Blockchain Mining

as a

Stochastic Process

December 7th, 2018

Benjamin Craver

bhcraver@ncsu.edu

Abstract

Cryptocurrency has surged in popularity the past year due to meteoric rises in prices. Turning some computer nerds into multimillionaires in the span of months. Now that prices are coming back down to Earth, people are realizing that the underlying technologies of cryptocurrency is where the true value lies. This technology is commonly referred to as *blockchain* but what does this word mean and how does it work? This paper uses stochastic models to describe blockchain mining and its properties. More specifically the incentive structure and its potential traps.

Table of Contents

Introduction 3

Origin of Blockchain 3

How Mining Works 4

Methodology 4

Geometric Distribution Convergence to Exponential Distribution 4

Alternative History Attacks Modeled as Gamblers Ruin Problem 5

Analysis 6

Mining as a Stochastic Process 6

Expected Time to Mine a Block 6

Block Difficulty 7

Expected Reward 8

Why Miners Join Pools 8

Mining Pools 9

Gamblers Ruin 10

Conclusion 12

References 13

Introduction

Origin of Blockchain

Bitcoin is credited with popularizing blockchain technologies. However, the genesis of blockchain technology actually dates back two decades prior to Bitcoin's release. The first implementation of a blockchain-like system was proposed by Haber & Stornetta [1] as a means of time-stamping documents. They This allowed them to cryptographically verify a document was valid at a specific time in the past. The use of blockchain technology for a means of storing value was first proposed in 2002 by Back [2]. The proposed *Hashcash* system was used to 'charge' an agent a small amount for sending an email in an attempt to stop email spam.

Words like blockchain and mining get overused and become buzzwords. Therefore, it is appropriate to rigorously define these words. The word blockchain was never used in the original Bitcoin whitepaper, only "a chain of blocks". The term wasn't used until a year after the whitepapers release in a forum post between Satoshi Nakamoto and Hal Finney [3], who is one of the potential candidates of Nakamoto's true identity.

The use of the word blockchain is due to blocks being 'chained' together by including the hash of the previous block in the current block. It allows users to verify the chain of blocks is valid without storing every single block. The calculation for verifying a block is not computationally expensive while mining a block requires large amounts of electricity and computational power. This creates a robust system that is easy to verify but hard to fake.

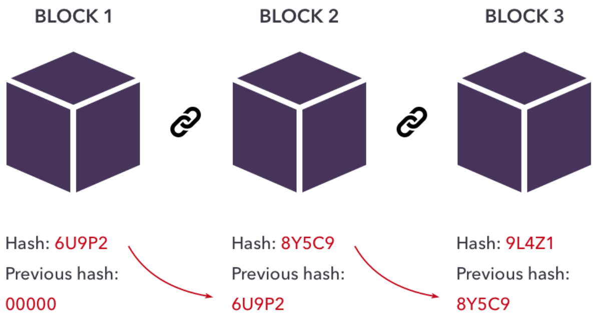

Figure 1: How Blocks are Linked Using Hashes

According to the Webster's dictionary a blockchain is "a digital database containing information (such as records of financial transactions) that can be simultaneously used and shared within a large decentralized, publicly accessible network". Another definition by bitcoin.org states "A chain of blocks with each block referencing the block that preceded it. The most-difficult-to-recreate chain is the best block chain". Both definitions leave out the most important aspects of a blockchain such as the algorithms used to verify blocks and the immutability of blocks.

How Mining Works

Another aspect of blockchain that is often misunderstood is the mining process. Mining refers to the process of continuously calculating hashes of blocks until the hashed value is below some threshold set by the network. By using computation power as a limited resource, mining creates a protection against malicious actors in the network. The first agent to mine the block is given a reward for their efforts. This currently amounts to 12.5 Bitcoins plus the transactions fees for all transactions included in the block.

The most fundamental technology behind blockchain is the hash algorithm. The hashing algorithm used in Bitcoin is the SHA-256 algorithm developed by the National Security Agency in 2001. The data to be contained in a block is first concatenated. This includes a random value called the nonce used to change the value of the hash among other data that can be used to alter the hash. Once a correct or winning nonce is discovered, the block and the hash is published to the network. All the network nodes quickly verify its validity and work on the next block begins.

Methodology

Geometric Distribution Convergence to Exponential Distribution

Guessing the correct nonce to produce a block hash below the threshold is describe by a geometric distribution. The nonce has 2^{32} or 4,294,967,296 possible values due to the 32 bits used to store the nonce in the block. This number grows even higher when the target hash of a block is below a threshold and other information in the block must be altered. The number of nonce for a single block that result in a hash lower than the target are unknown unless calculated by brute force. As discussed later, this would take millions of years on a standard computer. Therefore, it is not ideal to model nonce generation as a geometric distribution. Taking the limit as the number of possible nonce tends toward infinity results in an exponential distribution.

$$1 = \sum_{k=0}^{\infty} \mathbb{P}(X = k) = \sum_{k=0}^{\infty} \lambda \left(\left(1 - \frac{\lambda}{n}\right)^{n \cdot k/n} \frac{1}{n} \right) \overset{n\to\infty}{\to} \int_0^{\infty} \lambda \mathrm{e}^{-\lambda x} \mathrm{d}x$$

Figure 2: Convergence of Geometric to Exponential

Alternative History Attacks Modeled as Gamblers Ruin Problem

The blockchain protocol is meant to protect against the *double spend* problem. *Double Spend* is when an agent uses a resource more than they have available. In traditional finance a middle-man such as a bank or credit card company verify and clear transactions. In blockchain this verification is performed using math and the expenditure of resources such as compute power and electricity.

The chain with the most blocks is considered to be the correct chain. *Alternative History Attacks* try to exploit this rule. An *Alternative History Attack* is when a malicious actor pays in Bitcoins to receive some good. The seller, or receiver of the Bitcoins, then waits for a specified number of blocks to be mined before sending the product. This ensures the Bitcoin transaction is final and practically irreversible. However, if the attacker (the one who sent the Bitcoins) controls a significant portion of the Bitcoin mining networks computing power, they can secretly work on a longer chain and publish it once the unknowing client has sent the product. The trick is, in the blocks published by the attacker, the transaction where the Bitcoins were sent is missing.

Since mining can be modeled by the exponential distribution, the probability that an attacker will be able to mine X number of blocks in secret before the rest of the network can be modeled using a binomial random-walk and gamblers ruin. The gamblers ruin formula has three inputs and one output which is the probability of success. The three inputs are p = probability of success of the attacker, q = probability of failure by the attacker, z= the number of necessary successful attempts before the game is won, and λ = z*(q/p). This formula was included in the original Bitcoin whitepaper [4].

$$1-\sum_{k=0}^{z}\frac{\lambda^{k}e^{-\lambda}}{k!}\left(1-(q/p)^{(z-k)}\right)$$

Figure 3:Formula for Probability of a Successful Alternative History Attack

Analysis

Mining as a Stochastic Process

Imagine this, block 151 was published and verified 5 minutes ago. You have been trying to mine block 152 since block 151 was released. Your friend just booted up his new mining rig, with the same hash rate as your machine, and is starting to mine block 152. Who is more likely to mine the block? The answer, they're equally likely. What is the expected time until block 152 is mined given it has been available 5 minutes? It is 10 minutes, the same expected time as when it first started. These are both examples of the lack of memory property, a key feature of the exponential distribution. The standard deviation of the exponential distribution is equal to the mean resulting in a lot of uncertainty.

Expected Time to Mine a Block

The expected time to calculate a block is modeled by the formula

*time (λ)= difficulty * 2^32 / hashrate*

where the units of time correspond to the units of hashrate. The hashrate of an average GPU is around 25 MH/s (million hashes a second). The current difficulty is 6,653,303,141,405. With a single GPU the expected time to mine a block is over 36 million years. Buying lottery tickets is a better bet than mining Bitcoin. Therefore, miners join together in mining pools. If a miner in the pool wins a block, then the administrators take a small fee and the profits are distributed across the pool in proportion to hash rate. This results in a drastic reduction in variance turning what was once a lottery style gamble to a predictable and steady stream of income.

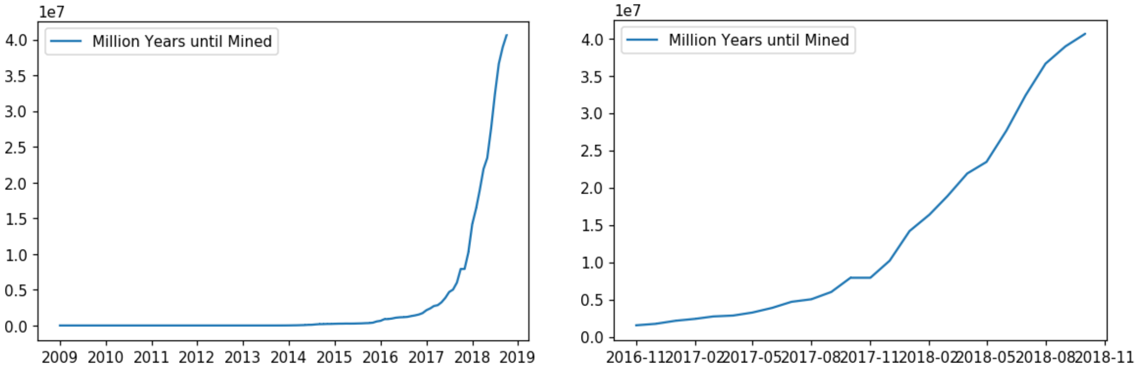

Figure 4: Expected Time Until Block is Mined Using Single GPU

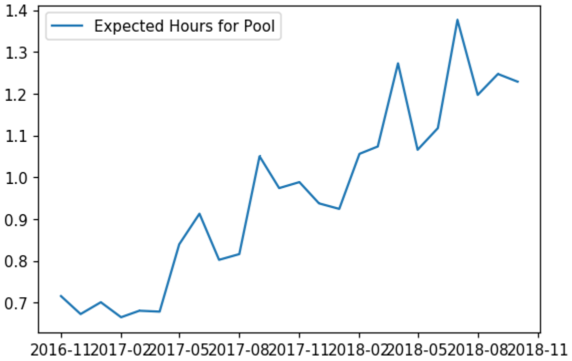

Figure 5:Expected Time Until Block is Mined by Pool with ~20% of Network Hash Power

Block Difficulty

The probability of successfully mining a block changes every 2016 blocks or roughly two weeks because there are 20,160 minutes in a week.

*new_difficulty = current_difficulty * (20160 / time_for_last_2016_blocks)*

This is done to ensure that a block is mined about every 10 minutes even when the total computation power of the network can fluctuate. The reward for mining a block is fixed, until sometime in 2020, at 12.5 Bitcoins per block. Therefore, when the price of Bitcoin increases mining becomes more and more profitable. More people start mining Bitcoin and the total hash power of the network increases. Consequently, the time to mine blocks decrease. The network adapts to this change by increasing the difficulty of finding a correct nonce to ensure that blocks are mined on average every 10 minutes.

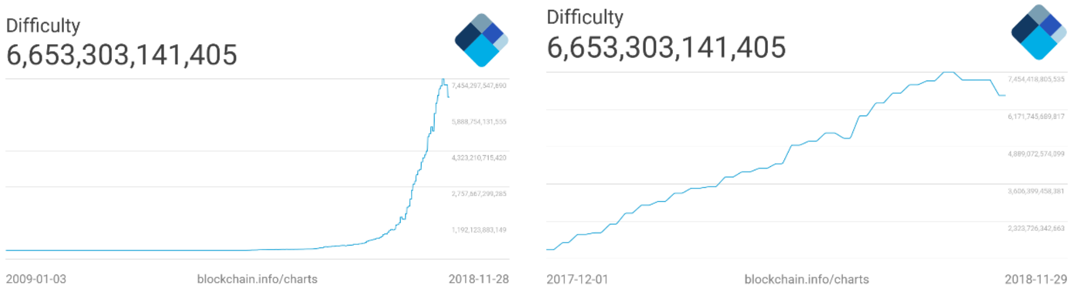

Figure 6: Block Mining Difficulty Increase [5]

Expected Reward

The reward for finding a winning nonce is currently 12.5 Bitcoins plus the transaction fees of all the transactions included in the block. This value is currently fluctuating around 4,000 USD per Bitcoin which equates to $50,000 per block mined. Transaction fees are paid by users to have their transaction included into a block. The higher the fee the higher the probability your transaction will be included in the next block. The miner includes an address in the block for the Bitcoin reward to be sent to. This address is usually the miners address. However, in the case of pool mining, the address is that of the mining pool administrators.

Why Miners Join Pools

To reduce the variance of expected payoff, miners join together in mining pools. If a miner in a pool finds a correct nonce, the block is published, and the Bitcoins are sent to the pool's administrator wallet. The reward is then distributed to the miners in the pool according to hash power. To calculate a miner's hash power, miners constantly send blocks that are greater than the network wide target but less than some specified value.

For a miner with a single GPU, joining a pool with ~18% of total network hash power reduces the variance of payment arrival from 1,296 Trillion years^2 to 1.93 hours^2.

Mining Pool Dynamics

Mining pools create a problem because they can make up large portions of a networks overall power. Currently the largest two pools account for at least 18% and 12% percent of all blocks mined. This can cause problems because the network was designed to be decentralized to protect against double-spend attacks.

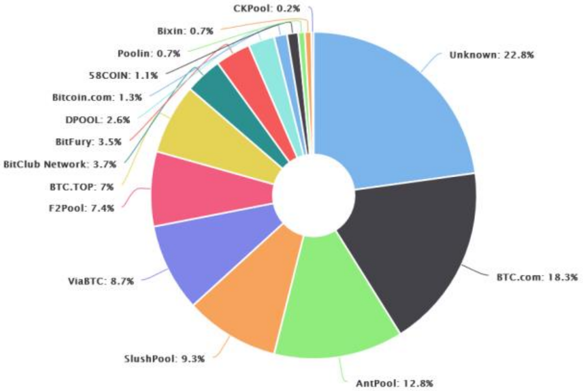

Figure 7: Pool Hash Power Distribution as of Dec 1st, 2018 [6]

The exact hash rate of a mining pool isn't available, but we can estimate it by calculating the total hash rate of the total network and then multiplying this value by the proportion of blocks mined by a pool.

*Total hash rate * percent of blocks mined by pool*

Currently the total estimated hash rate of the entire network is around 35 million-trillion hashes per second or 35,000,000,000,000,000,000. The largest mining pool controls 18.3% of the network or roughly standard 256,200,000,000 standard GPUs. Now they aren't running that many GPUs across the world. There are technologies that are more efficient such as Application Specific Integrate Circuits also known as ASIC miners.

Gamblers Ruin

It is in the interest of miners to join the largest pool because this results in the largest variance reduction of their pay. However, Bitcoin was designed with the principles of decentralization in mind to protect against malicious actors. If a pool becomes too large, the administrators could attempt an alternative history attack with an unnerving high probability of success. Below is a graph showing the probability of an alternative history attack after 6 blocks by the pool with the highest proportion of the network.

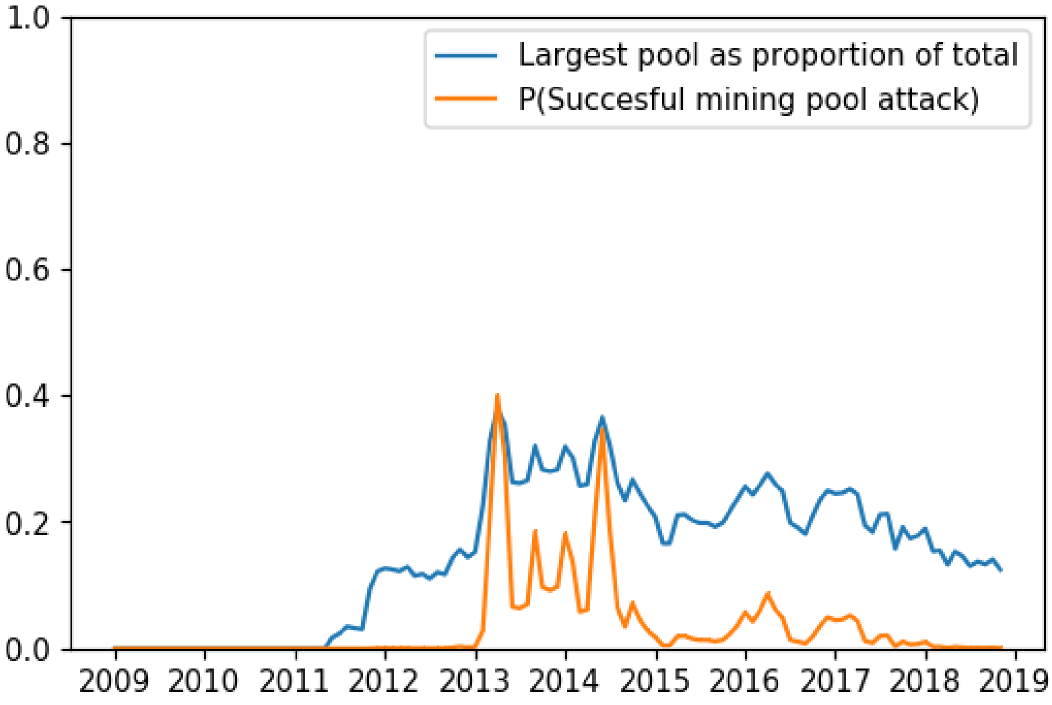

Figure 8: Proportion of Largest Pool's Hash Power to Total Networks Hash Power Compared to Attack Success Probability

When looking at this graph it is clear there is not a linear relationship between pool proportion and the probability of a successful attack. Below is a graph of a pools proportion of total network power to the probability of a successful alternative history attack. At roughly 30% of total network power there is a knee point where the probability of a successful attack starts to exponentially increase. Therefore, it can be concluded that a pool shouldn’t be allowed to become larger than 30% of the total network. At 40% of the total network, a pool has slightly greater than 50% probability of a successful attack.

This means, neglecting transaction and mining costs, in the long run they will be more successful than not, and an attack strategy will be profitable.

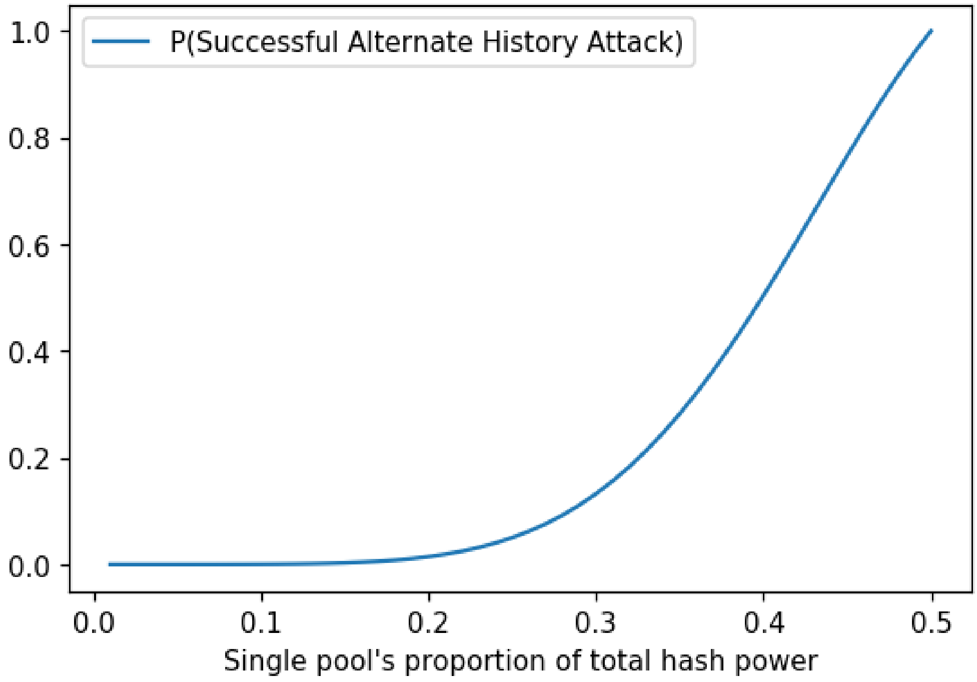

Figure 9:Attack Success Probability Compared to Pool Hash Power Proportion to Total Network Hash Power

At 50% of the total network, a pool has a 100% chance of a successful attack. At this point the malicious pool would control all of the voting power of the network. The nodes of the malicious pool could choose to verify whichever transactions they please. However, if a pool was found to control this much of the network all trading would surely cease and the price of Bitcoin would collapse. Therefore, it is in the interest of miners to self-regulate to ensure confidence in the authenticity of the blockchain record keeping protocol.

Conclusion

Cryptocurrency and Bitcoin are built on the most revolutionary technology since the inception of artificial intelligence. Cooperative networks based on blockchain protocols will continue to proliferate after the importance of Bitcoin has waned. Therefore, it is important to build models to study blockchain and stochastic models are an appropriate way to accomplish this. Interarrival times of Blockchain mining can be modeled by an exponential distribution which allows for the application of other models and formulas of stochastic modeling. Now that the Bitcoin network has become so large, the only profitable way to mine Bitcoins is through a mining pool. However, this contradicts Bitcoins founding principle of decentralization. To protect confidence in the blockchain protocol underlying Bitcoin, miners should not join a pool that commands over 20% of the networks total hash power. This ensures a pool cannot successfully perform an alternative history attack.

References

[1] S. Haber, W.S. Stornetta, "How to time-stamp a digital document," In Journal of Cryptology, vol 3, no 2, pages 99-111, 1991.

[2] A. Back, "Hashcash - a denial of service counter-measure," http://www.hashcash.org/papers/hashcash.pdf, 2002.

[3] https://satoshi.nakamotoinstitute.org/emails/cryptography/6/

[4] N. Satoshi, “Bitcoin: A Peer-to-Peer Electronic Cash,” https://bitcoin.org/bitcoin.pdf ,2008.

[5] Blockchain.com provided data for graphs

[6] https://www.blockchain.com/en/pools