Blockchain 101

An introduction to blockchain technology and an analysis of if it's worth the hype

Zack Holmberg

I. Introduction

Blockchain technology is arguably the most disruptive technology to surface in the last decade. This paper serves as an introduction to blockchain technology for those who understand nothing about blockchain technology, except for the fact that it's related to Bitcoin. More specifically, this paper will discuss a brief history of blockchain technology, modern design and implementation of a blockchain, the advantages and disadvantages of the technology and finally a brief evaluation of whether the technology is worth the hype.

II. History

On October 31, 2008, the world was changed forever when a whitepaper, titled "Bitcoin: A Peer-to-Peer Electronic Cash System" was published under the pseudonym Satoshi Nakamoto. In this paper, Nakamoto introduced the cryptocurrency known as "Bitcoin". Additionally, Nakamoto detailed the implementation of the cryptocurrency known as the Blockchain (Nakamoto, 2008). This seemingly novel technology was officially implemented on January 3, 2009 at 12:15 pm with the first block, referred to as the "Genesis Block", being mined on the Bitcoin Blockchain, according to its public record (Block 0, n.d.). The "birth" of blockchain is widely accredited to Nakamoto's original paper. However, Nakamoto's paper was not the first to discuss concepts similar to the blockchain technology, and many technologies levered by blockchain are the result of earlier research.

In 1991, 17 years before Nakamoto released their paper, Stuart Haber and W. Scott Stornetta published the article "How To Time-Stamp a Digital Document" in the Journal of Cryptology. Haber and Stornetta's article proposed using hash values of documents and saving those values alongside a timestamp (Stornetta, 1991). Then, records would be linked in a chain-like data structure, akin to a linked list, by storing hashes of previous records. Additionally, their timestamping protocol makes use of private key digital signatures to authorize submitted data (Stornetta,

1991). In fact, this is exactly how Bitcoin and blockchains in general sign transactions and build the chain, which we will discuss in further detail in the next section. Shortly thereafter, Haber and Stornetta improved their technique to enable multiple documents to be added simultaneously to a single block (Bayer, 1993).

Nevertheless, we know for sure that Nakamoto's Blockchain is the product building upon Haber and Stornetta's idea because Nakamoto cites Haber and Stornetta three times in their Bitcoin paper (Nakamoto, 2008).

III. Modern Design and Implementation

a. Definition

First and foremost, an important distinction must be made between "the Blockchain" and "a blockchain". On one hand, the Blockchain defines the specific implementation of the blockchain technology as used by Bitcoin. On the other hand, a blockchain merely defines an implementation of the general blockchain technology and is what will be discussed in this paper. Now that we have clarified that differentiation, let's take dive into understanding exactly what a blockchain is, and furthermore what blockchain technology as a whole is.

Essentially, a blockchain is merely a data structure, specifically a glorified linked list. We will formally define a blockchain as an *append-only, immutable, distributed and digital ledger composed of blocks containing data that are linked together by a digital chain*.

b. Ledger

Ever since the inception of transactions, whether it be an exchange of goods, services or currency, a transaction is (often required by law) recorded in some sort of medium, whether that be on a piece of paper, on a spreadsheet or database. In these cases, the medium on which the transaction records are contained is defined as the *ledger*. Ledgers were created in order to provide a source of truth if disputes about a transaction ever came about, which can be common due to the lack of

trust in human nature. Furthermore, this trust issue is worsened by the exchange of digital assets. Formally, this issue is referred to as the “Double-Spending Problem” (DSP). With regards to digital transactions, one can spend the same value (in the form of a digital asset) more than once because digital assets, such as an image or program, can be duplicated and copies of a digital asset can be used in different transactions (Voshmgir, 2019). This is the DSP. In contrast, physical values don’t face the same problem because they cannot be duplicated (albeit, counterfeit is possible, but it is expensive). More specifically, the parties involved in a transaction can immediately verify the physical token, whether that be currency, a commodity or a collectable (Voshmgir, 2019). The DSP has plagued the digital economy for decades, however blockchain technology offers to solve that problem. With the technology, the blockchain itself represents the ledger. However, in order to fully understand this and the rest of the definition, we will first have to understand the blockchain architecture.

c. Architecture

As mentioned earlier, a blockchain is a data structure that is composed of blocks containing data that are linked together by a digital chain. This is pretty obvious given the name blockchain. However, the structure of the chain and content of the blocks is not so obvious. In general, a blockchain is made up of blocks that can contain different forms of data, however there are a few standard components of a block that are required to achieve the goals of the technology, such as immutability.

First, a block must contain a hashed (usually using SHA-256) reference to the block that came directly before a given block. The “chain” is formed as a result of every block containing a reference to its predecessor. The first block of any blockchain is known as the “genesis block” or “Block 0” (Michael Nofer, 2017).

Second, a block must contain its own hash value and other identifying information. This self-identifying hash value (hashed using SHA-256) is usually composed of other pieces of identifying information within the block. Most often, a

hash is created by using a timestamp of when the block was created, the hash value of the previous block and a “nonce”, which is a random number generated for verifying the hash, as the hash input (Michael Nofer, 2017). More specifically, for a block’s hash to be considered valid, it must satisfy some sort of condition as defined by the blockchain. For example, let’s suppose that a block’s hash value is only valid if the first three digits of the hash are ‘0’. Then, the algorithm will take the timestamp and hash of the linked block and try to find a nonce value that, given the other two inputs, will provide a hash that satisfies the condition when all three inputs are used in the hash function. The genius part of this implementation is the immutability guarantee that it provides. Due to the fact that every hash value in the chain is computed using a hash value of another block and the data within a block, the implementation guarantees that if one single block is altered or tampered with, the hash value of that block will change (because one of its hash input values is being changed), and thus every hash value in the entire chain will change, rendering the whole chain invalid. Recalculating the hash of every block would be an impossible task with regards to computing power. This design decision provides security and the append-only nature of the data structure.

Lastly, a block must contain some sort of content or piece of valuable data, such as a transaction record. However, this component of a block is extremely customizable. More specifically, the content of a block is determined by the developers of a specific blockchain implementation. For example, blocks in the Blockchain contain data regarding a Bitcoin transaction as well as other metadata (Block, n.d.). However, the content of a block can be anything that can be digitally represented, such as any piece of media (Michael Nofer, 2017). Furthermore, in the case of Ethereum, a block can contain even a piece of executable software, but that’s a discussion for a different paper.

A blockchain lives on a distributed online system composed of nodes, where a node is merely a computer. Every node gets copy of the blockchain on the system. An example of a block in the Blockchain is depicted in Fig. 1, where TX represents a transaction record. Thus, as a byproduct of the architecture described

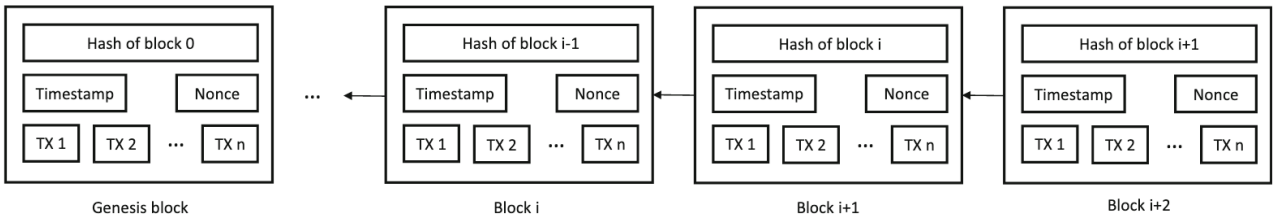

Figure 1 (Michael Nofer, 2017) – Visual representation of a blockchain

above, we can now see clearly how the design of a blockchain satisfies our definition of an *append-only, immutable, distributed and digital ledger composed of blocks containing data that are linked together by a digital chain*. However, now that you know what a blockchain looks like, you're likely wondering how blocks, and furthermore the blockchain, are created. Let's look at that next.

d. Consensus Algorithms

Blockchain construction is handled differently by various blockchain implementations. There is no standard or one-size-fits all protocol for blockchain creation, thus for this section we are going to examine the Bitcoin Blockchain's implementation, as it is the most popular and widely used. Additionally, we will discuss another popular consensus algorithm.

As Nakamoto describes in their paper, Bitcoin transactions occur on a purely distributed peer-to-peer (P2P) system (Nakamoto, 2008). More specifically, the system is a graph of nodes that are all connected to each other. Each node, which is merely a computer, contains a copy of the current state of the Blockchain. When a new Bitcoin transaction occurs, it is added to the "transaction pool", which is a list of unconfirmed transactions that need to be added to the Blockchain ledger. Every transaction requires a new block to be added to the Blockchain. The process of adding transaction records to the Blockchain is called "mining" (Mining, n.d.). Essentially, once a transaction is added to the transaction pool, miners compete in mining a block for the transaction, which boils down to a competition for who can solve a difficult math problem first. In order for a miner to be successful, they must discover a *Proof-of-Work* (PoW) for the given block (Nakamoto, 2008). A PoW is a

piece of data which is difficult, and more importantly expensive, to solve but easy for the system to verify. The PoW is a consensus algorithm, and it is the genius that actually makes Bitcoin work. Any node in the distributed P2P can attempt to mine a block, however only the first miner to do so receives a (sizeable) monetary reward for their work, which is the incentive for miners to contribute to the system (Nakamoto, 2008). The math problem that miners attempt to solve is the finding of the hash value for the new block, which they attempt to do by trying different nonce values for the hash function, as we discussed prior. Once a the PoW is achieved and the network validates the new block, then the transaction is changed to confirmed, a new block is created and the Blockchain is updated. Then, all nodes on the network get sent a new copy of the data structure.

The difficulty of the PoW is managed by the blockchain owners and is often increased based on the number of blocks in the chain. The difficulty is increased to account for Moore's Law. As Moore's Law insinuates, CPU power increases predictably over time, and thus the PoW difficulty must increase proportionately to maintain a blocks-per-minute type of goal. For example, the Bitcoin Blockchain's difficulty is adjusted every so often to achieve a target of 0.1 blocks/min, or a block being mined every ten minutes.

However, this time window introduces the possibility for a number of miners to all achieve the PoW within similar amounts of times while competing on the same block. When this occurs, the longest chain rule comes into effect (Nakamoto, 2008). As Nakamoto states in their paper,

"The PoW also solves the problem of determining representation in majority decision making. If the majority were based on one-IP-address-one-vote, it could be subverted by anyone able to allocate many IPs. Proof-of-work is essentially one-CPU-one-vote. The majority decision is represented by the longest chain, which has the greatest proof-of-work effort invested in it. If a majority of CPU power is controlled by honest nodes, the honest chain will grow the fastest and outpace any competing chains." (Nakamoto, 2008).

However, the PoW consensus algorithm has received negative feedback due to fact that, even individually, miners use a lot of resources (read: fossil fuels) to find the PoW, making it a wasteful option. To solve this problem, the Proof of Stake (PoS) consensus algorithm was developed as an eco-friendly alternative to PoW (Zibin Zheng). In the PoS system, miners are replaced by validators. Candidate validators must prove ownership of a certain amount of an asset, most often the cryptocurrency they are mining for (Zibin Zheng). The system verifies asset ownership because instead of mining, nodes put forward bids for the unchallenged access to creating a new block. What's at "stake" is an amount of the previously validated asset that a candidate validator must submit with their bid. Once all bids are in, the system will choose a validator who gets the only rights to creating the new block. With this system, only one node will spend the energy to create a new block, vastly reducing overall costs of adding a new block to the chain compared to PoW. The theory behind the PoS consensus algorithm is that validators with a greater amount of equity, or those who are more invested, would be less likely to attack or be a dishonest node in the system (Zibin Zheng). More importantly, if a validator misbehaves, they lose the stake that they committed to the bid. This is the incentive for nodes to be honest.

In summary, consensus algorithms define and fuel the way new blocks are added to a blockchain whilst simultaneously guaranteeing security and validity for the system. However, despite the genius behind the use of consensus algorithms, *technically* speaking blockchain technology still has vulnerabilities.

e. Vulnerabilities and Attacks

In this section, we will talk about the main type of blockchain attack and what vulnerabilities of a blockchain that it exploits. This attack is an umbrella term for many different attacks that all leverage the same principle. (Marco-Gisbert, 2019). This attack is called the **51% attack**.

As mentioned earlier, the use of consensus algorithms, especially PoW, makes tampering with a blockchain extremely difficult. However, difficult is not

impossible; there is one case that would allow hackers to tamper with a blockchain that uses the PoW consensus algorithm. The 51% attack's intention is to abuse the DSP. The attack occurs when an attacker is able to control *at least 51%* (a majority share) of the total hashing power of a distributed blockchain network. An attacker starts by secretly mining a separate chain which is fully isolated from the public chain. Once the dishonest chain is created and is longer than the public chain, it is pushed to the network for consensus and will be accepted as the new real chain (Marco-Gisbert, 2019). This is what enables the double-spending. Since the blockchain policy complies with the longest chain rule as previously discussed, if attackers are able to get at least 51% of the hashing power, then they will be able to drive the longest chain because they can create blocks faster than any other miner on the network. This guarantees that they will always have the longest chain. However, as also talked about previously, achieving at least 51% of the hashing power would require an *insane* amount of computing power and resources, which makes this type of attack extremely difficult to pull off.

51% attacks can also be used on blockchains that use the PoS consensus algorithm. When a node can consistently stake a large portion of wealth, they become a powerful entity within the network and develop the ability to influence the well-being of the network (Marco-Gisbert, 2019). By achieving the majority of the share of whatever asset pertains to that blockchain, a node can ensure that they consistently win the rights to add new blocks to the chain. This occurs because the PoS algorithm essentially acts as a lottery, with odds given to each node equal to the portion of the asset that a node has compared to the entire pool. Thus, to conduct a 51% attack, an attack must acquire at least a majority share of the entire asset pool. However, similar to PoW, the cost of achieving this majority can be *immense*. Therefore, the threat level of the 51% attack on PoS systems is comparable, or even considered lower, than PoW systems (Marco-Gisbert, 2019).

IV. Evaluation

a. Advantages

Blockchain technology has many advantages that puts it a class above other buzzwordy tech. I'll only list the main three. Firstly, a blockchain's distributed architecture provides a reliable system, as there is no single point of failure within the system; each node in the system contains a copy of the chain. Thus, any single node going offline will have no effect on the rest of the system. Secondly, due to the complexity of mining blocks, it is extremely difficult to alter the chain. This feature enables blockchain to be a great candidate for systems that require an auditable data trail, since the ledger is immutable and completely public. Additionally, the architecture of blockchain (virtually) solves the DSP, which is an extremely impressive feat. Finally, the idea behind blockchain technology gives it so much potential. More specifically, blockchain removes the trust factor from transactions. Currently, there are so many different trusted intermediaries, such as banks, that people utilize in order feel confident in their transactions. Blockchain removes the need for such intermediaries, and all the costs that come with them.

b. Disadvantages

Despite the hype, blockchain technology, like any technology, has its disadvantages, Firstly, as discussed in Section III Part e., blockchain consensus algorithm are still potential vulnerable to the 51% attack, however unlikely. Second, as also mentioned as an advantage, modification of a blockchain is extremely difficult. This feature of a blockchain can be an incredible pain if maintenance on the chain is desirable. Finally, as discussed in Section III Part d., blockchain implementations that us the PoW consensus algorithms are extremely inefficient and costly, primary monetarily speaking but these inefficiencies are ultimately environmentally unsustainable.

c. Final Thoughts

Even with considerations to the disadvantages of blockchain, I must ultimately state that blockchain technology is more than likely the future of distributed systems. The advantages of the technology are mostly unheard of when it comes to distributed systems, and I cannot wait to see how much blockchain technology expands into both the private and public sectors over the next decade(s).

V. Conclusion

In short, invest in Bitcoin*. In all seriousness and conclusion, this paper served as an introduction to blockchain technology for those who previously understood relatively nothing about blockchain. More specifically, this paper discussed a brief history of blockchain technology, modern design and implementation of blockchain, the advantages and disadvantages of the technology and finally a brief evaluation of whether the technology is worth the hype.

Hopefully by now you have a relatively complete understanding of blockchain technology and are inspired to do some further research on your own.

*I am a Computer Scientist, not a financial advisor.

VI. Bibliography

Block. (n.d.). Retrieved from Bitcoin Wiki: https://en.bitcoin.it/wiki/Block

Block 0. (n.d.). Retrieved from Blockchain.com: https://www.blockchain.com/btc/block/000000000019d6689c085ae165831e934ff763ae46a2a6c172b3f1b60a8ce26f

Bayer D., Haber S., Stornetta W. S. (1993). Improving the Efficiency and Reliability of Digital Time-Stamping.

Marco-Gisbert, S. S. (2019). Assessing Blockchain Consensus and SecurityMechanisms against the 51% Attack. *MDPI applied sciences*, 1-17.

Michael Nofer, P. G. (2017). Blockchain.

Mining. (n.d.). Retrieved from Bitcoin Wiki: https://en.bitcoin.it/wiki/Mining

Nakamoto, S. (2008, October 31). *Bitcoin: A Peer-to-Peer Electronic Cash System.* Retrieved from bitcoin.org: https://bitcoin.org/bitcoin.pdf

Stornetta, S. H. (1991). How To Time-Stamp a Digital Document. *Journal of Cryptology*, 99-111.

Voshmgir, S. (2019). What is Blockchain? In S. Voshmgir, *Token Economy.*

Zibin Zheng, S. X. (n.d.). An Overview of Blockchain Technology:Architecture, Consensus, and Future Trends. *2017 IEEE 6th International Congress on Big Data.* 2017.