Trend of Centralization in Bitcoin's Distributed Network

Alireza Beikverdi, JooSeok Song
Department of Computer Science,
Yonsei University, Seoul, Republic of Korea
Email: {abeikverdi, jssong}@emerald.yonsei.ac.kr

***Abstract*—Bitcoin, a distributed, peer to peer crypto currency, has gained a significant popularity among different users around the world by promising users a fully decentralized network with an inherently independence from governments and without the influence of any central authorities and organizations. Mining is the fundamental concept in Bitcoin which must be done in checking all monetary transactions and verifying them which in return generates Bitcoin as a reward to encourage this work. While the qualitative nature of this unique system is clearly accepted and understood, there are some issues regarding to its decentralized network environment. Based on our analysis due to the high variance of solo mining, the number of users joining top most famous Bitcoin mining pools are increasing due to the fact that users together under a Bitcoin pool will have a higher chance of generating next block in the Bitcoin's blockchain by reducing the variance and earning the mining reward. Furthermore, emerging huge mining farms with strong mining resources and fast processing power is another trend toward centralization. Although some might argue that, the protocol itself is purely decentralized and these are market-based centralization, this trend clearly illustrates that the pure, decentralized protocol of Bitcoin is going toward centralization in its distributed network, where any kind of centralization should be considered carefully due to the 51% attack. By analysing all the created blocks from 2009 to 2014 we proposed a centralization factor which shows how centralized is the state of Bitcoin's network in different years. Centralization, due its simplicity, is a phenomenon that happens to any disciplined and organized system automatically, which in case of Bitcoin is against the pure initial decentralized nature of it and might arise some concerns and threats to the Bitcoin's unforeseeable future.**

***Index Terms*—Bitcoin, Decentralized network, Centralization, Bitcoin pool, Mining**

I. Introduction

Recent improvements in technology and computer industry such as emergence of web, online shopping and online banking consequently, has created some alternatives for traditional money exchange. As reports in first half of 2014 in [1] illustrates, number of users using online payment services and mobile banking have increased and gained a lot of popularity. However, having a centralized trusted counter party that issues and stores these transactions such as banks and services like PayPal has always been controversial. People need to trust these third parties and since they are built as a centralized system, any breach in the network can result in information loss. There has been many reports about stolen credentials and credit cards information from the banks and other services which concern people about their privacy and security. Additionally, although we are living in a digital world, our monetary system is still analog and it is based on fiat currencies such as US Dollar, Euro, Chinese Yuan, Korean Won and etc. which are all backed by different governments and are all inflationary.

In any system no matter how trusted the counter party is, existing an element of trust or dependency to an individual or any organization or even governments as a central authority can cause inevitable issues. Centralization not only centralizes the trust factor but it also centralizes the information and processes which in that case any breaches in security of the centralized database would cause huge information leak.

By having internet as a great platform where everyone has an access to, distributing information has become significantly simpler. Therefore many distributed, decentralized technologies can easily benefit from internet as a platform. An example of it is peer to peer networks such as torrents where users can share their files and information with each other freely. As a result many others started implementing similar services in a distributed, decentralized network.

In January of 2009, Satoshi Nakamoto released the first widely used implementation of peer-to-peer distributed trustless electronic cash, replacing the central servers signature with a consensus mechanism based on proof of work protocol, with economic incentives to act cooperatively [2]. Introducing Bitcoin as the first decentralized peer to peer crypto currency [3], received a lot of attentions and brought many new opportunities to the online payment system. It introduces some unique features which has never done before. It is claimed to be a decentralized network which means there is no central authority or any organizations, controlling over the network. There is a ledger file on each node consisting all the transactions ever made in the network which introduces some transparency and accountability to all the transactions. New transactions are synchronized based on a consensus protocol called *proof of work* [3]. Furthermore it provides some levels of anonymity over transactions which can be perceived as a useful and at the same time harmful feature which is explained an analysed in [4] [5] in details. Its low transactions fees as well as universality which enables anyone with an access to the internet to send or receive money anywhere instantly, can be a great infrastructure for many developing and small countries.

Apart from the beautiful, fair nature of decentralized payment systems, there are many technical challenges that all distributed payment systems must face. While these problems

978-1-4799-8676-7/15/$31.00 copyright 2015 IEEE
SNPD 2015, June 1-3 2015, Takamatsu, Japan

are varied, ripple lab in their white paper [6] grouped them into three main categories: correctness, agreement, and utility. Correctness is an important issue and any payment system should be able to discern the difference between a correct and fraudulent transaction. Agreement refers to an agreement among all the nodes in decentralized network over one single truth and utility is defined generally as the user-friendliness and usefulness of the service.

In order to provide correctness attribute with security over transactions in Bitcoin, unlike any other payment systems that are based on fiat currencies, users themselves can contribute to the security of the system by verifying the transactions. Since there is no central authority to secure and verify transactions in Bitcoin's network, Satoshi Nakamoto in [3] introduces a concept called *mining*.

Bitcoin mining in overall exists to first determine initial distribution and generation of coins and secondly to synchronize all the transactions.

In this paper we show that Bitcoin, especially in its core engine ,mining, is getting centralized which can be considered as a threat and in order to make it viable as a future payment protocol, this issue should be resolved.

This paper is organized as follows. In section II, we provide a brief description of the concept of mining and difficulty in Bitcoin. Section III introduces new trends in Bitcoin which are leading to centralization. Section IV is the analysis over the whole blockchain since 2009 and proves centralization trend in Bitcoin's mining and define a centralization factor on yearly basis. Finally, we conclude the paper in section V.

II. Bitcoin Mining

A. Mining

As we mentioned earlier in this paper, mining is the core engine of the network which verifies and synchronizes all the transactions among all the nodes in the peer to peer distributed network. Each user can decide to become a Bitcoin miner. Miner's task is to verify new transactions. Thereafter network uses a consensus-based agreement over the verified transactions and keep the verified, agreed transactions in a block roughly every ten minutes and attach the block to the previous blocks and call the whole chains of blocks a blockchain. This agreement among the nodes is achieved through the proof of work process using hashcash function. Hashcash function was initially proposed by Adam Back [7] which is integrated into the proof of work concept in Bitcoin. Bitcoin's hash-based proof of work, works in a way that miners should find a *nonce* value that results in a value below a given target (calculated based on difficulty) when hashed with additional fields in equation 1. Header *h* is calculated by summation of version *v*, previous block hash *pb*, root hash of *r*, time *t*, bits *b* which is the target and random *nonce*.

$$h = v + pb + r + t + b + nonce \tag{1}$$

Then it calculates the final hash based on equation 2. If such nonce in this brute force is found, miner proves his work and adds his proposed block to the blockchain. This protocol is call *proof of work*.

$$FinalHash = hash(SHA256, hash(SHA256, h)) \tag{2}$$

SHA-256 is used as the hash function and it is applied twice to the input as an extra security layer.

In this scheme based on [7] there is a target value *T* which is adjusted and calculated periodically by the network. Miners goal is to find a nonce value where:

$$FinalHash < T \tag{3}$$

Once equation 3 is satisfied, the block is claimed and added to the blockchain and there is a reward as proposed in the original paper in [3] for the block finder as an incentive for users to do this process. Thus this process of finding suitable nonce is called *Bitcoin mining*.

Therefore finding this nonce creates a competition among miners in the Bitcoin's network. Basically anyone who can generate hashcodes faster, would have a higher chance of finding the small hash and earning the reward.

B. Difficulty

The Bitcoin's hashcode generation is controlled and adjusted by the network. However, the target value *T* depends on the current number and speed of miners in the network, and is quoted in terms of the difficulty *D* [8]. *D* can be calculated based on this equation:

$$D = \frac{T_{max}}{T} \tag{4}$$

And T_{max} is calculated based on equation 5

$$T_{max} = (2^{16} - 1)2^{208} \simeq 2^{224} \tag{5}$$

SHA-256 generates a value between 0 to $2^{256} - 1$. Thus for any nonce value the probability P of finding the small hash which satisfies equation 3 is:

$$P = \frac{T}{2^{256}} \simeq \frac{1}{D2^{32}} \tag{6}$$

Difficulty *D* is recalculated every 2016 blocks and adjusted based on the networks behaviour. As described in [8] difficulty has almost been increasing ever since which means the speed of hash generation has increased which indicates the mining resource power has always been growing. The real driving force behind that, is the high value and demand for each Bitcoin as of this day, there are many users who are willing to do mining because of its financial incentives.

The difficulty and reward of mining is adjusted to be at an equilibrium, otherwise miners do not find enough incentives for this process.

III. Centralization Process

Starting by 2013 where the value of each Bitcoin has risen, we observed many different trends in Bitcoin that will be discussed in details in this section.

Two of the most significant trends are: popularity of Bitcoin cloud wallets and also centralized mining. Both elements will lead to centralization and can be observed in different parts of Bitcoin's network.

A. Bitcoin cloud wallets

As it was mentioned earlier utility is one of the key factors in any crypto currency. It is clear that using Bitcoin's protocol is not a simple task for non-technical users to learn. The lack of user friendliness and the difficulty of using the core blockchain technology and the size of blockchain are significantly limiting users from using the core technology. Thus centralized organizations and services such as MtGox arose. These services host user's private keys and simplify the process with nice user interface.

The most significant issue with these cloud services normally is that, these services take the responsibility of Bitcoins private keys and any security breach in them would cause a disruptive damage to the users which is exactly the same as any other centralized services. These services consider this as a trade-off between security and usability. MtGox failure is a clear instance of this issue in which almost half a billion worth Bitcoin vanished and eventually MtGox announced and filed for bankruptcy protection in Japan and the US [9] [10]. Lessons are not learnt from this incident yet and we can still see similar services around and users who might be the future victims of these services.

Another example can be Coinbase; one of the most successful Bitcoin wallet and payment systems. Considering the structure of Coinbase, there is nothing decentralized about it. It resembles PayPal and like any other traditional banking system, they have control over your information and can spy and invade your privacy. They get all of their clients information and check what their users do with their Bitcoin and possibly block a client if he/she doesn't follow certain regulations.

Fortunately with the technology of multi-signature where user can have an address that is associated with more than one ECDSA private key, to some extent, issue of centralized cloud wallets are solved [11]. The simplest type is an m-of-n address. It is associated with n private keys, and sending Bitcoins from this address requires signatures from at least m keys [12]. A multi-signature transaction is one that sends funds from a multi-signature address. In these cloud services normally usage of 2-of-3 keys are popular where user holds two keys and service owns one key. In this case the service provider can not issue any transactions without the client's consent. Unfortunately using multi-signature is still not really popular and only few service providers provide these types of services as of this date.

B. Centralized mining

Mining is considered as the core engine of Bitcoin and it's undoubtedly the most important component in the system. According to the constant increase in difficulty for mining as discussed in the previous chapter, mining is getting more and more competing. Due to the really high difficulty value, currently a single individual has a very low chance of getting any reward from Bitcoin's mining, hence there has been some interesting trends such as increasing the popularity in mining pools and mining server farms which will be discussed as follows:

1) Mining pool: Mining pool's idea is to combine each individuals computing power to increase the probability of earning rewards by mining process. As more and more miners competing for the limited supply of blocks, individuals found that they were working for months without finding a block and receiving any reward for their mining efforts. This made mining something of a gamble. To address the variance in their income, miners started organizing themselves into pools so that they could share rewards more evenly [13].

A miner with hashrate h mining for a period of time t, will calculate a total of ht hashes and based on the difficulty formula that was explained in previous section, the number of blocks that user will find in average is:

$$ht/2^{32} * D \tag{7}$$

His expected payout is thus:

$$htB/2^{32} * D \tag{8}$$

Where B is the reward for mining.

As explained in [14] block finding with a constant hashrate h is a Poisson distribution. Since mean and variance are equal in Poisson distribution the variance of solo mining is:

$$Lamda = \frac{ht}{2^{32} * D} \tag{9}$$

However, if this operation is done in a mining pool, the variance will be much lower although the payout is still the same.

Due to this fact solo miners join mining pools to reduce the variance of mining to make it more practical as time goes by and more individuals get into the Bitcoin's network. In next section we analyse mining pools and centralization in their mining in details.

2) Mining server farm: There is another important factor in mining apart from the server pools, which is emerging huge server farms and facilities. Recently due to the value increase of Bitcoin in the market, hashing process is getting more competitive and faster from the hardware perspective. By using ASIC hardware, 1TH/s is something normal and cost effective as of this day. There is also another concern about some governments or some organizations and huge companies with the right resources and money with different reasons which have the power to establish server farms to start obtaining more resource shares in the network which leads to a centralized mining eventually. There is no proper solution to this issue with proof of work protocol yet.

In overall, due to inefficiency and lack of incentives with high variance in mining for single individuals, there has been a decline in the number of individual nodes who take part in mining and it has been replacing by giant mining farms facilities as well as mining pools. This process appears to be a trend in Bitcoin's network which is going toward a centralized point where mining is out of normal people's power in the network.

IV. Analysis

Centralization process can be seen in different components of Bitcoin's network. However, centralization in mining, as the core of Bitcoin's network, should be taken into consideration more precisely.

In this section we analyse mining and mining pools specifically and define a mining centralization factor which illustrates how centralized the mining is in Bitcoin's network.

Mining pools as discussed earlier, reduce the variance of mining and it's a natural process. In order to address each individual miner's portion of the block reward, a *share* is awarded by the mining pool to the miners who present a valid proof of work of the same type as the proof of work that is used for creating blocks, but of lesser difficulty, so that it requires less time on average to generate. Thus miners receive a reward based on their shares in the block creation. Depending on the mining pools protocol, the reward distribution varies.

Although organizing these pools would help to involve single miners, with recent trends it has become more of a business which takes charge and controls the mining process. With the original mining protocol of *getwork*, mining pool organizers issue a block header for miners to solve. In this approach miners are kept in dark without having any influences over the transaction verification process and eventually miners are solely used for hash generation of the proof of work protocol.

As discussed in [13] some people proposed *getblocktemplate* to address this issue. With this protocol miners are able to create their own block by verifying transactions so they can freely choose what they participate in mining.

There are different mining pools available currently. According to Bitcoin wiki [13] there are eleven types of mining pools such as PPLNS (Pay Per Last N Shares) used in GHash.IO, PPLNSG (Pay Per Last N Groups or Shifts) used in BTC Guild, PPS (Pay Per Share), Prop (Proportional), SMPPS (Shared Maximum Pay Per Share) CPPSRB (Capped Pay Per Share with Recent Back Pay) used by Eligius based on Luke-Jr's approach.

A general differentiation among these mining pools would be that some pools share the reward once it's generated so this means if pool doesn't get any reward, all the miners would get nothing whereas some mining pools guarantee miners a regular payouts regardless of their Bitcoin's reward. Thus pool organizers are taking the risk rather than individuals and they get more shares since they are taking this risk. Also there are hybrid versions of both. Generally, PPS (Pay Per Share) results in the least possible variance for miners while transferring all risk to the pool operator.

As a result, it can be seen that most of the mining pools are in charge of reward distribution. Some of these mining pools keep the transaction fees while some others share that among users. Additionally depending on the pools, they have specific requirements on the minimum speed of hash generation. It is also very important to take into consideration that these mining pools mostly do not allow miners to cash out their reward once they get any shares. Miners are required to mine and receive a specific amount of Bitcoin in order to get their reward which limits short-term miners. There were reports of some attacks to different mining pools which in one case *50BTC* minig pool was not able to pay the miners payout on time [15].

Gervais.et.al in [16] describes pools as a good phenomenon in Bitcoin since still to some extent its single individuals that are mining, however, having mining pool as an entity where centralizes users in a decentralized network is a threat.

As explained earlier, a clear ambiguity can be seen in most of these mining pools processes. Joining a mining pool requires some levels of trust for each individuals, in the network that is claimed to be purely trust less. Additionally some of these mining pools offer services such as cloud mining. Cloud mining refers to services in which anyone can buy or rent a mining hardware from their server facility and these cloud pools pay them according to their contracts. Hence in this case individual miners who rent or buy these hardware are not aware of the hashing processes and it is all under pools control.

We used the API provided by BlockTrail [17] and by implementing a JavaScript program, created a dataset which is shared in public at [18]. The dataset consists of more than 6 million data over 12 attributes including the miners name (mining pool) for each block in Bitcoins blockchain from the beginning of 2009 starting with genesis block to 22nd of October 2014. We have done a yearly analysis over these raw data and identified top 39 mining pools that had an effect in the mining process of Bitcoin and their frequencies over this period to analyse centralization in mining.

According to our dataset mining in Bitcoin has started in 3rd of January 2009 and it continued for 2 years almost quietly. It was purely decentralized and anyone could join mining with a normal computers processing power during those early days. The earliest mining pools activities can be seen during the beginning of 2011 by Deepbit. Following by Deepbit six other mining pools emerged during 2011. Mining became quite popular by mining pools during this year and almost 30 percent of blocks were added by these seven mining pools in 2011. This 30 percent mining pools dominance is a sign of centralization but since there is a different frequency distribution among these mining pools out of 30 percent miners, we used uniformity formula and calculated a ratio of uniformity which infers centralization.

$$\frac{\sqrt{\frac{\sum_{i=1}^{N}(x-\mu)^2}{N}}}{\mu} \tag{10}$$

Centralization factor is calculated and shown in figure 1 in each year. First two years of 2009 and 2010 are ignored in this figure since Bitcoin's network was in its early stages and there were no mining pools or any centralized mining competition[1].

[1] In this paper we assume all other unknown miners are all single or small groups of individuals.

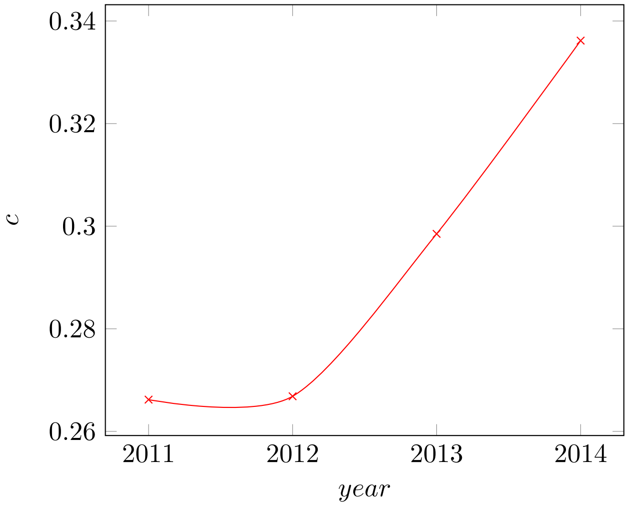

Fig. 1. Shows trend of centralization in Bitcoin Mining. X-axis shows different years and Y-axis indicates centralization factor c. 0 shows an absolute decentralized system and 1 shows an absolute centralized system.

Consequently it can be seen that during 2012, centralization trend is almost constant. There was gradual slight increase in 2013 and 2014 where by October 2014 centralization factor is at its peak by 33 percent.

As Satoshi's paper [3] claims, as long as half of miners in the system are honest miners, the system is secure therefore, from the first day, there has always been this attacks possibility called as 51 percent attack. This attack can partially shut down the whole network by manipulating all the transactions which makes a decentralized mining really important.

Centralized mining is considered as a big threat. As an example GHash.io as the biggest mining pool this year, has passed 50% of block creation in the blockchian as explained in [19] and some people DDoS this service to respond this issue and show their concerns. Cex.io which the first cloudhashing service to be adopted by the masses works alongside with GHash.io and although these mining pools and services claim that they do not have any intentions on 51% attack, it is still an unanswered question.

Fortunately recently new mining pools have emerged such as the protocol that mentioned earlier which uses getblocktemplate and also peer to peer mining pools (P2Pool) which offer a decentralized peer to peer mining. P2Pool uses a concept called *sharechain*. Its uses the same technology as blockchain with a lower difficulty that is generated every 30 seconds. Each miner can get shares for this process which will eventually get rewarded distributively based on their shares. Other miners can also see the miners suggested block on each share generation on the sharechain. These protocols still have many inefficiencies but are still considered as better options toward decentralization. Additionally there is another significant fact that should be taken into consideration for any future inferences on mining. Bitcoin's mining reward decreases over time and by the year 2140 the mining reward will be zero [13]. As discussed at [20] transaction fees are going to replace the reward fee in future to keep the incentives for miners. In [20] they illustrated the best-case adoption in future.

In near future the reward value is going to be halved which obviously would have a big influence on the mining process and would obviously affect centralization factor. The future mining with the current scheme is very unpredictable due to the fluctuations in the value of Bitcoin but according to the current trend based on the last three years, centralization in mining is predicted to be increased in 2015.

V. Conclusion

Decentralized peer to peer network requires decentralized verification and adjustment. Mining process as the engine of Bitcoin, which verifies and secures all the transactions is a really important component in the system.

Based on our work, recent trends especially in mining illustrates a centralization trend which is against the nature of Bitcoin as a decentralized network.

We introduce a centralization factor in Bitcoin's mining which shows the state of centralization in the network. Centralization factor zero indicates absolute decentralization in mining while centralization factor 1, shows absolute centralization same as banks and PayPal payment system. Based on our mining pool dataset, it is illustrated that mining in Bitcoin has an increasing centralization factor which reaches 0.33 in 2014.

As shown in this paper, in order to keep Bitcoin as a distributed and decentralized network, mining process as the core engine of mining where all the transaction verification happens should be clear to users. Moreover, any users with current normal processing power should be able to contribute in the mining process. Centralization phenomenon is something that happens to any disciplined system by nature to makes things simpler which in case of Bitcoin, centralization is considered as a big concern because of the 51 percent attack and it can destroy the beautiful nature of Bitcoin as a decentralized payment system which is one of its important properties.

Ultimately, this centralization can be considered as a market-based centralization and its not coercive. However, it is very important for anyone in the network (especially miners) to be aware of this and try to preserve and improve this beautiful technology accordingly.

As for future work, we will find and propose methods and solutions to the centralization and improvements in the mining protocol.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2014R1A1B3004161).

References

[1] Global Online Payment Methods: First half 2014, ReportLinker, [Online; accessed 12-October-2014].

[2] A. Back, M. Corallo, L. Dashjr, M. Friedenbach, G. Maxwell, A. Miller, A. Poelstra, J. Timn, and P. Wuille, “Enabling Blockchain Innovations with Pegged Sidechains,” pp. 125, 2014.

[3] S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system,” *Consulted*, vol. 1, no. 2012, p. 28, 2008.

[4] M. Mser, “Anonymity of Bitcoin Transactions An Analysis of Mixing Services,” pp. 1718, 2013.

[5] F. Reid and M. Harrigan, “An Analysis of Anonymity in the Bitcoin System,” 2011 IEEE Third Int’l Conf. Privacy, Secur. Risk Trust 2011 IEEE Third Int’l Conf. Soc. Comput., pp. 13181326, Oct. 2011.

[6] D. Schwartz, N. Youngs, and A. Britto, “The Ripple Protocol Consensus Algorithm,” 2014.

[7] A. Back, “Hashcash - A Denial of Service Counter-Measure,” no. August, pp. 110, 2002.

[8] K. J. O. Dwyert and D. Malone, “Bitcoin Mining and its Energy Footprint,” 2014.

[9] Y. Takemoto and S. Knight, “Mt. Gox files for bankruptcy, hit with lawsuit,” Reuters, 28-Feb-2014.

[10] E. Warnock, T. Mochizuki, and A. Martin, “Mt. Gox files for bankruptcy protection,” The Wall Street Journal, Feb-2014.

[11] V. Buterin, “Multisig: A Revolution Incomplete,” Bitcoin Magazine, Jul-2014.

[12] M. Rosenfeld, “Multi-signature addresses,” Bitcoin StackExchange, 2014. [Online]. Available: http://bitcoin.stackexchange.com/questions/3718/what-are-multi-signature-transactions.

[13] “Bitcoin wiki,” 2014, [Online; accessed 25-November-2014]. [Online]. Available:https://en.bitcoin.it/wiki/MainPage

[14] M. Rosenfeld, “Analysis of hashrate-based double-spending,” pp. 114, 2012.

[15] “Transactions failed due to 11th February attack are refunded,” 50BTC, 2014. [Online]. Available: https://50btc.com/en/news/view/161.

[16] A. Gervais, G. O.Karame, V. Capkun, and S. Capkun, “Is Bitcoin a Decentralized Currency ?,” no. Zurich, ETH, 2014.

[17] “Block Trail,” 2014. [Online; accessed 23-October-2014][Online]. Available: https://www.blocktrail.com/api.

[18] “Mining pool dataset over Bitcoin Blockchain,” 2014. [Online]. Available: https://www.dropbox.com/s/ctgwlr3whbwmalt/Dataset.csv.

[19] A. Quentson, “Bitcoin Mining Pool Ghash.io DDos-ed in Response to threat of 51% attack?,” Cryptocoins news, 2014.

[20] K. Kaskaloglu, “Near Zero Bitcoin Transaction Fees Cannot Last Forever,” pp. 9199, 2014.