# A Coefficient of Variation Method to Measure the Extents of Decentralization for Bitcoin and Ethereum Networks

Keke Wu[1], Bo Peng[2], Hua Xie[2], and Shaobin Zhan[1]

*(Corresponding author: Shaobin Zhan)*

Shenzhen Institute of Information Technology, Shenzhen, 518172, China[1]

(Email: {wukk, zhansb}@sziit.edu.cn)

Shenzhen Silico Design Technology Co., LTD, Shenzhen, China[2]

*(Received Nov. 20, 2018; Revised and Accepted Aug. 5, 2019; First Online Aug. 5, 2019)*

## Abstract

The most primary advantage of Bitcoin and Ethereum systems is widely understood to be decentralization. However, despite the widely acknowledged importance of this property, most studies on this topic lack quantification, and none of them performs a measurement on the extent of decentralization they achieve in practice. In this paper, we present a coefficient of variation method in probability theory and statistics to quantify decentralization. Using the coefficient of variation, we calculate the dispersion extents of blocks mined and address balances to quantify the extents of decentralization for Bitcoin and Ethereum systems, and the results of calculations indicate that Bitcoin's mining is more approximately 27.3% decentralized than Ethereum with top 19 pool samples, and Bitcoin's wealth is more approximately 16.5% decentralized than Ethereum with 100 samples. Our method can be used to measure the extent of decentralization for any blockchain system.

*Keywords: Bitcoin; Blockchain; Coefficient of Variation; Decentralization; Ethereum*

## 1 Introduction

Bitcoin is a digital currency implementation based on blockchain technology that was invented by Satoshi Nakamoto in 2008 [14]. Bitcoin network is the first digital currency system that has been tested in large scale and long time in history. As a public blockchain platform, for adapting to more complex and flexible application scenarios, Ethereum [3] has further extended the functions of Bitcoin for digital currency transactions, supporting the important feature of smart contract. The common advantage of Bitcoin and Ethereum systems is widely understood to be decentralization that does not have any central authority or server and their networks are peer-to-peer. By storing data across its decentralized network, the blockchain eliminates a number of risks that come with data being held centrally.

Since decentralization is the most important property in blockchain, many studies about the decentralization were proposed. Croman and Gencer *et al.* proposed the technical evaluation of blockchain decentralization systems, mainly focusing on the network congestion or delay to evaluate the performance of the blockchain distributed network [5, 8, 12, 16, 20]. They analyze how fundamental and circumstantial bottlenecks in Bitcoin limit the ability of its current peer-to-peer overlay network to support substantially higher throughputs and lower latencies. Their results suggest that reparameterization of block size and intervals should be viewed only as a first increment toward achieving next-generation, high-load blockchain protocols, and major advances will additionally require a basic rethinking of technical approaches. They offer a structured perspective on the design space for such approaches. Within this perspective, they enumerate and briefly discuss a number of recently proposed protocol ideas and offer several new ideas and open challenges.

Gervais *et al.* revealed that there are many important operations and decisions in Bitcoin system which is not decentralized, and they revealed that some nodes control services, decision-making, transactions and mining in Bitcoin system, finally they gave a way to optimize the decentralization of Bitcoin network [4, 9, 10, 13, 17]. They show that the vital operations and decisions that Bitcoin is currently undertaking are not decentralized. They also show that third-party entities can unilaterally decide to "devalue" any specific set of Bitcoin addresses pertaining to any entity participating in the system. Finally, they explore possible avenues to enhance the decentralization in the Bitcoin system. Ron and Shamir analyzed the transaction data of Bitcoin and revealed the occurrence of large transactions in the Bitcoin system at a certain point in time [11, 15, 18].

These existed research papers above are based on data analysis of Bitcoin transaction behavior to illustrate the drawbacks of the low extent of decentralization in Bitcoin system. Although these studies are mainly concerned about the decentralization of blockchain, none of them performs a measurement on the extent of decentralization they achieve in practice. The closest research work to ours is the paper [19] that only focuses on evaluating a critical value of the number of nodes needed to control over 51% of the network by using a Nakamoto coefficient, rather than quantifying the dispersion of a set of data of blockchain systems, such as blocks mined and address balance and so on.

Herein, we must be able to measure the data dispersion extents of the targets of nodes in blockchain systems before we improve the decentralization. In this paper, we present a coefficient of variation method in probability theory and statistics to measure and quantify the extents of decentralization for blockchain systems. Using the coefficient of variation, we measure the dispersion extents of blocks mined and address balances to quantify the extents of decentralization for blockchain systems. The reminder of this paper is organized as follows. In section 2, we introduce the theory of the coefficient of variation in probability theory and statistics, and the meanings of decentralization in blockchain systems. In section 3, we propose a quantitative measurement method to measure the data dispersion extent based on the coefficient of variation, and in section 4, we calculate the dispersion extents of blocks mined and address balances by using the measurement method. In section 5, we compare the results of the coefficient of variation between Bitcoin and Ethereum systems. Finally, we conclude the paper in section 6.

# 2 Background

In this section, we first introduce the theory of the coefficient of variation in probability theory and statistics, and we illustrate the meanings of decentralization in blockchain systems to introduce the measurements of decentralization.

## 2.1 Coefficient of Variation

In probability theory and statistics, the coefficient of variation, also known as relative standard deviation, is a standardized measure of dispersion of a probability distribution or frequency distribution. The coefficient of variation ($c_v$) is defined as the ratio of the standard deviation ($\sigma$) to the mean ($\mu$): $c_v = \sigma/\mu$. It shows the extent of variability in relation to the mean of the population. The coefficient of variation should be computed only for data measured on a ratio scale, as these are the measurements that allow the division operation.

The coefficient of variation is useful because the standard deviation of data must always be understood in the context of the mean of the data. In contrast, the actual value of the coefficient of variation is independent of the unit in which the measurement has been taken, so it is a dimensionless number. For comparison between data sets with different units or widely different means, one should use the coefficient of variation instead of the standard deviation. The value of coefficient of variation is larger, the greater the degree of dispersion.

In this paper, we use the coefficient of variation to measure the extents of decentralization with dimensionless numbers for Bitcoin and Ethereum systems. We consider the extents with the two targets: blocks mined, and address balance.

## 2.2 Decentralization

Decentralization is the process by which the activities of an organization, particularly those regarding planning and decision making, are distributed or delegated away from a central, authoritative location or group.

In blackchain systems, the decentralization means that no single individual can destroy transactions in the network, and any transaction request requires the consensus of most participants. Bitcoin and Ethereum also have a peer-to-peer network for disseminating block and transaction information. Both Bitcoin and Ethereum also contain full nodes, which serve two critical roles: (1) to relay blocks and transactions to miners (2) and to answer queries for end users about the state of the blockchain. In the Bitcoin and Ethereum protocols, users submit transactions for miners to sequence into blocks. Better decentralization of miners means higher resistance against censorship of individual transactions. Specifically, a decentralized system (like Bitcoin or Ethereum) is composed of a set of decentralized subsystems (like mining, exchanges, nodes, developers, clients, and so on). Srinivasan *et al.* used these six subsystems to calculate a critical value with a Nakamoto coefficient, and to illustrate how many nodes needed to control over 51% of the network in Bitcoin or Ethereum [19].

In this paper, we will calculate the dispersion degrees by two targets (blocks mined and address balance) to measure the extents of decentralization for Bitcoin and Ethereum systems. Please note: you may decide to use different subsystems or targets based on which ones you consider essential to decentralization of the system as a whole.

# 3 Measurement Method

As mentioned above, we use the coefficient of variation to measure the dispersion degree for dimensionless data sets in Bitcoin and Ethereum systems. Herein, we elaborate the inferring process of coefficient of variation according to the variance and standard deviation, and then we present the formula of the coefficient of variation as a measurement method.

In probability theory and statistics, variance is the expectation of the squared deviation of a random variable from its mean. Informally, it measures how far a set of (random) numbers are spread out from their average value. Variance is a central role in statistics, where some ideas that use it include descriptive statistics, statistical inference, hypothesis testing, and Monte Carlo sampling. Variance is an important tool, where statistical analysis of data is common. Variance is the square of the standard deviation, the second central moment of a distribution, and the covariance of the random variable with itself, and it is often represented by $Var(X)$. If the generator of random variable $X$ is discrete with probability mass function $x_1 \vdash\!\rightarrow p_1, x_2 \vdash\!\rightarrow p_2, \cdots, x_n \vdash\!\rightarrow p_n$ then

$$Var(X) = \sum_{i=1}^{n} p_i \cdot (x_i - \mu)^2,$$

where $\mu$ is the expected value, i.e. $\mu = \sum_{i=1}^{n} p_i x_i$. When such a discrete weighted variance is specified by weights whose sum is not 1, one divides by the sum of the weights. Therefore, in statistics, the variance of a set of $n$ equally likely values can be written as

$$Var(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2,$$

where $\mu$ is the average value, i.e. $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$.

In statistics, the standard deviation (SD, also represented by the lower case Greek letter sigma $\sigma$) is a measure that is used to quantify the amount of variation or dispersion of a set of data values. A low standard deviation indicates that the data points tend to be close to the mean (also called the expected value) of the set, while a high standard deviation indicates that the data points are spread out over a wider range of values. The standard deviation of a random variable, statistical population, data set, is the square root of its variance, i.e.,

$$\sigma = \sqrt{Var(X)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}$$

where $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Coefficient of variation is another statistic to measure the degree of variation of observed values in data. When comparing the degree of variability of two or more data, the standard deviation can be used directly if the unit of measurement is the same as the average. If the unit and/or average are different, the standard deviation could not be used to compare the degree of variation, but the ratio of the standard deviation to the average (relative value) should be used to compare. The ratio of standard deviation to average is called coefficient of variation ($c_v$), i.e.,

$$c_v = \frac{\sigma}{\mu}$$

where $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}$ and $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$. Coefficient of variation can eliminate the effect of unit and/or average differences on the comparison of variability between two or more data sets. Therefore, coefficient of variation can be used to calculate the different dimensionless data sets between Bitcoin and Ethereum. More theories about variance, standard deviation, and coefficient of variation, please refer to the probability theory and statistics textbooks.

# 4 Calculations

According to the measurement method presented above, let's now calculate coefficients of variation for the blocks mined and address balance in Bitcoin and Ethereum networks. We can calculate the decentralized extents each of them according to coefficients of variation.

## 4.1 Blocks Mined

The quantity of blocks mined reflects the priority to account in blockchain networks. The data is more dispersed (or polarized), the ability of the miners controlling the entire blockchain network is more powerful, and the extent of decentralization of the blockchain network is lower. On the contrary, the data is more average, the ability of the miners controlling the entire blockchain network is weaker, and the extent of decentralization of the blockchain network is higher.

Hence, we use the coefficient of variation as the measurement method to calculate and quantify the degrees of data dispersion for Bitcoin and Ethereum networks, and we can compare the extents of decentralization between them.

### 4.1.1 Coefficient of Variation of Bitcoin Blocks Mined

We catch the data of Bitcoin blocks mined over the last 7 days from the website btc.com on Oct. 25, 2018, where Bitcoin's data will be updated in real time, as show in Figure 1.

The green frame in Figure 1 is the data top list of Bitcoin blocks mined. The data distribution of top 19 blocks mined in Bitcoin network is as show in Figure 2. We can see that the top 7 miners mined most blocks, and they can influence the decentralized extent of entire Bitcoin network.

According to the measurement method presented above, we use the random variable $X = \{161, 136, 110, 101, 95, 90, 73, 16, 15, 15, 12, 11, 10, 10, 7, 6, 4, 2, 1\}$ and sample number $n = 19$ to calculate the coefficient of

| | Pool | Hashrate Share | Hashrate | Blocks Mined | Empty Blocks Count | Empty Blocks Percentage | Avg. Block Size (Bytes) | Avg. Tx Fees Per Block (BTC) | Tx Fees % of Block Reward |
|---|---|---|---|---|---|---|---|---|---|
| 1 | BTC.com | 16.65 % | 8.20 EH/s | 161 | 4 | 2.48 % | 1,055,771 | 0.13199030 | 1.06 % |
| 2 | AntPool | 14.06 % | 6.93 EH/s | 136 | 1 | 0.74 % | 730,825 | 0.09653624 | 0.77 % |
| 3 | ViaBTC | 11.38 % | 5.60 EH/s | 110 | 1 | 0.91 % | 953,433 | 0.10927257 | 0.87 % |
| 4 | SlushPool | 10.44 % | 5.14 EH/s | 101 | 1 | 0.99 % | 974,467 | 0.11446046 | 0.92 % |
| 5 | BTC.TOP | 9.82 % | 4.84 EH/s | 95 | 1 | 1.05 % | 1,005,207 | 0.11535621 | 0.92 % |
| 6 | F2Pool | 9.31 % | 4.58 EH/s | 90 | 0 | 0.00 % | 1,062,745 | 0.12640051 | 1.01 % |
| 7 | Poolin | 7.55 % | 3.72 EH/s | 73 | 1 | 1.37 % | 1,069,483 | 0.12907499 | 1.03 % |
| 8 | Huobi.pool | 1.65 % | 814.82 PH/s | 16 | 0 | 0.00 % | 940,827 | 0.11471356 | 0.92 % |
| 9 | BitClub | 1.55 % | 763.89 PH/s | 15 | 0 | 0.00 % | 877,048 | 0.10521577 | 0.84 % |
| 10 | DPOOL | 1.55 % | 763.89 PH/s | 15 | 0 | 0.00 % | 959,886 | 0.17730335 | 1.42 % |
| 11 | BitFury | 1.24 % | 611.11 PH/s | 12 | 0 | 0.00 % | 683,110 | 0.12251172 | 0.98 % |
| 12 | Bixin | 1.14 % | 560.19 PH/s | 11 | 0 | 0.00 % | 1,000,779 | 0.07636004 | 0.61 % |
| 13 | 58COIN | 1.03 % | 509.26 PH/s | 10 | 0 | 0.00 % | 633,567 | 0.05980556 | 0.48 % |
| 14 | WAYI.CN | 1.03 % | 509.26 PH/s | 10 | 0 | 0.00 % | 1,207,203 | 0.14989612 | 1.20 % |
| 15 | Bitcoin.com | 0.72 % | 356.48 PH/s | 7 | 0 | 0.00 % | 800,584 | 0.07289596 | 0.58 % |
| 16 | BWPool | 0.62 % | 305.56 PH/s | 6 | 0 | 0.00 % | 1,149,830 | 0.10385478 | 0.83 % |
| 17 | KanoPool | 0.41 % | 203.70 PH/s | 4 | 0 | 0.00 % | 992,993 | 0.19477824 | 1.56 % |
| 18 | BTPOOL | 0.21 % | 101.85 PH/s | 2 | 0 | 0.00 % | 1,125,270 | 0.08888516 | 0.71 % |
| 19 | CKPool | 0.10 % | 50.93 PH/s | 1 | 0 | 0.00 % | 831,358 | 0.09131175 | 0.73 % |

Figure 1: Miner distribution sorted by blocks mined in Bitcoin network over the last 7 days (Data from https://btc.com/ on Oct. 25, 2018 [1])

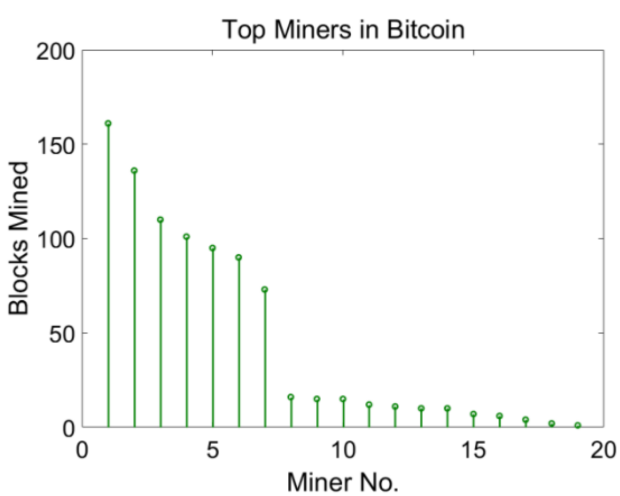


Figure 2: The distribution of top 19 blocks mined in Bitcoin network

variation that is introduced above as follows.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \approx 46.05$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2} \approx 51.38$$

$$c_v = \frac{\sigma}{\mu} \approx 1.12.$$

We obtain the value of coefficient of variation of Bitcoin blocks mined is approximately equal to 1.12.

#### 4.1.2 Coefficient of Variation of Ethereum Blocks Mined

In the same way, we obtain the Ethereum block data over the last 7 days on Oct. 25, 2018 from the website etherscan.io, where Ethereum's data will be updated in real time, as show in Figure 3.

The red frame in Figure 3 is the data top list of Ethereum blocks mined. The data distribution of top 19 blocks mined in Ethereum network is as show in Figure 4. We can see that the only top 5 miners mined most blocks, and they can influence the decentralized extent of entire Ethereum network.

We also use the random variable $Y$ = {11389, 9569, 5711, 4334, 3748, 816, 729, 591, 589, 461, 378, 329, 315,

| Rank | Address | Blocks Mined | Percentage |
|---|---|---|---|
| 1 | 0xea674fdde714fd979de3edf0f56aa9716b898ec8 (Ethermine) | 11389 | 26.6073% |
| 2 | 0x5a0b54d5dc17e0aadc383d2db43b0a0d3e029c4c (SparkPool) | 9569 | 22.3554% |
| 3 | 0x829bd824b016326a401d083b33d092293333a830 (F2Pool_2) | 5711 | 13.3422% |
| 4 | 0x52bc44d5378309ee2abf1539bf71de1b7d7be3b5 (Nanopool) | 4334 | 10.1252% |
| 5 | 0xb2930b35844a230f00e51431acae96fe543a0347 (MiningPoolHub_1) | 3748 | 8.7562% |
| 6 | 0x2a65aca4d5fc5b5c859090a6c34d164135398226 (DwarfPool_1) | 816 | 1.9064% |
| 7 | 0xd4383232c8d1dbe0e03bdfab849871fa17e61807 | 729 | 1.7031% |
| 8 | 0x52e44f279f4203dcf680395379e5f9990a69f13c (bw) | 591 | 1.3807% |
| 9 | 0x70aec4b9cffa7b55c0711b82dd719049d615e21d | 589 | 1.3760% |
| 10 | 0x2a5994b501e6a560e727b6c2de5d856396aadd38 | 461 | 1.0770% |
| 11 | 0x35f61dfb08ada13eba64bf156b80df3d5b3a738d | 378 | 0.8831% |
| 12 | 0xcc16e3c00dbbe76603fa833ec20a48f786dfe610 | 329 | 0.7686% |
| 13 | 0x09ab1303d3ccaf5f018cd511146b07a240c70294 (MinerallPool) | 315 | 0.7359% |
| 14 | 0x005e288d713a5fb3d7c9cf1b43810a98688c7223 | 301 | 0.7032% |
| 15 | 0xb75d1e62b10e4ba91315c4aa3facc536f8a922f5 | 288 | 0.6728% |
| 16 | 0x84a0d77c693adabe0ebc48f88b3fff010577051 | 256 | 0.5981% |
| 17 | 0x6a7a43be33ba930fe58f34e07d0ad6ba7adb9b1f (Coinotron_3) | 241 | 0.5630% |
| 18 | 0x4bb96091ee9d802ed039c4d1a5f6216f90f81b01 (Ethpool_2) | 229 | 0.5350% |
| 19 | 0x4c549990a7ef3fea8784406c1eecc98bf4211fa5 | 198 | 0.4626% |

Figure 3: Miner distribution sorted by blocks mined in Ethereum network over the past 7 days (Data from https://etherscan.io/ on Oct. 25, 2018 [6])

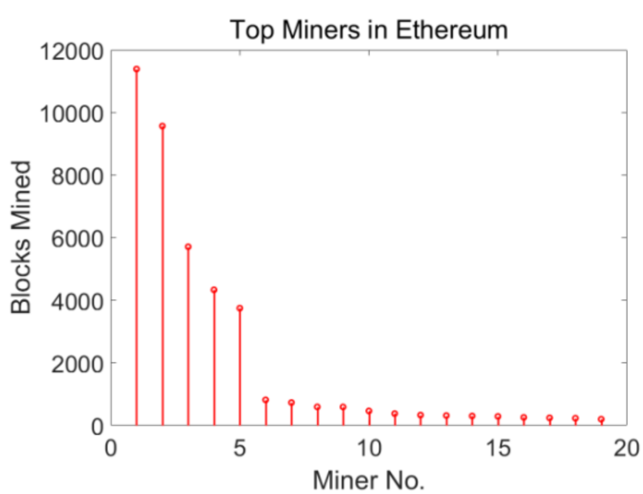


Figure 4: The distribution of top 19 blocks mined in Ethereum network

301, 288, 256, 241, 229, 198} and sample number $n = 19$ to calculate the coefficient of variation as follows.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} y_i \approx 2130.11$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \mu)^2} \approx 3271.55$$

$$c_v = \frac{\sigma}{\mu} \approx 1.54.$$

We obtain the value of coefficient of variation of Ethereum blocks mined is approximately equal to 1.54.

## 4.2 Address Balance

This index examines the addresses of the first 100 tokens, and accumulative total tokens as a percentage of the total tokens in the blockchain. We believe that the decentralized blockchain should also decentralize wealth, and the more centralized tokens means that institutions or individuals with a large number of tokens are more likely to manipulate token prices.

### 4.2.1 Coefficient of Variation of Bitcoin Address Balance

We catch the data of Bitcoin address balance (token) from the website btc.com on Oct. 25, 2018 as show in Figure 5.

The green frame in Figure 5 is the data top list of Bitcoin address balances (tokens). The data distribution

| # | Address | Balance | Last 30 Days Tx Count | First Tx Time | Last Tx Time |
|---|---|---|---|---|---|
| 1 | 3D2oetdNuZUqQHPJmcMDDHYoqkyNVsFk9r | 133,317.23128155 | 83 | 2017-01-05 20:34:15 | 2018-10-25 02:53:50 |
| 2 | 16ftSEQ4ctQFDtVZiUBusQUjRrGhM3JYwe | 129,234.33723943 | 19 | 2017-12-08 15:51:10 | 2018-10-25 11:39:32 |
| 3 | 16rCmCmbuWDhPjWTrpQGaU3EPdZF7MTdUk | 107,203.07546044 | 6 | 2016-02-28 02:00:09 | 2018-10-25 11:39:32 |
| 4 | 3Cbq7aT1tY8kMxWLbitaG7yT6bPbKChq64 | 98,042.49937302 | 8 | 2017-09-09 00:41:05 | 2018-10-25 11:39:32 |
| 5 | 3Nxwenay9Z8Lc9JBiywExpnEFiLp6Afp8v | 97,848.28496144 | 6 | 2015-10-16 22:43:06 | 2018-10-25 11:39:32 |
| 6 | 183hmJGRuTEi2YDCWy5iozY8rZtFwVgahM | 85,947.34736772 | 6 | 2018-07-01 21:29:21 | 2018-10-25 11:39:32 |
| 7 | 1FeexV6bAHb8ybZjqQMjJrcCrHGW9sb6uF | 79,957.19635956 | 6 | 2011-03-01 18:26:19 | 2018-10-25 11:39:32 |
| 8 | 1HQ3Go3ggs8pFnXuHVHRytPCq5fGG8Hbhx | 69,370.12335943 | 6 | 2013-04-10 05:03:36 | 2018-10-25 11:39:32 |
| 9 | 1PnMfRF2enSZnR6JSexxBHuQnxG8Vo5FVK | 66,452.07925125 | 7 | 2013-11-23 03:06:31 | 2018-10-25 11:39:32 |
| 10 | 1AhTjUMztCihiTyA4K6E3QEpobjWLwKhkR | 66,378.81086167 | 6 | 2014-02-25 13:33:06 | 2018-10-25 11:39:32 |
| 11 | 1DiHDQMPFu4p84rkLn6Majj2LCZZZRQUaa | 66,235.82586355 | 6 | 2013-11-23 08:08:37 | 2018-10-25 11:39:32 |
| 12 | 1EBHA1ckUWzNKN7BMfDwGTx6GKEbADUozX | 66,233.75898250 | 7 | 2013-11-23 01:05:19 | 2018-10-25 11:39:32 |
| 13 | 18rnfoQgGo1HqvVQaAN4QnxjYE7Sez9eca | 63,600.04702741 | 20 | 2014-10-24 18:40:08 | 2018-10-25 11:39:32 |
| 14 | 34xp4vRoCGJym3xR7yCVPFHoCNxv4Twseo | 55,482.91464262 | 28 | 2018-10-18 20:59:18 | 2018-10-25 11:39:32 |
| 15 | 1LdRcdxfbSnmCYYNdeYpUnztiYzVfBEQeC | 53,880.05876207 | 6 | 2014-05-28 06:49:42 | 2018-10-25 11:39:32 |
| | ... | ... | | ... | |
| 97 | 1aXzEKiDJKzkPxTZy9zGc3y1nCDwDPub2 | 10,900.00003664 | 5 | 2016-08-04 18:29:39 | 2018-10-23 19:50:56 |
| 98 | 33ZNiyx5Z5CMkULX7ENvcKKxFNCzGJv5vQ | 10,885.20660310 | 4 | 2018-07-06 18:17:44 | 2018-10-23 19:50:56 |
| 99 | 155fzsEBHy9Ri2bMQ8uuuR3tv1YzcDywd4 | 10,845.61928398 | 90 | 2015-01-28 09:25:06 | 2018-10-26 04:27:50 |
| 100 | 1F34duy2eeMz5mSrvFepVzy7Y1rBsnAyWC | 10,770.52537305 | 3 | 2011-08-09 06:14:47 | 2018-10-23 19:50:56 |

Figure 5: Address Balance in Bitcoin network (Data from https://btc.com/ on Oct. 25, 2018 [2])

of top 100 address balances (tokens) in Bitcoin network is as show in Figure 6.

Herein we can ignore the decimal digits since the values of address balances are very huge. Therefore, we use the random variable $X = \{133317$, 129234, 107203, 98042, 97848, 85947, 79957, 69370, 66452, 66379, 66236, 66234, 63600, 55483, 53880, 53000, 52431, 51830, 48500, 45899, 40593, 40474, 40438, 40414, 40054, 40000, 36000, 35612, 34010, 32957, 32841, 32796, 32500, 32490, 31925, 31270, 31085, 31000, 30108, 29999, 29772, 29683, 28151, 27833, 27683, 27496, 26215, 25489, 25409, 25403, 25378, 25302, 25272, 25160, 25064, 24000, 23228, 22891, 22211, 22173, 22100, 21603, 20934, 20263, 20008, 20000, 19414, 17955, 17817, 16252, 16224, 16000, 15746, 15500, 15000, 15000, 15000, 14850, 14627, 14500, 14316, 14000, 13900, 13576 ,13000 12800, 12553, 12000, 11927, 11837, 11800, 11337, 11251, 11102, 10960, 10910, 10900, 10885, 10846, $10771\}$ and sample number $n = 100$ to calculate the coefficient of variation as follows.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \approx 32906.88$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2} \approx 25055.40$$

$$c_v = \frac{\sigma}{\mu} \approx 0.76.$$

We obtain the value of coefficient of variation of Bitcoin address balances is approximately equal to 0.76.

### 4.2.2 Coefficient of Variation of Ethereum Address Balance

We also catch the data of Ethereum address balance from the website etherscan.io on Oct. 25, 2018 as show in Figure 7.

The red frame in Figure 7 is the data top list of Ethereum address balances (tokens). The data distribution of top 100 address balances (tokens) in Ethereum network is as show in Figure 8.

Herein we still can ignore the decimal digits since the values of address balances are very huge. Therefore, we use the random variable $Y = \{1538423$, 1510066, 1507810, 1483159, 1378754, 1024185, 1004999, 1000000, 988888, 959123, 825000, 817061, 801053, 672785, 672524, 670941, 658443, 560000, 558117, 552124, 549774, 530000, 505000, 493015, 483000, 450000, 450000, 450000, 436000, 427828, 403085, 395433, 380000, 369023, 365003, 350001, 345741, 325000, 319500, 306276, 281380, 275000, 267786, 254248, 250000, 245342, 245300, 234322, 232419, 221195, 220523, 219824, 207438, 204364, 204176, 203527, 203468, 200782, 195524, 193737, 190905, 190121, 189000, 187068, 185591, 183371, 180001, 176650, 172224, 169032, 166602, 164998, 163197, 150000, 142943, 141354, 137476, 135284, 132930, 132288, 131340, 130379, 130000, 128529, 126850, 125266, 123450, 122862, 121861, 120347, 114939, 113762, 110195, 109488, 109381, 108761, 107866, 107371, 106712, $105114\}$ and sample number $n = 100$ to calculate the coefficient

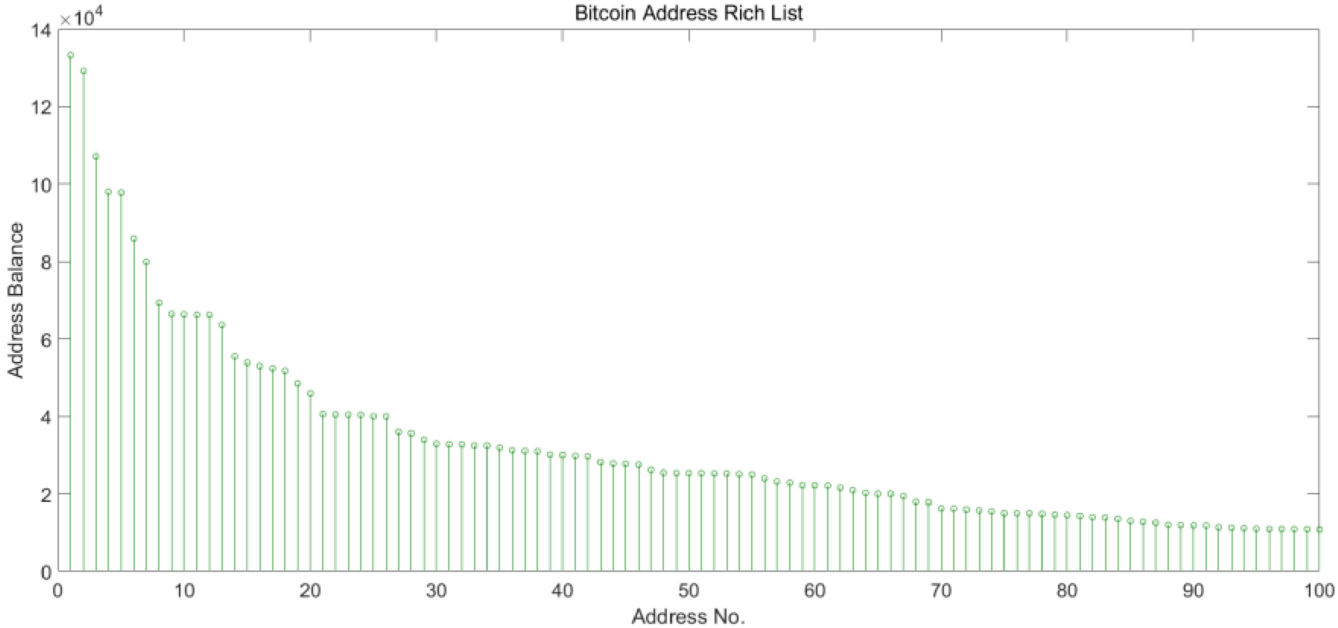


Figure 6: The distribution of top 100 address balance in Bitcoin network

| Rank | Address | Balance | Percentage | TxCount |
|---|---|---|---|---|
| 1 | 0x281055afc982d96fab65b3a49cac8b878184cb16 | 1,538,423.10656596 Ether | 1.49652056% | 519 |
| 2 | 0x6f46cf5569aefa1acc1009290c8e043747172d89 | 1,510,065.64213014 Ether | 1.46893548% | 499 |
| 3 | 0x90e63c3d53e0ea496845b7a03ec7548b70014a91 | 1,507,810.43875773 Ether | 1.46674170% | 446 |
| 4 | 0x742d35cc6634c0532925a3b844bc454e4438f44e | 1,483,159.05734310 Ether | 1.44276176% | 2163 |
| 5 | 0x53d284357ec70ce289d6d64134dfac8e511c8a3d | 1,378,754.09306818 Ether | 1.34120050% | 14994 |
| 6 | 0xc02aaa39b223fe8d0a0e5c4f27ead9083c756cc2 | 1,024,180.85887157 Ether | 0.99628490% | 192870 |
| 7 | 0x61edcdf5bb737adffe5043706e7c5bb1f1a56eea | 1,004,999.00001000 Ether | 0.97762550% | 118 |
| 8 | 0xab7c74abc0c4d48d1bdad5dcb26153fc8780f83e | 1,000,000.01146312 Ether | 0.97276267% | 403 |
| 9 | 0xbe0eb53f46cd790cd13851d5eff43d12404d33e8 | 988,888.05476810 Ether | 0.96195338% | 5 |
| 10 | 0xfbb1b73c4f0bda4f67dca266ce6ef42f520fbb98 | 959,169.02268714 Ether | 0.93304381% | 6966260 |
| 11 | 0xfca70e67b3f93f679992cd36323eeb5a5370c8e4 | 824,999.89932905 Ether | 0.80252910% | 36 |
| 12 | 0xdc76cd25977e0a5ae17155770273ad58648900d3 | 817,060.63884765 Ether | 0.79480608% | 138 |
| 13 | 0xe853c56864a2ebe4576a807d26fdc4a0ada51919 | 801,052.79895972 Ether | 0.77923425% | 147 |
| 14 | 0xf27daff52c38b2c373ad2b9392652ddf433303c4 | 672,784.62216252 Ether | 0.65445976% | 97 |
| 15 | 0x3d2e397f94e415d7773e72e44d5b5338a99e77d9 | 672,524.35429759 Ether | 0.65420658% | 81 |
| | ... | ... | ... | |
| 95 | 0x955a27306f1eb21757ccbd8daa2de82675aabc36 | 109,380.81181740 Ether | 0.10640140% | 34 |
| 96 | 0x21346283a31a5ad10fa64377e77a8900ac12d469 | 108,761.24794052 Ether | 0.10579871% | 38 |
| 97 | 0xe8507b1532fc44e41b48efe45cf4abf92c5767c3 | 107,866.00000000 Ether | 0.10492785% | 1 |
| 98 | 0x3bc643a841915a267ee067b580bd802a66001c1d | 107,371.19229881 Ether | 0.10444652% | 108 |
| 99 | 0xdb8c6862ea4f5cc843c4b3ed75eb8951714b7635 | 106,711.69360026 Ether | 0.10380498% | 1309 |
| 100 | 0x692190b4a5d3524b6fed0465e7400c07d09db954 | 105,113.55510868 Ether | 0.10225038% | 28 |

Figure 7: Address Balances in Ethereum network (Data from https://etherscan.io/ on Oct. 25, 2018 [7])

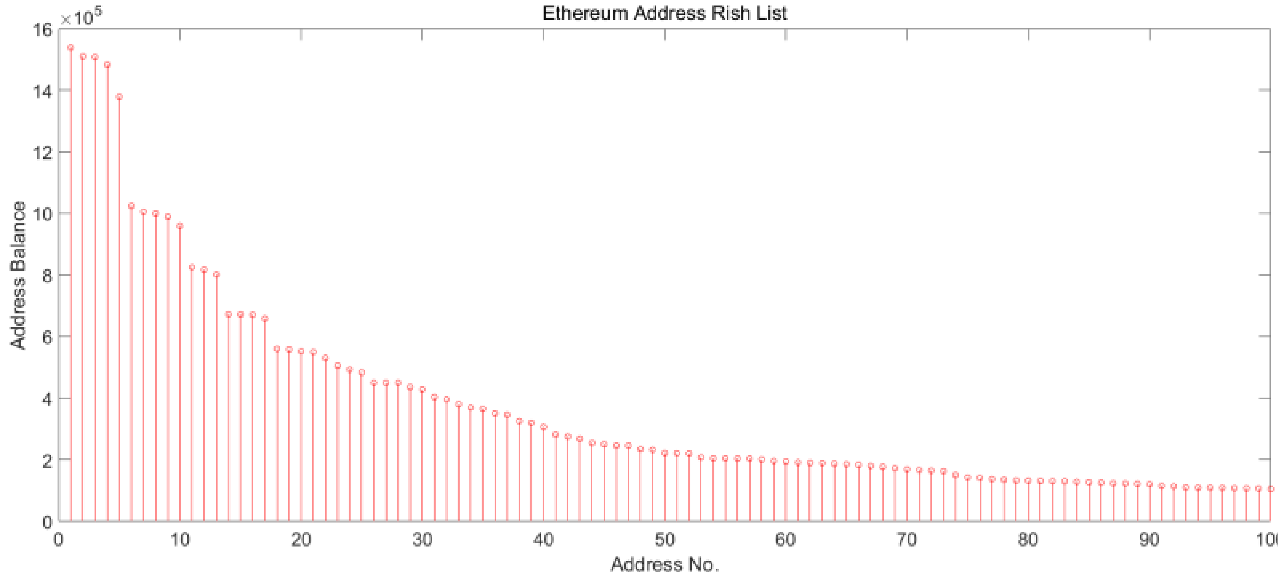


Figure 8: The distribution of top 100 address balances in Ethereum network

of variation as follows.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} y_i \approx 377230.01$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\mu)^2} \approx 345119.60$$

$$c_v = \frac{\sigma}{\mu} \approx 0.91.$$

We also obtain the value of coefficient of variation of Ethereum address balances is approximately equal to 0.91.

# 5 Comparison and Analysis

We present the comparison of decentralization extents between Bitcoin and Ethereum, and we analyze the centralized and decentralized influences in Bitcoin and Ethereum networks.

## 5.1 Blocks Mined Index

This index examines how many individual or organizational unions are needed to control more than 50% account power. For example, how many pools in PoW will add up to 50% of the total net power. This index intuitively reflects the difficulty of controlling a digital currency through 51% attacks. We believe that the more decentralized the blockchain, the less likely it is to control the entire blockchain by controlling a few individuals or organizations.

According to the results of calculations of coefficients of variation above, the value of coefficient of variation of Ethereum blocks mined is larger than the value of coefficient of variation of Bitcoin blocks mined, as show in Figure 9.

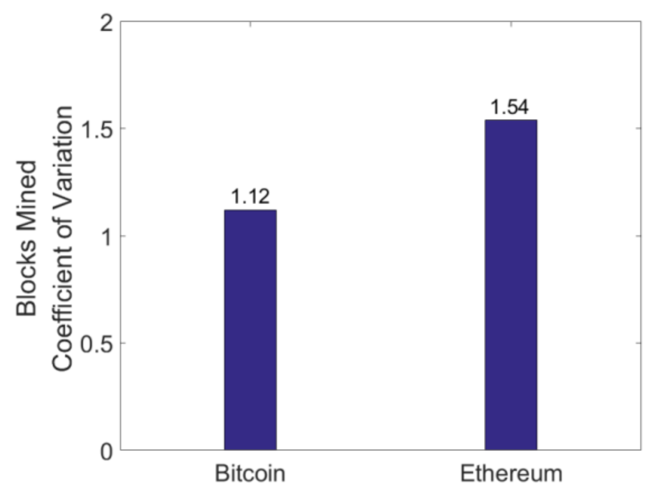


Figure 9: Coefficients of variation of blocks mined comparison between Bitcoin and Ethereum

As shown in Figure 9, Bitcoin mining is more decentralized than Ethereum as measured by blocks mined over the past 7 days. Ethereum mining is somewhat more centralized.

## 5.2 Address Balance Index

This index is a more controversial indicator, because many people would argue that addresses with a large number of tokens may be exchanges. Actually, those tokens are not exchanges, but are temporarily deposited in exchanges. Herein, we still believe that address balance decentralization is an important factor in the real decentralization of digital money.

According to the results of calculations of coefficients of variation above, the value of coefficient of variation of Ethereum address balance is also larger than the value of coefficient of variation of Bitcoin address balance, as show in Figure 10.

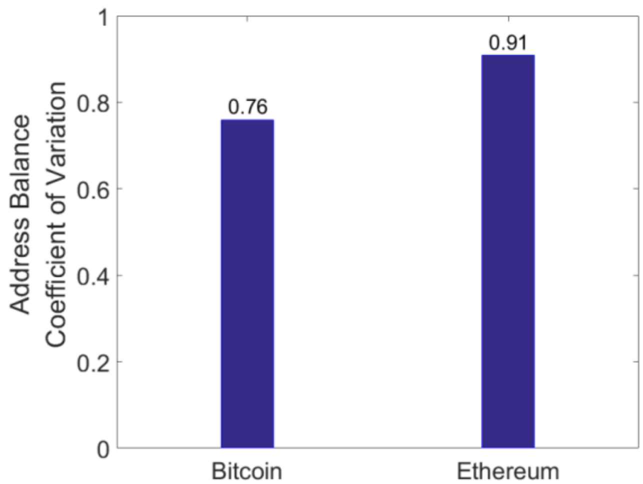


Figure 10: Coefficients of variation of address balances comparison between Bitcoin and Ethereum

As shown in Figure 10, Bitcoin's wealth is more decentralized than Ethereum as measured by address balances. Ethereum's wealth is somewhat more centralized.

### 5.3 Extents of Decentralization

To sum up the coefficient of variation results of calculations above, we list the quantitative extents of decentralization for Bitcoin and Ethereum networks as the following table.

Table 1: Extents of decentralization

| Coefficient of Variation | Blocks Mined (Top 19) | Address Balance (Top 100) |
|---|---|---|
| Bitcoin Network | 1.12 | 0.76 |
| Ethereum Network | 1.54 | 0.91 |
| Comparison (Bitcoin is less than Ethereum) | 27.3% | 16.5% |

This table shows that the extents of decentralization of Bitcoin network are 1.12, 1.54 respectively, and the extents of decentralization of Ethereum network are 0.76, 0.91 respectively. Hence, the extents of decentralization of Bitcoin network are more 27.3%, 16.5% large than Ethereum respectively.

## 6 Conclusions

Decentralization is the most important property of networks like Bitcoin and Ethereum. It is critical to be able to measure the extents of decentralization. More importantly, the coefficient of variation method is one such general measurement method adapting to quantify dispersion for any data set. Therefore, given a proposed blockchain network, we can calculate its coefficient of variation for kinds of targets which you think they are important, and analyze whether this is plausibly a decentralization bottleneck for the network.

## Acknowledgments

This work was supported by the Guangdong Natural Science Foundation (Grant No. 2018A030313746), and the Basic Research Project of Shenzhen (Grant No. JCYJ20170817114239348).

## References

[1] BTC, *Pool Distribution*, Oct. 25, 2018. (`https://btc.com/stats/pool`)

[2] BTC, *Address Rich List*, Oct. 25, 2018. (`https://btc.com/stats/rich-list`)

[3] V. Buterin, "A next generation smart contract and decentralized application platform," *Ethereum White Paper*, 2013. (`http://blockchainlab.com/pdf/Ethereum_white_paper-a_next_generation_smart_contract_and_decentralized_application_platform-vitalik-buterin.pdf`)

[4] M. Conoscenti, A. Vetr, J. C. D. Martin, "Peer to peer for privacy and decentralization in the internet of things," in *IEEE/ACM 39th IEEE International Conference*, 2017. DOI: 10.1109/ICSE-C.2017.60.

[5] K. Croman, C. Decker, and I. Eyal, *et al.*, "On scaling decentralized blockchains," in *Financial Cryptography and Data Security*, pp. 106-125, 2016.

[6] Etherscan, *Top Miners by Blocks*, Oct. 25, 2018. (`https://etherscan.io/stat/miner?range=7&blocktype=blocks`)

[7] Etherscan, *Top Accounts by ETH Balance*, Oct. 25, 2018. (`https://etherscan.io/accounts`)

[8] A. E. Gencer, S. Basu, and I. Eyal, *et al.*, "Decentralization in bitcoin and ethereum networks," *Financial Cryptography and Data Security (FC'18)*, 2018. (`https://fc18.ifca.ai/preproceedings/75.pdf`)

[9] A. Gervais, G. O. Karame, and S. Capkun, *et al.*, "Is bitcoin a decentralized currency?," *IEEE Security & Privacy*, vol. 12, no. 3, pp. 54-60, 2014.

[10] I. Grishchenko, M. Maffei, and C. Schneidewind, "A semantic framework for the security analysis of ethereum smart contracts," in *Principles of Security and Trust*, pp. 243–269, 2018.

[11] L. Guo, X. Li, and J. Gao, "Multi-party fair exchange protocol with smart contract on bitcoin," *International Journal of Network Security*, vol. 21, no. 1, pp. 71-82, 2019.

[12] Z. Li, J. Huang, D. Gao, Y. Jiang and L. Fan, "ISCP: An improved blockchain consensus protocol," *International Journal of Network Security*, vol. 21, no. 3, pp. 359-367, 2019.

[13] I. Lin and T. Liao, "A survey of blockchain security issues and challenges," *International Journal of Network Security*, vol. 19, no. 5, pp. 653-659, 2017.

[14] S. Nakamoto, *Bitcoin: A Peer-to-Peer Electronic Cash System*, 2008. (https://bitcoin.org/bitcoin.pdf)

[15] D. Puthal, N. Malik, S.P. Mohanty, E. Kougianos, and C. Yang, "The blockchain as a decentralized security frame-work," *IEEE Consumer Electronics Magazine*, vol. 2, no. 2, pp. 18-21, 2018.

[16] M. Risius, K. Spohrer, "A blockchain research framework," *Business & Information Systems Engineering*, vol. 59, no. 6, pp. 385–409, 2017.

[17] B. Rodrigues, T. Bocek, A. Lareida, D. Hausheer, S. Rafatil, and B. Stiller, "A blockchain-based architecture for collaborative DDoS mitigation with smart contracts," in *Security of Networks and Services in an All-Connected World*, pp. 16–29, 2017.

[18] D. Ron and A. Shamir, "Quantitative analysis of the full bitcoin transaction graph," in *International Conference on Financial Cryptography and Data Security*, pp. 6-24, 2013.

[19] B. S. Srinivasan and L. Lee, *Quantifying Decentralization*, 2017. (https://news.earn.com/quantifying-decentralization-e39db233c28e)

[20] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An overview of blockchain technology: Architecture, consensus, and future trends," *IEEE 6th International Congress on Big Data*, 2017. DOI: 10.1109/BigDataCongress.2017.85.

## Biography

**Ke-Ke Wu** received the Ph.D. degree from Institute of Computing Technology, Chinese Academy of Sciences in 2011. He is currently an associate professor in the Shenzhen Institute of Information Technology. His research interests include blockchain, information security, Cryptography, and side channel analysis technology. (E-mail: kk.wu@sziit.edu.cn)

**Bo Peng** received the Ph.D. degree from South China University of Technology in 2001. He is currently an advanced engineer in the Shenzhen Silico Design Technology Co., LTD. His research interests include blockchain, information security, Cryptography, and Internet of Things. (Email: pengbo@silico.design)

**Hua Xie** received the bachelor degree from Tsing hua University in 2004. He is currently an advanced engineer in the Shenzhen Silico Design Technology Co., LTD. His research interests include blockchain, information security, Cryptography, and Internet of Things. (Email: xiehua@silico.design)

**Shao-bin Zhan** received the Ph.D. degree from Jilin University in 2007. He is currently an associate professor in the Shenzhen Institute of Information Technology. His research interests include blockchain, information security, and Cryptography. (E-mail: zhansb@sziit.edu.cn)