A Comprehensive Study on the Scalability Challenges of the Blockchain Technology

En omfattande studie på Blockchain teknologins skalbarhetsutmaningar

Edin Ademi
Abstract

Blockchain is a decentralized ledger structure that attracted the attention of the public because of Bitcoin, and since then the interest in the technology has risen rapidly. The reason for the success of the blockchain technology is that it gives benefits of centralized networks regarding security, anonymity and data integrity in a decentralized network where no authority or organization is governing the transactions. There are however limitations and challenges of the technology that have to be considered, such as scalability, security and privacy when the implementation scales upwards. We take a look at the current research on challenges and limitations of blockchain, and how these challenges affect the user. Recommendations on future research directions are provided for researchers.
Sammanfattning

Acknowledgements

To my family, I deeply thank you from the bottom of my heart for your continued support. Thank you!

I express gratitude to my supervisor, Mia Persson, for her willingness to go above and beyond to guide and help me throughout this process. Furthermore, I would also like to extend my gratitude to the professors at Malmö University that gave me every possible opportunity to reach my goals and lastly I would like to thank my friends that I met during this incredible time for their continued support and always cheering me up!

Edin Ademi, October 28, 2017
# Contents

**Keywords**  
6

1 **Introduction**  
1.1 Study Idea and Objective  
1.2 Motivation  
1.3 Research Question  
1.4 Hypothesis and Study Goals  
1.5 Outline of Thesis  
1.6 Limitations  
8

2 **Previous Work**  
2.1 Blockchain Challenges  
11

3 **Research Methodologies**  
3.1 Research Methodology  
3.2 Conducting the Literature Review  
3.3 Alternative Research Methods  
3.3.1 Quantitative Approach  
3.3.2 Arguing for the Qualitative Approach  
13

4 **Results**  
4.1 The Blockchain Technology  
4.2 Consensus Protocols for Blockchain  
4.2.1 Proof-of-Work  
4.2.2 Proof-of-Stake  
4.2.3 Practical Byzantine Fault-Tolerance  
4.2.4 Selfish Mining  
4.3 Blockchain Scalability  
4.3.1 Double Spend Attack  
4.3.2 Block Size Incrementation  
4.3.3 Segregated Witness  
4.3.4 The Greedy Heaviest-Observed-Sub-Tree  
4.4 Blockchain Security  
4.4.1 Security Risks  
4.4.2 Security Enhancements  
16

5 **Discussion and Future Works**  
27

References  
29
Keywords

**Bitcoin** - The first introduced cryptocurrency that implemented the blockchain technology.

**Ethereum** - A popular blockchain implementation that handles transactions of smart contracts.

**Smart Contracts** - Transaction of contracts that consist of properties, money and shares done in the blockchain technology.

**PoW** - Stands for Proof-Of-Work, which is a popular consensus algorithm for blockchain implementations.

**BFT** - Stands for Byzantine Fault Tolerance. This is a replication algorithm to tackle the Byzantine Generals Problem.

**DAO** - A digital decentralized autonomous organization used in Ethereum.
Chapter 1

1 Introduction

The way we conduct transactions over the internet has changed very little. We rely on organizations and authorities as intermediaries to exchange the things of value we are in possession of (1). These intermediaries are governments, banks, digital platforms and colleges that establish and provide our identity and vouch for us being honourable (1).

These intermediaries are doing a great job, however, they are limited because they use centralized servers and these are targets for being compromised (1). To provide us with the service of transactions, they usually take a fixed rate fee or in some cases, a percentage based fee (1). These centralized servers are governed by intermediaries that collect and regulate the data that they are in possession of, this means that we cannot access our own data, our privacy is undermined. The intermediaries are usually slow and unreliable due to the personnel in some cases having to validate transactions (1).

During 2008, Bitcoin published their white paper and they delivered a prototype of the first implementation of Bitcoin two months later (1, 2). The individual or group behind Bitcoin called ”Satoshi Nakamoto” was able to create a new class of decentralized currency. The release of the proof of concept implementation of Bitcoin and the white paper release was crucial to the growth of Bitcoin (1, 2). Unlike previous proposals, the working prototype confirmed that the system they introduced was feasible. Along with that, it was also introduced as a open-source project which was an excellent way of attracting developers to maintain and support the growth of the system (2).

The blockchain technology gathered the attention of the technology world on the basis of Bitcoin. However, its capabilities extend far beyond the crypto currencies. Its implementation vastly improves already existing applications and it facilitates new applications that previously could not be deployed (1, 2). The goal of the blockchain technology is to create a decentralized environment where no third part is governing the data and transactions occurring within the implementations (1).

Blockchain provides us with benefits such as immutability, transparency and it redefines trust which enables secure and transparent solutions which could be made public or private (2). The technology is based on a distributed ledger structure. The distributed ledger structure is a way to make the transactions transparent for the populace and circumvent the intermediaries. The structure allows a distributed ledger to be created and shared between distributed computers on a network (2). The ledger is not administered by authorities or companies and the transactions can be viewed by any user within the network, hence the middleman for transactions of users and companies are cut out which in turn reduces costs of transactions (2). All of the transactions that are made within a distributed ledger structure are encrypted and they are all verified by other computer nodes on the distributed network. If the consensus amongst these computers is that the transaction is valid, a new block of data is added to the chain and is then shared to all on the network (2).

The digital ledger which is also referred to as a blockchain is called that because it is built upon a design where changes can only occur when new information is added. Each new addition, a block, contains a set of new transactions that references a previous transaction
in the chain (4). The data is recorded in a public ledger which includes the information of the transaction and every transaction that has been completed, this information is shared to the other nodes (2, 4). These characteristics of blockchain makes it more transparent than centralized networks where a third party is governing the network and sets their own regulations and restrictions (3). Unlike tradition centralized ledgers, such as the financial institutions these are replicated globally within networked computers and it is accessible anywhere as long as internet connection and a computer is attainable (4). A part of these participants in the network are called miners, they are responsible for detecting transaction requests from users - aggregating, validating and adding them to the blockchain as a new block (4).

The blocks in a blockchain are only added if the consensus amongst the miners is that the transactions are valid (4). The miners are a crucial part of the blockchain because they provide the blockchain with the security that each new block addition is valid and each transaction is fulfilled completely. They are also implemented to avoid attacks, such as double spending that would otherwise be possible in the blockchain if miners did not validate each transaction that occur in the blockchain.

With the rapid development of the blockchain technology there are some studies that focus on the security and privacy aspects of blockchain. However, the majority of them are mainly focused on blockchain implementations such as Bitcoin. There are insufficient research based on a systematic examination on the security of the blockchain technology as of current (5). Major problems can occur if organizations implement blockchain within their system such as smart contracts (where the block contains information of properties, pharmaceuticals etc.) with crucial and valuable information and new threats of being compromised are discovered (5). Based on previous research, Li X et al. discovered that of the consisting 19,366 smart contracts in Ethereum (which is the second most popular blockchain implementation), 8,833 of them are vulnerable. This is due to them consisting of valuable information such as property transactions. If the system is being compromised it can bring great harm to the end user (5). In June, 2016 a decentralised autonomous organization that was instantiated on Ethereum where the objective was to provide a new decentralised business model for organizing commercial and non-profit enterprises, was attacked, resulting in DAO losing of about 60 million dollars (5). In March 2014, a hacker group exploited transaction malleability which is an attack that lets the attacker change the unique ID of a transaction before it is confirmed on the network. This allows the attacker to pretend that a transaction never happened. The attack took place in a trading platform called MtGox, which caused the collapse of that platform and a loss of 450 million dollars worth of Bitcoin stolen (5). This is grounds for further investigation and calls for a review of the different proposals for maintaining the security within the blockchain technology and identifying possible advantages and disadvantages of the proposals.

1.1 Study Idea and Objective

Blockchain is an emerging technology that is vastly implemented into existing applications and new applications to facilitate the advantages of the blockchain technology. We are going to conduct a study on the blockchain technology and its capabilities, this will be fulfilled with a qualitative research methodology on previous research in the same field. The focal point of our study is on the challenges in terms of scalability of the blockchain and its limitations at this present time. We also take a look at the proposed ideas to
deal with the scalability challenges of the technology and our aim is to discover how the end user is affected and if the proposed solutions confront these limitations.

1.2 Motivation

The motivation for undertaking this work is to primarily investigate the blockchain technology in terms of scalability. As a result of the field being in the early stages, more research is required. The majority of the research in the field of blockchain are solely focused on the digital currencies and very few of them are done within the underlying technology. The secondary factor is to review the scalability issues of the blockchain and if the issues affect the users privacy when the network scales upwards.

1.3 Research Question

Based on the research gaps that were defined as part of the literature reviews, we defined one primary research question paired with two sub-research questions. Our research questions are:

1. Are there any scalability issues within the blockchain technology that affect the end user in terms of privacy and integrity?
   (a) Are there any existing solutions proposals for the aforementioned scalability issues? Identify possible advantages and disadvantages of the proposed solutions.

1.4 Hypothesis and Study Goals

The goal is to conduct a comprehensive study on the limitations and challenges of blockchain in terms of scalability so the field is understood more thoroughly by providing qualitative analysis of proposals of solutions for the scalability issues of blockchain. We also identify the advantages and disadvantages of the proposed solutions. Therefore we can break down the project goals to a number of steps to conduct the study:

1. Conduct a literature review on previous research on blockchain challenges.
   (a) Evaluate the findings based on the qualitative research on previous works in terms of scalability of the blockchain.
   (b) Analyse the proposals for the solution of the scalability issues of the blockchain and then identify the advantages and disadvantages of the proposals.

1.5 Outline of Thesis

The remainder of this study is organized as follows. In Chapter 2, we introduce the related works within the field of blockchain. The chapter is aimed to describe the problem of scalability within blockchain and to highlight why this study was undertaken. In Chapter 3, we introduce the research methodologies, techniques and tools we utilize to conduct the study. Chapter 4 we publish our results from the study based on previous works. In Chapter 5 we analyse and evaluate the data and the qualitative research on solutions for the scalability issues of the blockchain technology and we analyse the different proposals with possible advantages and disadvantages.
1.6 Limitations

The limitations of our study are on the blockchain technology, scalability, challenges and proposals to tackle the scalability issues. The reasons for these limitations are because we are solely focused on the underlying technology. Blockchain implementations are introduced frequently so investigating each new blockchain implementation is not feasible. The inner works of Bitcoin and other crypto currencies are touched upon to investigate the proposals for handling the scalability issues. However, a comprehensive study on the digital currencies that implement blockchain is not done, and therefore are outside of the scope of this study.
Chapter 2

2 Previous Work

This chapter serves as a literature review whose purpose was to primarily identify the current state of the art regarding research within the blockchain technology, but also incorporate and highlight the problems and issues of the technology.

2.1 Blockchain Challenges

A systematic map was used to conduct a study by Yii-Humoo J. et al. in which the current state regarding research within the blockchain technology is investigated (3). A search protocol was utilized to gather all the relevant scientific papers from the databases. A screening of the papers was done, to filter out irrelevant papers that were not within the blockchain technology (3). The findings based on this study were that the majority of the papers were primarily focused on the challenges and limitations as presented by Swan (3, 7). They identified three categories of research within the blockchain technology, these are blockchain reports, blockchain improvement and blockchain application (3). The blockchain report category encompasses previous reports on ideas and solutions within Bitcoin and Blockchain.

The security and privacy aspect of blockchain were the major research topics (3). In the study of Jesse Yii-Humo et al. they were able to identify four research gaps and they call for more research within these gaps (3). The first major research gap is within the latency, size, bandwidth, versioning, hard forks and multiple forks, they identified this gap due to not having any previous research done within these challenges (3). The second gap they identified was the usability aspect of the blockchain technology from the perspective of the developer (3), this gap is identified by looking at the Bitcoin API and the analysis is that the API is not developer-friendly and it is difficult to use. The third research gap is that the majority of the research within the scope of blockchain, is done on Bitcoin, they call for more research on smart contracts and other implementations of blockchain (3). This hinders the knowledge of the blockchain environment outside of the cryptocurrencies (3). Jesse Yii-Humo et al. call for more research in other environments where blockchain can be applied because it can reveal and produce better models for doing transactions in different industries (3). The final gap they identified was the low number of high quality publications within this field (3). Based on their conducted study, they provided recommendations for future research topics within the blockchain technology, a topic that they recommended was within the challenges of blockchain and how to overcome them and the scalability issues of blockchain (3).

Xiaoqi Li et al. presents a survey on the security of the blockchain systems (5). Their argument for conducting their study is because the studies performed on the security of blockchain lack a systematic examination on the risks of blockchain, the corresponding real attacks and the security enhancements, their survey focuses a comprehensive perspective of the risks and challenges (5). They divide the risks into nine categories and detail the causes and the consequences of the risks (5). By conducting their comprehensive study, they are able to identify security enhancements that mitigate the risks of these attacks, they present solutions such as SMARTPOOL, QUANTITATIVE FRAMEWORK, OYENTE, HAWK and TOWN CRIER.
Marko V. presents a study where he reviews the recent efforts towards improving scalability within the blockchain technology (24). He introduces two concepts in blockchain, Proof-of-Work and Byzantine Fault Tolerance (24). Blockchain implementations that are based on PoW offer good node scalability, but with poor performance. BFT based implementations offers good performance but to a limited set of nodes (24). A concept introduced by Sompolinski et al. which resolves the conflicts in PoW blockchain, called GHOST is presented in this study in which he gives an overview of the concept (24, 8).
Chapter 3

3 Research Methodologies

This chapter inspect the research strategies and clarifies the systematic research approach used to conduct the study, so it can be replicated to ensure validity and reliability.

3.1 Research Methodology

The research methodologies we considered in the study were qualitative and quantitative. With the qualitative approach we were able to understand the context of our problem and gain insight on previous research that had focused on the different challenges within the blockchain field. Based on the literature review and dissecting the different research papers that were published on the blockchain technology, we had an idea that the research would be more in-line with the theoretical approach rather than the statistical or experimental approach, which is more in-line with the quantitative research method.

The reason we chose the qualitative approach is to get a comprehensive understanding of the problems we were about to conduct a study on. This allowed us to focus on the literature we had gathered and dissect the different problems of the blockchain and then form our own problem areas in the field based on the literature review. The qualitative approach is utilized to dive deeper into a problem to better understand the problem area. There are many research papers on the field of blockchain, therefore the qualitative approach was used to gather information of previous research.

3.2 Conducting the Literature Review

The literature review used in this study was done systematically by starting with a number of keywords to identify relevant literature to the domain we are researching. We used a number of databases, such as ACM, IEEE, Google Scholar to aid us in extracting relevant research papers within the field. The starting keywords were "blockchain", "blockchain scalability", "blockchain vulnerabilities", "blockchain challenges", "blockchain security". The result of these queries were used as a starting foundation with relevant literature within the field of blockchain, and then we used their references, citations and literature review to extract information that were relevant to our study. The process was iterative and it started early in the research phase and we continued to build upon the literature as the research project went on.

1. Techniques

(a) To conduct our study we used a systematic research process that focuses on being objective and gathering information. This systematic process is followed so that the reader can replicate this study by following the same steps. The steps to conduct our study were:

i. We started our research process by identifying a problem area within the blockchain technology.
ii. Our next step was to review the literature within the field to get a better grasp of the problem. This step served as a foundation of knowledge for our problem area, this also helped us identify how previous research were conducted and what their conclusions were.

iii. In this step we had to adjust our initial problem that we identified in step (i) due to the field being too broad and we had to narrow down the scope of our study. Once we adjusted the initial problem, we also defined the target group for our study.

iv. This step served as a planning phase for our research, here we discussed how we are going to collect the data and this plan was followed throughout the research process.

v. The last step of our study was the collection of the data and analysing the data that we used in our study. This data is analysed so that the research questions from the study can be answered.

2. Tools

(a) A big part of our study was the gathering of literature and academic papers were our primary necessity, therefore we utilized tools such as Google Scholar, ACM, IEEE and Malmö University’s Online Library to collect our academic literature.

3.3 Alternative Research Methods

There are three research approaches — qualitative, quantitative and mixed methods. Research designs are apart of these, and they provide specific directions and procedures that are implemented into the study. Based on our literature review, and studying the research approaches we came to the conclusion of using the qualitative research approach. This section serves as an explanation to the selection of the qualitative approach over the quantitative approach.

3.3.1 Quantitative Approach

In the quantitative approach, there are research designs such as true experiments, quasi-experiments and applied behavioural analysis and single-subject experiments (9). If we decided to perform experiments on the different blockchain implementations and then analyse their performance, this research approach would be more suitable than the qualitative research approach. However, our study was theoretical and because this approach is more suitable for experiments or surveys, it is not applicable to the study we wanted to conduct.

3.3.2 Arguing for the Qualitative Approach

In the qualitative approach a key element to collecting data is by observing (9). The study we conducted was based on a qualitative research approach because we were reliant on previous works within the field. We analysed the studies we had gathered during the literature review and we observed the different proposals that had been suggested for the scalability issues of the blockchain. Based on the focus and direction of our study, we investigated studies and the research approaches that had been used within the
blockchain technology. The conclusion was that the majority of the research papers that were theoretical and more in-line with our study, had used the qualitative research approach. We did not conduct any experiments on the scalability of the blockchain or the proposals, therefore the theoretical approach of the qualitative research method was more in-line with the study we conducted.
Chapter 4

4 Results

This section presents the qualitative data collection that we gathered throughout the study. Based on the findings here, we are able to understand and answer the research questions fully.

4.1 The Blockchain Technology

The blockchain technology got the attention through Bitcoin, when it was first introduced (7). The white paper that was released by the group "Satoshi Nakamoto", introduced a new concept of a distributed digital currency system (13). The system worked by chaining cryptographic blocks of data together (7).

In digital networks, the data is usually transmitted through multiple parties, the primary issue with this concept is the need to verify if the data received is the most up-to-date or if it is valid (7, 13). The solutions that tackle this problem is usually done by giving someone the rights to handle the data, this could be an authority or an organization. However, constantly relying on an intermediary is expensive (7). The reason blockchain is popular is because it disrupts this idea that an individual need to rely on a intermediary to authenticate a transaction. The data is verified through mathematical algorithms and computational powers (7, 13). The blockchain allows any computer node to validate the transactions, the incentive to perform validation of transactions is that the miner gets compensated for the computational power provided (7, 13). The difference between centralised networks and the blockchain is that, in the centralised network there are third parties that regulate and confirm transactions (7). In the blockchain, the computer nodes work together to perform the validations instead of relying on one single point entity to do it (7). The difference is that centralised networks charge for auditing and keeping records, which is not available to the public. In blockchain implementations, the records are public, which means they are visible to anyone in the network. The blockchain transactions consist of crypto currencies and in smart contracts. The transactions are made by users or firms that want to avoid having a centralized network that is able to keep track of the transactions they have performed.

Banks and governments are implementing blockchains to revolutionize the way transactions and information is stored. Their goals are speed, cost, security, fewer errors and elimination of a single point of attack or failure. The blockchain is not a database or a file, it is like a global spreadsheet or a ledger that holds the resources of a peer-to-peer network to verify and approve each transaction that occurs in the ledger. Each blockchain implementation, is distributed. The way it is distributed is by allowing volunteers to provide computational power, so there is no central database to attack. The volunteers are also called miners that get rewarded for the computational power they provided. The blockchain is encrypted, it has public and private keys that maintain virtual security. The public key is the identifier for a user when a transaction is performed. The transaction is then visible to the public. The private key is used to perform transactions in which the rightful owner should only have access to.

The blockchain is the public ledger of all blockchain implementation transactions that have been performed (13). The platform is constantly growing as miners add new blocks.
every ten minutes to record the most recent transaction (7, 13). The blocks are added in a linear order, and every full node (all the computers connected in the blockchain network, that performs validation for transactions) has a copy of the blockchain, this is automatically downloaded once the miner joins the network (13). The blockchain has complete information from the very first transaction recorded, to the most recent one (13). The key innovations are the decentralized trustless transactions.

The blockchain is maintained by a network called miners, which are compensated for their efforts because they provide their computational powers (21). In the case of Bitcoin, the compensation is in forms of Bitcoin. For every block that has been added to the blockchain, the miners get a share of the bitcoins provided for that block. The compensation differs based on the blockchain implementation. The transactions are protected with cryptographic techniques which only allows the rightful owner of a address to transfer funds (21). The miners are in charge of recording the transactions into the blockchain, which also determines who owns the transactions (21). Because of the difficulty of mining a block, it is unrealistic for a single computer node to hash a block within reasonable time. Therefore there are mining pools organized that gather miners with single computer nodes which then collaborate to hash the blocks at a faster rate and the reward is shared amongst the miners in the mining pool (21).

![Figure 1: Blocks that contain the previous hash, timestamp and nonce that is used as an input for the complex algorithm.](image)

The blockchain transaction records are stored in blocks. Each block includes a unique ID and the ID of the preceding block (as seen on figure 1 and figure 3) and the first block is called the genesis block (13, 17). A valid block contains a cryptopuzzle of the preceding hash, the hash of the transactions of the current block and a list of transactions that have been verified by the miners (13). Timestamping is a basic function that permanently registers on the block-chain that an action was performed at that specific time. Timestamping is crucial to seeking the truth when validating if a transaction has been performed previously with the same ID. A block also contains nonces, which is an arbitrary random number that is only used once. The timestamp, nonce and previous block hash is taken as input to solve the complex algorithm.

Any miner may add a valid block to the chain, by publishing the block to the network of miners (17). If two miners create two blocks with the same preceding block, the chain is forked into two branches, which forms a tree (13, 17, 21). Other miners may then add new valid blocks to either branches. The formation of branches is not desirable since the miners have to follow the globally agreed set of transactions. To resolve forks, miners are advised to add valid blocks to the longest chain (as seen on figure 2), or if they are of equal length, to the first one they encountered (17, 21).
The blockchain consists of a series of blocks, however figure 3 shows three only. The block shows the transaction the users have sent, the Proof-of-Work shows the complex unique number that is created by the miners that serves as an identifier for that specific block and the previous block has a reference pointer to the previous block Proof-of-Work identifier (11, 13). This concept ensures the integrity of the entire blockchain, because it provides the possibility of tracking the genesis block, which is the first block ever added to the blockchain (13). Fraudulent attempts can easily be detected because a change in the Proof-of-Work hash will disrupt the entire blockchain chain (11, 13). To avoid dishonesty amongst the participants of the mining, a consensus mechanism is implemented to assure the authenticity of the transactions, and if the consensus amongst the majority is that the transactions are valid the miner gets rewarded in crypto-currency (13, 11).

New transactions are not directly added to the blockchain, during the consensus period, they ensure that these transactions are stored in a block for a certain time before they are being transferred to the blockchain (13). Thereafter the information can no longer be altered. Using cryptography to validate the transactions, people all around the world can trust each other by transferring assets peer-to-peer (13).

One advantage that blockchain brings over centralized systems is that they don’t have a single-point of failure for the architecture, the functionalities of the network still persists even if a computer node has complete or partial failure (13, 14). If a computer node has partial failure or complete failure, it is excluded from the consensus mechanism of the blockchain implementation (13). The blockchain still persists if there is a failure in a computer node, because there are several computer nodes working on the same block. the blockchain is never reliant on one computer node. This increases the trust of the people if they believe in the underlying system, due to them not having to rely on
intermediaries to handle their transactions that are targets for security breaches because they possess valuable information (13).

4.2 Consensus Protocols for Blockchain

How to reach consensus in blockchain is inspired from the Byzantine Generals Problem which is defined as if a group of generals who command a group of the Byzantine army circle the city, the attack will only work if the consensus amongst the generals is that they will attack the city, if there is a deviation from that command, then the attack will fail (30). If there’s a traitor amongst the generals then he might rely different decisions to different generals (30). This environment is not trustful, which is also a challenge amongst the blockchain technology. The problem is that there is not a central entity that ensures that the distributed nodes are all trustworthy, so there might be traitors in the consensus process that deviate from the consensus of the majority to avoid a new block being added to the blockchain. These traitors can cause attacks such as selfish mining and double spending. To handle the issue, protocols for the nodes are applied to ensure that ledgers in different nodes are consistent and do not deviate (30).

4.2.1 Proof-of-Work

PoW is a consensus mechanism in blockchain implementations. The strategy is that each node in the network is calculating a hash value of the constantly changing block headers. A block header is the entire block and depending on the blockchain implementation, in most cases the block header contains timestamp, nonce, hash and previous hash. The consensus requires that the value must be equal or smaller than the given value (17, 18, 19). In the decentralized network, all the participants calculate the hash value, and by using different nonces for each block until the value is within the limit of the consensus protocol. When one of the participants finds the relevant value, all the other nodes that participated need to confirm the correctness of the value (19). The miners verify the validity of the value by taking the hash-value found by one of the miners, and then hashing the same block and if the value is below the target-value, then a consensus is agreed upon by the miners. After that the transactions in the block are validated and the collection of the transactions for the calculations are approved to be the authenticated result, which is done by adding a new block to the blockchain (19). The nodes that calculate the hashes are called miners and the procedure to solve the complex mathematical algorithms is called mining, since the calculations for authentication are time consuming and requires computational resources, a mechanism to reward the participants is also proposed (17, 19).

A block contains nonces that a miner must set in a way that the hash of the entire block is smaller than a set value, and the value is dynamically set based on the difficulty of the mining, the adjustments are made based on the block-mining rate and the computational power of nodes participating in the blockchain, this is done to maintain the expected block-mining rate of one block for each 10 minutes (18). This latency is called the block frequency and is a critical number for blockchain implementations together with the size of the block which is set to 1MB per block (18).
4.2.2 Proof-of-Stake

Proof-of-Stake is a more energy saving alternative to Proof-of-Work and it works by demanding proof of ownership on the amount of currency the possess, the reason for this is that it is to believed that the person with more currency would be less likely to attack the network (19). In PoS there’s a concept of currency age. The age is determined by the value multiplied with the time period of its creation (17, 19). Currency holders are encouraged to hold coins for a longer period, because the reward is that the holder gets more rights in the network and a certain reward depending on the age of the coin. Since PoS encourages people to hold their currency and with the introduction of the coin age, it means that the blockchain is not entirely reliant on PoW (19). The security of the blockchain that utilizes the PoS approach increases with value in the blockchain. The attacks are also less likely due to the coin holders having to possess a large number of coins and hold them for a long time, which decreases the incentive to attack the blockchain (17, 19).

4.2.3 Practical Byzantine Fault-Tolerance

PBFT is a replication protocol to tackle the Byzantine Generals Problem. In distributed systems, PBFT can be a practical method to solve the transmission errors (20). In 1999, Miguel C. and Barbara L. presented their PBFT system in which their algorithmic complexity consisted of several terms in which improved efficiency of the algorithm. A new block is determined in a round and in every round a primary leader would be assigned based on protocols and these are the ones for ordering the transactions (20). The process could be divided into three phases, the pre-prepared phase, the prepared phase and the commit phase as seen on the study of Miguel and Barbara (20). PBFT requires that every node is known by the network, so in each phase a node would only enter the next phase if it has received votes from over 65% percent of the nodes (20). In contrast to PoW, modern BFT protocols have been proven to sustain upwards of ten thousand transactions with the limitations of network-speed, and these are not only concepts, these are practical systems (20).

4.2.4 Selfish Mining

The probability of mining a block is proportional to the computational power used for solving the specified block. A single miner is unable to mine a block for years, therefore typically miners organize mining pools where the miners work together to mine a block and when they successfully mine a block, they share the revenue associated with that block.

Eyal and Sirer (21) introduced a Selfish Mining attack that made it possible for miners to obtain a larger revenue than the fair share by not presenting their discovered blocks. The primary insight into the selfish mining approach is that the attacker wants the honest miners to waste computational resources on blocks that are destined to not be on the blockchain (21). This is done by selfish miners revealing the blocks they have mined so that the honest miners work is invalidated (21). The selfish miners keep their blocks private, while the honest miners keep working on the public branch (21). Because the computational power is still higher in the public branch, the selfish miners will eventually be surpassed, however what they do is that they reveal blocks from the private branch.
to the honest miners so that honest miners will switch to the recent revealed block, abandoning the blocks on the public branch that they were working on (21).

4.3 Blockchain Scalability

In December 2014, Bitcoin's network was able to process roughly 90,000 transactions per day, which has since been growing, however still far from the 150 million transactions per day of VISA (17, 24). The problem that arises with Bitcoin is that if it is not able to scale upwards to match the demand, it will increase the transaction fees and users might look at other forms of payment. The reason for the low amount of transactions in Bitcoin is due to it being a relatively new platform, but once the demand rises and the need to process transactions at a higher rate, the guarantees of security may not hold and attacks may become more common (23).

Today’s representatives of blockchain take up to 10 minutes to confirm transactions. The scalability is dependant on a number of factors, these are:

Maximun throughput: This is the maximum rate of which the blockchain can confirm transactions.

Latency: Time for a transaction to be confirmed. The transactions are confirmed once they are included in a block, which is roughly 10 minutes (32).

Bootstrap time: The time it takes for a new computer node to download the history necessary to validate new transactions (32).

Cost per Confirmed Transaction: The cost in USD of consumed resources to confirm a single transaction (32).

Based on the research of Kyle C. et al. the cost per confirmed transaction can be between $1.4- $2.9, where 57% of the cost is on electricity (32). The cost of transaction includes the mining (Proof-Of-Work), the hardware for mining, transaction validation, bandwith and storage (32). Mining consumes roughly 98% of the total cost of a single transaction (32).

According to the founder of Ethereum, Vitalik Buterin, the main challenge within the blockchain is the scalability and privacy concern (31). Current blockchain implementations process between 3-20 transactions per second, which is several magnitudes away from the amount of processing power required for the blockchain to go mainstream (31). Individuals or companies are not too keen on publishing their records onto a public database that can be read by governments, foreign governments and business competitors (31). There are no ”holy grail” technology that can be implemented that allows the users to do absolutely everything they can do now in blockchain but in privacy. Developers need to focus on partial solutions and mechanisms that are designed to bring privacy to specific parts of the application (31).

Despite the power of the blockchain, the present form of the blockchain lacks transactional privacy (29). The entire process of conducting a transaction is done over a public network which are therefore visible to the public (29). Parties can create multiple public keys to increase anonymity, the values and transactions are still publicly visible. Hawk is a proposal by Ahmed et al. (29) which is a framework to develop privacy preserving smart contracts, because in the era of blockchain 2.0, privacy is a bigger issue due to now having valuable information stored in smart contracts (29). Developers can use this framework to write private smart contracts (29). The Hawk contract is compiled by
three pieces, (1) The program that will be executed on all virtual machines of nodes (2) The program that will be executed by the users of smart contracts and (3) The program that will be executed by the manager, which is a trustworthy part in Hawk (29). The manager can see the private information of the contract but will not disclose it(29). If the manager aborts the protocol of Hawk, it will automatically be penalized. Hawk can be used to protect the privacy of the users of blockchain implementations (29).

The difference between the PoW consensus and the BFT consensus is a set of important blockchain properties. Marko V. identified a set of properties, that he then compares in terms of scalability (24). The properties are node identity management, consensus finality, scalability in terms of nodes, scalability in terms of clients, performance throughput and performance latency, power consumption, tolerated power of an adversary, network synchrony assumptions and correctness proofs (24).

Node identity management means how nodes are identified in PoW compared to BFT, and it is probably their most fundamental difference, in PoW the nodes are completely decentralized and open to anyone, while in BFT they are permissioned and every node needs to have the unique IDs to the other nodes (24). This is a powerful feature of the PoW implementation due to the nature of allowing anyone to participate and contribute to the system, and they are very common in permission less blockchains, PoW also prevents Sybil attacks (24). In contrast the BFT consensus needs to know every participating node in the network, this can also be identified as a "centralized" identity management due to a trusted party is issuing the identifications to the nodes (24). This puts the BFT approach at a disadvantage compared to the PoW approach, however due to more centralized organizations implementing blockchain, such as banks, in some cases require that permission only nodes and identified and certified with cryptography (24).

Consensus finality is a property that says that a valid block that has been added to the blockchain, should not be removed. PoW violates this property when two nodes append a block to the same block, which results in a fork (7, 13, 24). As in the case of Bitcoin the forks are handled by the longest chain rule, thus breaking the consensus finality by removing the shorter chain or the GHOST rule. BFT satisfies this property which gives BFT an advantage over PoW, this allows the blocks that have been added in a BFT implementation immediate confirmation if the transaction has been included (24).

When it comes to scalability with the number of clients, both PoW and BFT support thousands of clients all connected at once (24). Scalability of nodes in PoW has been proven due to it effectiveness in Bitcoin, however Bitcoin had to create miningpools, which in some aspects centralized the node management (24). BFT on the other hand has had the stigma of poor scalability, the protocols in databases and system communities are perceived as not scalable (24).

The performance of Bitcoin is very limited, which handles up to 7 transactions per second and a 1 hour latency with 6 block confirmation. PoW based blockchains face inherited performance problems (24). As mentioned before, the PoW scalability is reliant on two things, the block size and the rate of the block creations, and if the block size is increased, it also increases the security risks, such as the latency is increased and therefore potential trees in the blockchain are created and in which may lead to double spend attacks (24). BFT protocols have been able to sustain thousands of transactions with practically network speed latency (24).

With the amount of Bitcoin transaction that occur every day, the blockchain has become
heavy and it has reached upwards of 100GB already, this is done because the transactions need to be validated (24). And due to the block size and rate of block creations, Bitcoin is unable to match the demand. There are a number of efforts to tackle the problem of scalability, a proposed idea is to optimize the storage of the blockchain. In this proposal the old transactions that are stored in the blockchain are removed by the network and a database is used to hold the non-empty address trees (24). In this way, the nodes that are validating the transactions do not have to store the previous transactions that are not relevant to them. Another proposal was Bitcoin Next Generation which is a proposal where the blocks are now decoupled and there are leader blocks and micro blocks that handle the transactions (24). Miners would compete for the leader block, these would be the ones in charge of generation of new micro blocks (24).

<table>
<thead>
<tr>
<th>PoW consensus</th>
<th>BFT consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open for anyone to join the blockchain network.</td>
<td>Every node needs to know the other nodes in the network.</td>
</tr>
<tr>
<td>Violates against consensus finality, when two blocks are added to the same previous block.</td>
<td>Immediately confirms if a transaction has been added</td>
</tr>
<tr>
<td>Proven to be scalable in a larger scale, such as in Bitcoin.</td>
<td>Protocols that are used are not scalable.</td>
</tr>
<tr>
<td>Performance is very limited, handles 7 transactions per second, has inherent performance problems.</td>
<td>Sustains thousands of transactions in network speed latency.</td>
</tr>
</tbody>
</table>

Table 1: A table that concludes the different scalability issues in the PoW consensus and BFT consensus protocols

4.3.1 Double Spend Attack

Double spend attacks is a method to override the main chain to reverse transactions and it works by, if the attacker pays a person and then in secrecy builds a chain of blocks where the payment is not included (16). By releasing the chain he can cause a replacement in the ledger where the payments are erased or redirects the payment to somewhere else (21, 22).

This requires a lot of computational power which makes the attack unlikely since the honest nodes have a lot of computational power, however there was a case of a mining pool in Bitcoin having over 40% of the total computational power (24). If the attacker has a lot of computational power there is a possibility that the attacker can generate blocks that could replace the honest longest chain and that enables the attacker to replace the main chain at will (24). When the attacker has more computational power than the honest nodes, it is called the 51% attack (also known as the majority attack) (24).

The Bitcoin protocol becomes more susceptible to double-spend attacks when it scales upwards and tries to meet the demand. If we assume that the attacker creates blocks at a rate that is faster than the rate of the honest main chain, the attacker will always be successful with these types of attacks regardless of the length of the chain it aims to replace (24). The throughput of the blockchain can be affected by two elements, the size of the block and the block creation rate (11, 24). The difficulty of the computational
problem to create a valid block can also be lowered so that the creation of blocks is accelerated (13). Increasing the block size or increasing the rate of creation of blocks influence the blockchain in a negative way by introducing more forks which in turn reduces the security threshold of the blockchain (24).

4.3.2 Block Size Incrementation

The most discussed parameter in order to improve the blockchain scalability is the block size limit (33). Some proposals suggest that the limit should be removed or changed to a higher limit (33). A proposal that was suggested at BIP 100 (Bitcoin Improvement Proposals) suggested that the 1MB fixed limit should be changed to a new floating block size, which is determined by the consensus mechanism of the miners (33). Another suggestion was to increase the block size by 4.4% each 97 days until year 2063, which is a 17.7% increase per year (33). All of these proposals need to deployed through a hard fork (33). This means that the new blocks that have a bigger capacity of 1MB will be seen as invalid by the current version node (33).

4.3.3 Segregated Witness

Segregated witness also called SegWit, is a protocol to increase the block capacity and to provide protection from transaction malleability (33). It includes a wide range of features. Some blockchain implementations consist of 64-digit hexadecimal hash transaction identifiers (33). The transaction identifiers can be altered by the miners because the way the transaction identifiers are calculated the miners can make small changes, that will not change the meaning of the transaction, but it will change the transaction id. This is called third-part malleability (33). Segregated witness approach does not increase the block size limit, but it increases the amount of transactions that can be stored in a block (33). This approach in the best case scenario increases the throughput by four times (33). The segregated witness approach resolves the transaction malleability problem and thus allows new mechanisms to be implemented which could provide powerful tools for the scalability issues in blockchain implementations (33).

4.3.4 The Greedy Heaviest-Observed-Sub-Tree

A suggested implementation is proposed by Yonatan S. and Aviv Z. where the policy for the selection of the main chain in the block tree is differed (25). The advantage of this policy is that it maintains the security threshold against the 51% attacks even if the network is affected by delays and the attack is not (25). This allows the network to set higher rates for block creation and increase the block size without worrying about the 51% attacks which means a higher transaction throughput can occur within the network (25).

The basic concept behind the protocol is that blocks that are off the main chain can still contribute to its weight (25). It offers performance benefits over the standard longest chain protocol as is implemented in Bitcoin (25). It provides more secure ways to increase the block size and the rate of the block creations. A variant of GHOST is implemented into Ethereum although the performance has not been tested enough, however in 2016 the Ethereum throughput was less than 20.000 transactions per day (24, 25).
4.4 Blockchain Security

4.4.1 Security Risks

Blockchain has the reputation of being very safe, it’s decentralized, anonymous and transparent, and this is because they can use multiple generated addresses that identify an individual, instead of using their real identity (7, 13, 15). However in a study presented, it is shown that transactional privacy cannot be guaranteed by blockchain because the transactions balances is shown to the public by the public key (31). Another study also shows that a Bitcoin transaction can be linked to an individuals user information (15).

51% vulnerability is the a risk of blockchain. Blockchain relies on a distributed consensus mechanics to establish the trust of the nodes and blocks, however in some cases if a person controls more than 50% of the computational power, he can essentially control the blockchain (16, 17). In January of 2014, the mining pool gas.io reached 42% of the total computational power of Bitcoin. This raised concern that gas.io might exploit the situation if it reached higher than half of the computational power, however, they went out with a press release that they would avoid the 50% threshold. By launching the 51% attack, the attacker can manipulate and change the transactions that may lead to double spending, excluding transactions, hamper honest miners and impede the confirmation process from the honest miners (24).

In the case of Bitcoin, a user could technically have multiple addresses that do not link to the their real identity. This makes Bitcoin susceptible to criminal activities because it provides a safety net for criminals due to the authorities being unable to get the real identity of the criminal since the process is anonymous. There has been ransomware attacks where they would encrypt the persons files and ask for a ransom. The victim had to pay through Bitcoin to unlock their files and the reason for the transaction occurring on Bitcoin is because it is anonymous and cannot be tracked to them. In Bitcoin, a user is identified by the public key which can be seen by anyone, and a private key which is unique and only the rightful user is supposed to have it, this private key is the identifier and security credential. A discovery was made where a vulnerability was detected, where the private key could be acquired because the generated randomness of the key was not great. Once the private key of the user is lost it cannot be recovered. If the private key of a user is lost and retrieved by criminals, the user faces the possibility that his blockchain account will be tampered with.

Blockchain systems take measures to protect the privacy of the transactions that occur within the system. The user needs to assign a private key to each transaction. Andrew (27) conducted a empirical study on Monero which is a blockchain implementation, where he evaluates the weaknesses in mixins (a chaff coin introduced in Monero so that an attacker cannot interfere the linkage of actual coins spent in a transaction) (27). He found out that 66% of the transactions do not contain any mixins (27). 0-mixin transaction will lead to privacy leakage of the sender and by exploiting the weaknesses of Monero, they can interfere with the actual transactions with a 80% accuracy (27).

The DAO attack was targeted at the DAO contract which implemented a crowd-funding platform. This was deployed in Ethereum and after 20 days only the attack was launched towards DAO. Before the attack, DAO had already reached 150 million USD and the attackers were able to steal 60 million USD. The attack was launched by publishing a malicious smart contract which had a withdraw() function call to DAO in its callback.
4.4.2 Security Enhancements

A security measure to tackle the problem of mining pools reaching the 51% is by introducing something called SmartPool (26). It works by getting a transaction from node clients which contain mining tasks, then the miner performs hashing computation based on the tasks and returns the completed shares to the SmartPool client (26). When the number of completed shares reaches a specific amount they are committed to the SmartPool contract which is deployed on Ethereum (26). The SmartPool contract will verify the shares and reward the clients (26).

Loi et al. (28) proposed a concept to detect bugs in Ethereum smart contracts that is called Oyente (28). The concept leverages symbolic executions to analyze the bytecode of smart contracts and the execution model it follows is EVM (28). Ethereum stores the bytecode of its smart contracts in the blockchain, therefore this proposal is applicable to detect the bugs in the deployed smart contracts (28).
Chapter 6

5 Discussion and Future Works

In this chapter we discuss the results and answer the research questions. Possible future works within the field are presented at the end of the chapter.

The goal of our study was to get a better understanding of the blockchain technology and the challenges in terms of scalability it faces and analyse the possible proposals for the scalability issues. Based on the findings of our study we were able to answer our research questions and uphold our study goals.

From the study we conducted, we were able to address the scalability and privacy issues of the blockchain. We then identified possible solutions to the scalability issues of the blockchain. The results of our study shows that scalability of the blockchain is a major issue that needs to be addressed before the blockchain implementations can be utilized to their full capabilities. The scalability issues hinders the growth and solutions need to be implemented and actually be verified that it can handle large amounts of transactions for it to overtake the centralised networks by providing fast transactions and without a third party regulating and restricting the network. The findings of our study are relevant and important to the field of blockchain because it provides valuable findings on the scalability issues, such as the effect on the end users privacy and it also highlights the problem of the blockchain technology.

Our study differs from previous studies by providing new and updated information of scalability issues in the blockchain. Yii-Humoo et al. conducted a systematic study on the current state of the art research on the blockchain technology. Their findings were that the majority of the studies were within the Bitcoin, which is a adoption of the blockchain technology. Our study focused on the underlying technology, which is the blockchain technology, the consensus protocols and the limitations of the scalability within those aspects. Xiaoqi Li et al. conducted a study on the security of the blockchain and the different attacks that are possible within the blockchain technology. The study of Xiaoqi Li et al. was crucial to our study due to the security threats that are available in the blockchain technology. Based on the security threats we were able to identify if the privacy concerns were existent. Our study focused on the scalability aspects of the blockchain and if the privacy concerns affect the user when the blockchain scales upwards. Marko V. presented a study on the different consensus protocols. He compared the Byzantine Fault Tolerance consensus protocol that can be implemented in the blockchain and the most common consensus protocol, which is Proof-Of-Work. He identified a set of properties where the BFT protocol is better than the PoW protocol. However, the BFT protocol makes the blockchain more centralised in terms of all computer nodes running in a BFT consensus protocol need to know eachother, which is not the case in PoW. The work of Marko V. is mainly focused on the two consensus protocol and the scalability of these protocols. Our study is focused on the scalability of the blockchain, and not solely on the consensus protocols. We also take a look at the proposals for the scalability issues that are presented in the study, and we also identify the advantages and disadvantages of the proposals.

The methodology we used for the study was a qualitative approach. We were able to get a comprehensive understanding of previous works within the field in terms of scalability and this was the foundation of our study. The data collection from previous works was
done on varying time periods and due to the field being in the relatively early stages
and new studies being published frequently, the previous studies were relevant to the
study we conducted. A suggestion for future data collection within the field is to query
relevant studies based on the time period due to some of the studies presented in the
eyearly stages may not be relevant in the future.

Possible future works in the field of blockchain are presented based on the findings of
our study. Here are a few suggestions of studies that could contribute to the field:

Based on the findings of our study it would be an interesting study to see if the end
users that have performed transactions through blockchain implementations are aware
of the privacy concerns of blockchain implementations. The quantitative data collection
from the study could serve as a measurement of the awareness of the user and based on
the finding maybe more information needs to be presented to the public regarding the
concerns.

(i) A quantitative survey about the awareness of the end users that have performed
transactions through blockchain implementations regarding their privacy.

As online platforms for digital wallets gain millions of dollars worth of crypto curren-
cies flowing through their platform, there has been an increasing threat to the users
possession in these platforms. Based on the findings of our study, we noted that many
platforms had been compromised, such as MtGox, DAO, Tether where millions of dol-
lars had been stolen from users without reimbursement. Therefore a study on the online
platforms and the security of their system is vital to the field.

(ii) A comprehensive study on the online platforms for digital currency wallets and their
security.
References


