}(t)}-1\right)^2 dt = -1+\exp(\lambda T_w)+(\lambda T_w)^{-2(\lambda T_w-1)}\cdot\exp(2\lambda T_w)\cdot\Gamma(2\lambda T_w-1,\lambda T_w) \tag{16}$$

$$\int_0^\infty t\cdot g_{\mathrm{lb}}(t)\,dt = C\cdot\lambda^{-1-\lambda T_w}\Big(\Gamma\big(\lambda T_w+2,\lambda T_w\big)-\lambda T_w\Gamma\big(\lambda T_w+1,\lambda T_w\big)\Big) \tag{17}$$

$$\left(T_w=\frac{-\log(1-P_F)}{\lambda}\ ,\ C=\exp\big(\lambda\cdot T_w(1-\log\ T_w)\big)\right)$$

TABLE I: The Pearson distance between $g_{lb}(t)$ and $g_{nd}(t)$

Fork rate	The Pearson distance
0.0010	0.001998
0.0025	0.004994
0.0050	0.009990
0.0100	0.020018
0.0250	0.050559
0.0500	0.103501
0.1000	0.219171
0.2000	0.503638
0.4000	1.453386

TABLE II: An upper bound for Mean time to update the global block height ($\lambda = 1/600$)

Fork rate	An upper bound ($\lambda = 1/600$)	An upper bound ($\lambda = 1/15$)
0.0010	604.42 [sec]	15.11 [sec]
0.0025	609.71 [sec]	15.24 [sec]
0.0050	617.47 [sec]	15.44 [sec]
0.0100	631.15 [sec]	15.78 [sec]
0.0250	666.15 [sec]	16.65 [sec]
0.0500	716.33 [sec]	17.91 [sec]
0.1000	805.19 [sec]	20.13 [sec]
0.2000	968.91 [sec]	24.22 [sec]
0.4000	1305.18 [sec]	32.63 [sec]

4) An upper bound for the mean time to update the global block height: We derive a closed-form lower bound $G_{\mathrm{lb}}(t)$ by considering the scenario that miners only extend the earliest found block at each height and don't mine blocks if they don't know the head block of the longest chain. Therefore, the average time to update the global block height in the scenario is an upper bound for the mean growth time in the scenario that miners extend their local longest chain. The upper bound can be calculated by Equation (17) (A derivation of Equation (17) appears in Appendix C). Calculation results are shown in Table II.

5) A lower bound for the CDF of the time for the longest chain to grow by $n(>1)$ blocks: Since it is difficult to derive a closed-form lower bound $G_{\mathrm{lb}}(n,t)$ for the CDF of the time it takes for the chain to grow by n blocks from Equation (13), we numerically computed $G_{\mathrm{lb}}(n,t)$ by a Monte Carlo simulation that uses the inverse transform method. We generate a random variable by the inverse function of its CDF $G_{\mathrm{lb}}(t)$. The inverse function can be calculated as follows ($G_{\mathrm{lb}}(t)=u\ (u\in[0,1))$).

$$G_{\mathrm{lb}}^{-1}(u) = -T_w\times W\left(-\frac{\big((1-u)/C\big)^{1/(\lambda\cdot T_w)}}{T_w}\right) \tag{18}$$

$$\Big(C=\exp\big(\lambda\cdot T_w(1-\log(T_w))\big)\Big)$$

$W(\cdot)$ is Lambert W function. We can generate a random variable by substituting a uniform random variable $u\in[0,1)$ into $G_{\mathrm{lb}}^{-1}(u)$. By generating the sum of the n random variables repeatedly, we obtained the cumulative relative frequency graphs in Figure 5.

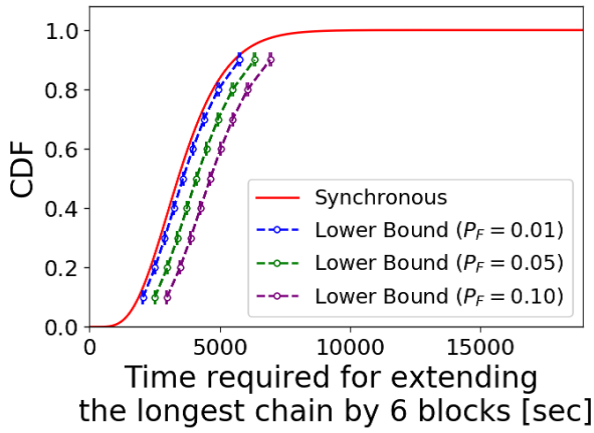


Fig. 5: Numerically calculated lower bound for the CDF of the time for the chain to grow by 6 blocks when $\lambda = 1/600$ (Points are means, error bars are 99.9% CIs).

B. Network simulation results

To calculate the CDF of the longest chain growth time interval, we simulated the growth of the main chain by emulating the topologies of Bitcoin's P2P overlay network. In our simulation, There are two types of nodes: full nodes and miner nodes. Following a behavior similar to Bitcoin's reference client[7], each full node maintains 8 outgoing connections and accepts all incoming connections. Each miner node connects to the 100 full nodes at random (In 2014, the highest degree nodes on the Bitcoin network have more than 90 degrees [19]). The number of full nodes equals 5,000 and the number of miner nodes is equal to 20. The propagation delays on the links were calculated from a normal distribution ($\mathcal{N}(\mu, \sigma), \mu = \sigma = 100$ [milliseconds]), and the bandwidth of each link was determined from a normal distribution $\mathcal{N}(\mu, \sigma), \mu = 4.0, \sigma = 0.8$ [Mbps]. Both values were re-sampled if there was a negative value. We performed 4 network simulations ($\lambda = 1/600$ and the block size equals 1 MB or 10 MB, and $\lambda = 1/15$ and the block size equals 25 kB or 250 kB). Each simulation was run for 50,000 sequential blocks. Simulation results are shown in Figure 6 (Table III and IV show example values of simulation results in Figure 6. (a) and (b)). We use the observed fork rate to compute our lower bound based on Equation (14). To calculate the lower bound $G_{\mathrm{cd}}(1, t)$, we use the longest time it takes for a block to be propagate to all miners as the upper bound D.

V. Discussion

A. Comparison between our lower bound $G_{\mathrm{lb}}(t)$ and the lower bound $G_{\mathrm{nd}}(t)$ based on the bounded delay model

Figure 6 shows that $G_{\mathrm{nd}}(t)$ is a tighter bound than $G_{\mathrm{lb}}(t)$ for most of the time. This is because the average weighted delay, which can be calculated from Equation (9), diverges infinitely if $F(t) = F_{\mathrm{lb}}(t)$. On the other hand, in the constant delay network which is used for deriving the lower bound based on the bounded delay model, the average weighted delay equals the delay diameter of the network D. However, our lower bound is more easily calculated because the observed fork rate is more easily acquired than D. Hence, the advantage of our lower bound is that it is easily and simply calculable at the expense of being tight.

We also compare the time t_1 and t_2 such that $G_{\mathrm{nd}}(t_1) = p$ and $G_{\mathrm{lb}}(t_2) = p$ $(0 < p < 1)$ in Table III and IV. In the case that the block size is 10 MB, $\lambda = 1/600$ and if the cumulative probability $p \leq 0.05$, our lower bound is tighter than the other one. This is because all the amount of computational resource is ignored before the block is propagated to all miners in the constant delay network. Therefore, if the delay diameter of the network D is large, our lower bound is expected to be relatively more befitting in a certain period from beginning to propagate the block to all miners.

B. How to apply our analysis to real-world PoW blockchains and its limitations

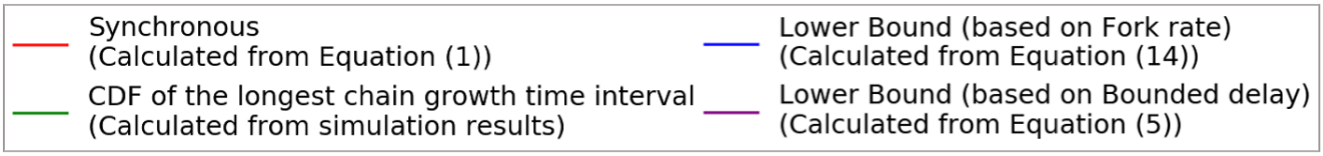
In the real-world network, it is not realistic for all miners to have the same amount of computational resource. Hence, the observed fork rate should be considered for each miner node and be calculated from blocks mined by the miner, but this reduces the reliability of the observed fork probability since the number of fork blocks which can be sampled decreases. Therefore, a method to estimate the fork rate with a small number of sample blocks is desired. However, our lower bound is calculated by only using the observed fork rate and the network block creation rate. If the fork rate seems very low, we can consider that the blockchain network is approximately synchronous by using the Pearson distance (Equation (16)) as the metric. In the future, we will consider the case that the fork rate is not low.

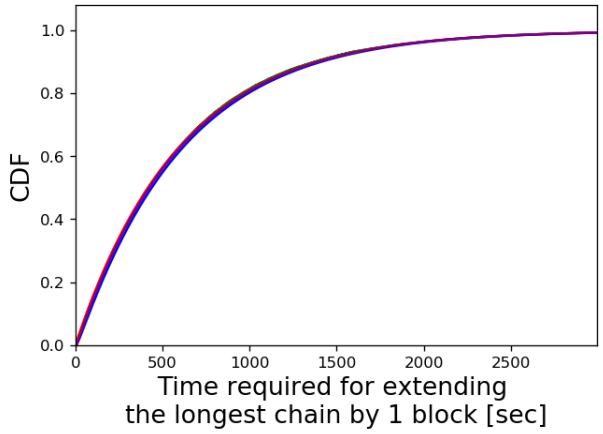
VI. Conclusion

It is desirable to estimate the distribution of the longest chain growth time interval by a simple and low cost method. In this paper, we derive a closed-form lower bound for the CDF by using the fork rate which can be easily observed in the network. Using this lower bound, we also obtain the Pearson distance that can be used as an indicator for judging whether the network can be approximately synchronous. By conducting network simulations, we compared our lower bound with the conventional lower bound based on the bounded delay model. The conventional one is a tighter bound but it requires real network parameters.

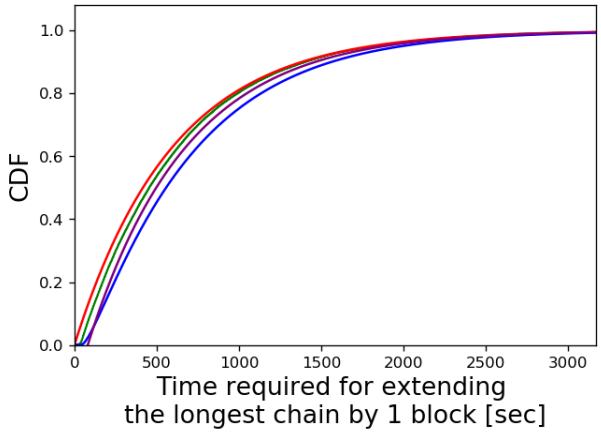
In this paper, we assumed that all nodes have the same amount of computational resource and the network topology is symmetric (These simple assumptions are also used for calculating the fork rate in [13]). In the future, we will explore another fork rate-based method to apply it into real-world PoW blockchains. In addition, we should consider the case that the block generation rate is dynamically changing because miners can join or leave the protocol and the difficulty level evolves over time [20].

[7]https://github.com/bitcoin/bitcoin/

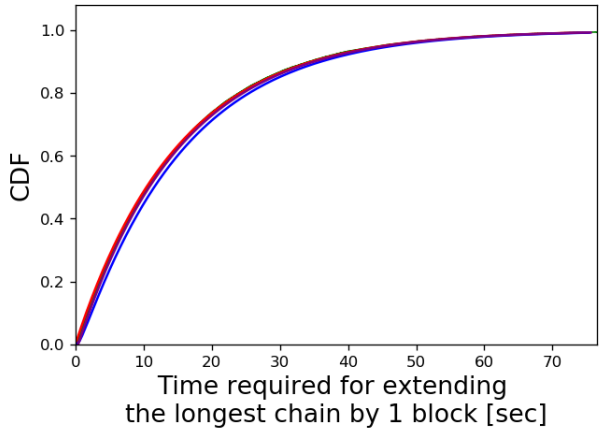




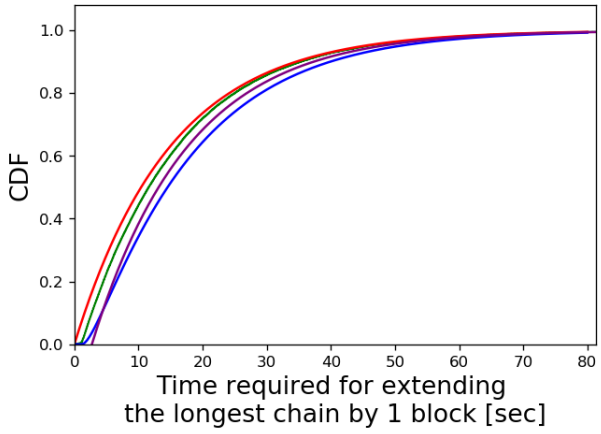
(a) $\lambda = 1/600$, Block Size: 1 MB
Observed fork rate $\hat{P}_F$: 0.00636
Observed delay diameter $\hat{D}$: 7.1626 [sec]



(b) $\lambda = 1/600$, Block Size: 10 MB
Observed Fork Rate $\hat{P}_F$: 0.06182
Observed delay diameter $\hat{D}$: 81.0510 [sec]



(c) $\lambda = 1/15$, Block Size: 25 kB
Observed Fork Rate $\hat{P}_F$: 0.01554
Observed delay diameter $\hat{D}$: 0.4242 [sec]



(d) $\lambda = 1/15$, Block Size: 250 kB
Observed Fork Rate $\hat{P}_F$: 0.07522
Observed delay diameter $\hat{D}$: 2.7291 [sec]

Fig. 6: Each figure shows the CDF of the longest chain growth time interval that was numerically computed by a network simulation, its upper bound, which is calculated in the synchronous model, and its two types of lower bounds. To obtain our lower bound which is built on top of Equation (14), we used the observed fork rate. To calculate the lower bound which is based on the bounded delay model, we used the longest time it takes for the block to be propagated into all miners on the network as the upper bound D.

TABLE III: $\lambda = 1/600, \hat{D} = 7.1626$ and $\hat{P}_F = 0.00636$ from the simulation result in Figure 6. (a).

Cumulative Probability	**Time** t_1 s.t. $G_{\mathrm{lb}}(t_1) = p$	**Time** t_2 s.t. $G_{\mathrm{cd}}(t_2) = p$
$p = 0.01$	15.1 [sec]	13.2 [sec]
$p = 0.05$	43.9 [sec]	37.9 [sec]
$p = 0.50$	437.9 [sec]	423.1 [sec]
$p = 0.95$	1824.9 [sec]	1804.6 [sec]
$p = 0.99$	2792.2 [sec]	2770.3 [sec]

TABLE IV: $\lambda = 1/600, \hat{D} = 81.0510$ and $\hat{P}_F = 0.06182$ from the simulation result in Figure 6. (b).

Cumulative Probability	**Time** t_1 s.t. $G_{\mathrm{lb}}(t_1) = p$	**Time** t_2 s.t. $G_{\mathrm{cd}}(t_2) = p$
$p = 0.01$	64.0 [sec]	87.1 [sec]
$p = 0.05$	109.2 [sec]	111.8 [sec]
$p = 0.50$	556.7 [sec]	496.9 [sec]
$p = 0.95$	1986.9 [sec]	1878.5 [sec]
$p = 0.99$	2968.0 [sec]	2844.2 [sec]

$$\int_{T_w}^{\infty} \exp(-\alpha t + \beta \log t) \cdot (1 - \frac{\gamma}{t})^2 dt = \alpha^{-(\beta+1)} \left((\alpha\gamma)^2 \Gamma(\beta - 1, \alpha T_w) - 2\alpha\gamma\Gamma(\beta, \alpha T_w) + \Gamma(\beta + 1, \alpha T_w)\right) \tag{22}$$

$$\int_{T_w}^{\infty} \exp(-\alpha t + \beta \log t) \cdot (1 - \frac{\gamma}{t}) dt = -\alpha^{-(\beta+1)} \left(\alpha\gamma\Gamma(\beta, \alpha T_w) - \Gamma(\beta + 1, \alpha T_w)\right) \tag{23}$$

$$\Gamma(a + 1, x) = a\Gamma(a, x) + x^a \exp(-x) \tag{24}$$

REFERENCES

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.
[2] Namecoin. https://namecoin.org/.
[3] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," *Ethereum Project Yellow Paper*, vol. 151, 2014.
[4] Y. Sompolinsky and A. Zohar, "Secure high-rate transaction processing in bitcoin," *Proc. 19th Int. Conf. Financial Cryptogr. Data Secur. (FC' 15)*, pp. 507-527, Jan. 2015.
[5] K. Croman, C. Decker, I. Eyal, A. E. Gencer, A. Juels, A. E. Kosba, A. Miller, P. Saxena, E. Shi, E. G. Sirer, D. Song and R. Wattenhofer, "On Scaling Decentralized Blockchains (A Position Paper)," *In 3rd Workshop on Bitcoin and Blockchain Research*, 2016.
[6] M. Ali, J. Nelson, R. Shea, and M. J. Freedman, "Blockstack: A Global Naming and Storage System Secured by Blockchains," *USENIX Annual Technical Conference*, p181-194, 2016.
[7] H. Kalodner, M. Carlsten, P. Ellenbogen, J. Bonneau, and A. Narayanan, "An empirical study of Namecoin and lessons for decentralized namespace design," *WEIS '15: Proceedings of the 14th Workshop on the Economics of Information Security*, June. 2015.
[8] H. Seike, T. Hamada, T. Sumitomo and N. Koshizuka, "Blockchain-Based Ubiquitous Code Ownership Management System without Hierarchical Structure," *2018 IEEE SmartWorld*, Guangzhou, 2018, pp. 271-276.
[9] C. Grunspan and R. P´ erez-Marco, "Double spend races," *International Journal of Theoretical and Applied Finance*, 2018.
[10] Y. Kawase and S. Kasahara, "Transaction-Confirmation Time for Bitcoin: A Queueing Analytical Approach to Blockchain Mechanism," *Queueing Theory and Network Applications: QTNA 2017*, Lecture Notes in Computer Science, vol. 10591, Nov. 2017.
[11] J. A. Garay, A. Kiayias, N. Leonardos, "The bitcoin backbone protocol: Analysis and applications", *Proc. 34th Int. Conf. Theory Appl. Cryptogr. Techn. (EUROCRYPT' 15)*, pp. 281-310, Apr. 2015.
[12] R. Pass, L. Seeman and A. Shelat, "Analysis of the blockchain protocol in asynchronous networks," *Advances in Cryptology - EUROCRYPT 2017*, Springer International Publishing, pp. 643-673, 2017.
[13] C. Decker and R. Wattenhofer, "Information propagation in the Bitcoin network," *IEEE P2P 2013 Proceedings*, Trento, 2013, pp. 1-10.
[14] BitcoinStats. http://bitcoinstats.com/network/propagation/.
[15] M. Rosenfeld, "Analysis of hashrate-based double spending," *ArXiv 1402.2009v1*, Feb. 2014.
[16] K. Pearson, "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *Philosophical Magazine Series 5,* 50(302): 157- 175, 1900.
[17] S. Kullback and R. A. Leibler, "On information and sufficiency," The Annals of Mathematical *Statistics*, vol. 22, no. 1, pp. 79-86, 1951.
[18] M. Sugiyama, T. Suzuki and T. Kanamori, *Density Ratio Estimation in Machine Learning*, Cambridge University Press, 2012.
[19] A. Miller, J. Litton, A. Pachulski, N. Gupta, D. Levin, N. Spring, and B. Bhattacharjee, "Discovering bitcoin's public topology and influential nodes," 2015.
[20] R. Bowden, H.P. Keeler, A.E. Krzesinski and P.G. Taylor, "Block arrivals in the Bitcoin blockchain," *ArXiv:1801.07447v1*, Jan. 2018.

APPENDIX

A. Derivation of Equation (6)

The event of a blockchain fork occurs if miners find blocks before the earliest mined block at one height is propagated to the entire network. Therefore, we calculate the probability that miners fail to find blocks which don't extend the longest chain during the time interval $[0, \infty)$. By defining an infinite increasing sequence of times $t_0 < t_1 < t_2 < \dots$ ($t_{i+1} - t_i = \Delta t,\ i \in \mathbf{N}$), the probability that a blockchain fork occurs during the period $[t_i, t_{i+1})$ can be approximated as follows[8].

$$P_b([t_i, t_{i+1})) \approx 1 - \exp(-\lambda(1 - F(t_i)) \cdot \Delta t) \tag{19}$$

Therefore, using this approximation formula, the probability that a blockchain doesn't occur during the time interval $[0, \infty)$ can be derived as follows.

$$\begin{aligned} P_F &\approx 1 - \prod_{i=0}^{\infty} (1 - (1 - \exp(-\lambda(1 - F(t_i)) \cdot \Delta t)), \\ &= 1 - \prod_{i=0}^{\infty} \exp(-\lambda(1 - F(t_i)) \cdot \Delta t), \\ &= 1 - \exp(-\lambda \sum_{i=0}^{\infty} (1 - F(t_i)) \cdot \Delta t), \\ &= 1 - \exp(-\lambda \int_0^{\infty} (1 - F(t)) dt) \quad (\Delta t \to 0). \end{aligned} \tag{20}$$

B. Derivation of Equation (16)

$$\begin{aligned} &\int_0^{\infty} g_{\mathrm{nd}}(t) \left(\frac{g_{\mathrm{lb}}(t)}{g_{\mathrm{nd}}(t)} - 1 \right)^2 dt = \int_0^{\infty} \lambda \cdot \exp(-\lambda t) dt \\ &+ \int_{T_w}^{\infty} \Bigg(-2\lambda C \exp(-\lambda t + \lambda T_w \log t) \big(1 - \frac{T_w}{t}\big) \\ &+ \lambda C^2 \exp(-\lambda t + 2\lambda T_w \log t) \big(1 - \frac{T_w}{t}\big)^2 \Bigg) dt \\ &\big(T_w = \frac{-\log(1 - P_F)}{\lambda}\ ,\ C = \exp\big(\lambda \cdot T_w (1 - \log\ T_w)\big)\big) \end{aligned}$$

We can simplify Equation (21) into Equation (16) by using Equations (22, 23, 24).

C. Derivation of Equation (17)

$$\begin{aligned} &\int_0^{\infty} t \cdot g_{\mathrm{lb}}(t) dt = \int_0^{T_w} t \cdot 0\ dt \\ &+ \int_{T_w}^{\infty} t \cdot \lambda C \exp(-\lambda t + \lambda T_w \log t) \big(1 - \frac{T_w}{t}\big) dt \\ &= \int_{T_w}^{\infty} \lambda C \exp(-\lambda t + (\lambda T_w + 1) \log t) \big(1 - \frac{T_w}{t}\big) dt \\ &\big(T_w = \frac{-\log(1 - P_F)}{\lambda}\ ,\ C = \exp\big(\lambda \cdot T_w (1 - \log\ T_w)\big)\big) \end{aligned} \tag{25}$$

We can simplify Equation (25) into Equation (17) by using Equation (23).

[8] In the short interval $[t_i, t_{i+1})$, we can assume the stale block generation rate is approximated by $\lambda \cdot (1 - F(t_i))$.