KU LEUVEN

FACULTY OF ECONOMICS AND BUSINESS

Feasibility of blockchain application as medium for collaborative systems or databases

How to replace a previously thought indispensable middleman with technology

Pelle Jacobs
r0364018

Thesis submitted to obtain the degree of

Masters in Information Systems Engineering
Majoring in Data Science

Promotor: Prof. Dr. Jochen De Weerdt
Assistant: Vytautas Karalevicius

Academic year: 2016-2017

Contents

Preface

Leuven, July 20, 2017.

Chapter 1

Introduction

THIS TEXT SHOULD BE UPDATED WHILE WRITING CHAPTERS

Hitting a market capitalization of $13.8 Billion in December 2013, it was clear that a once fringe cryptocurrency called “Bitcoin” had hit mainstream. By building on decades of research in decentralized currencies and cryptography, Bitcoin is the first successful implementation of a truly decentralized currency. Most of this success has been attributed to an innovation called “Blockchain”. However, this success has opened up a deeper concept: the power of technology to replace a previously thought middleman.

Now, the question is: “How can this idea to replace a middleman with technology be applied to collaborative, distributed databases and what is the role of blockchain is this concept?” The next three chapters provide an answer to this question:

In the next chapter, three main concepts that are essential aspects of this question are discussed. First public key encryption is explained, as it is the basis for modern encryption and digital signatures. Next, the concept of the blockchain, its properties and some well-known implementation are examined. Finally, the concept the thesis delves into distributed databases, the difference between distributed, decentralized and centralized databases and some non-blockchain examples of distributed databases and networks.

The third chapter explains how this thesis tries to solve the research question in the final chapter.

The final chapter investigates the viability of specific proposals to answer the research question: how every proposal tries to solve its specific problem, the advantages and the disadvantages of its approach.

Chapter 2

Prior research into relevant concepts

2.1 Public key cryptography

Public key cryptography is the backbone of most distributed systems as it provides both a way to encrypt a message as well as to validate the source of a message, without the need to agree upon a shared key. The most used implementation of this idea is the RSA encryption method, named after inventors Rivest, Shamir and Adleman [3].

2.1.1 Basics

In previous encryption systems before the public key cryptography, two parties who wanted to safely exchange messages needed to agree on a common cypher first. One of the more famous of such encryption systems is the Ceasar cypher [2], in which the cypher consisted of offsetting every character with a previously agreed upon number. However, if a third party manages to intercept transmission of this cypher itself, all encrypted messages can easily be decoded. Therefore, a trusted third party network is needed to securely exchange the cypher.

Public key cryptography uses a pair of hashes instead. These hashes are often referred to as the keys. These hashes are algorithmically generated so that they are cryptographically connected: data encrypted with one key can only be decrypted with the other key and vice versa [4].

One of these keys can be publicly broadcasted. As a result, everyone knows that this key is connected to the owner's identity. Therefore, it is referred to as the public key.

The other key will be held privately. It is very important that nobody except the owner knows the content of this key, as it will be used to prove the owner's identity. It is therefore referred to as the private key.

Because no cypher has to be exchanged, there is no need for the trusted middleman to exchange the cypher, eliminating a key weakness of cypher encryption.

2.1.2 Uses: encryption and source validation

A first use case of public key cryptography is encryption. If a person called Alice wants to send a message to Bob, she encrypts her message with Bob's public key. As a result, only Bob can decrypt this message as only Bob knows the corresponding private key. The message can now be send over any insecure network, such as the internet, email or even carrier pigeon, without any danger of leaking its contents.

Besides encryption, public key cryptography can also be used for source validation. If Alice wants to broadcast a message to the world, she can encrypt the message with her private key. As the encrypted message can only be decrypted with Alice's public key, everyone knows that only Alice could have encrypted it. This case is often referred to as a digital signature, as Alice signed the message with her private key.

For full security, a secure message is often first encrypted with the private key of the sender and then the public key of the recipient. Now only the recipient can decrypt the message and can then validate the origin of the message. This full process is modeled in figure 2.1.

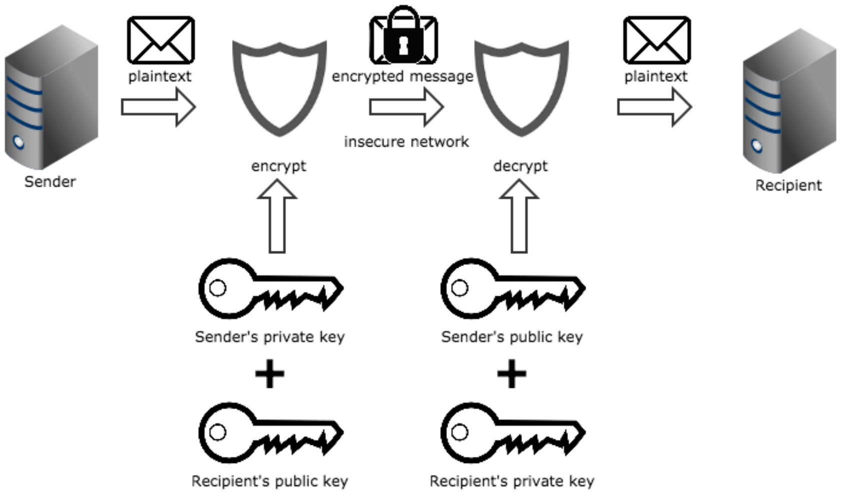

Figure 2.1

2.2 Blockchain

2.2.1 Definition of a blockchain

In the last few years, the blockchain as a concept has been given a wide range in meanings and definitions: some explain the blockchain as a history of events, some focus on blockchain as a bitcoin transaction ledger and some talk about the open decentralized database aspect of blockchains [5]. Therefore, we will start out this paper by thoroughly defining this concept, to avoid any misunderstanding or confusion further on.

Antonopoulos (2014) [1] defines the blockchain as follows: “The blockchain data structure is an ordered, back-linked list of blocks and transactions.” (p. 159). There are several aspects in this definition that need further explanation.

Back-linked lists as data structure

First of all, we are talking about a data structure: “A specialized format for organizing and storing data” [6]. There are several types of data structures such as arrays, lists, graphs, trees, etc.

Next, a blockchain is a specialized version of an “ordered, back-linked list”. A list is data structure that combines a number of ordered values (Abelson 1996 [7]). Incidentally, the “ordered” in the definition provided by Antonopoulos is superfluous, as by definition every list is ordered. An unordered list would no longer be a list, but would be a set instead. A blockchain will use back-linking to preserve the order of the values. This means that all values have a reference to the previous value in the list. For a blockchain, this means that every block in the blockchain contains an identification of the previous block. We will go further into hashes in a next section.

Transactions

Antonopoulos (2014) [1] defines transactions as: “data structures that encode the transfer of value between participants in the bitcoin system” (p. 109). Although it gives a good idea of what transactions are, even in the context of the bitcoin blockchain this definition is complete: despite most transactions being value transactions between participants, a transaction can also store data on a blockchain. Even the bitcoin blockchain allows non-monetary transactions to be included.

Therefore, we can conclude that there are two types of transactions: monetary transactions and non-monetary transactions. In a monetary transaction, there is a transfer of value (eg. and amount of bitcoin) between participants. It must be noted that this does not have to be an instant transfer. Several blockchain protocols allow for a custom condition to be scripted into the transaction.

On the other side, there are non-monetary transactions that store data onto the blockchain. An example of a common used non-monetary transaction is the encoding of a document on the blockchain.

Transactions are the raison d'être of the blockchain. All aspects that we will discuss in this chapter exist to make sure that transactions can be validated, propagated over the network and added to a distributed ledger.

Blocks

Antonopoulos (2014) [1] defines a block as: "a container data structure that aggregates transactions for inclusion in the public ledger, the blockchain." (p. 160). A block is literally a wrapper for multiple transactions to make it possible for them to be included into the blockchain. These blocks then form the ordered values of the back-linked list addressed in the beginning of this section, each linking to the previous block in the list.

To conclude, the blockchain is a list of blocks, with each block containing several transactions.

2.2.2 Properties of a blockchain

Securing the immutability of block headers

As the Bitfury Group explains [8], every implementation of a blockchain must secure itself against possible attacks on its immutability. Take for example a blockchain powering a cryptocurrency. If the immutability of this blockchain was not ensured, an attack could spend a coin and afterwards transmit a version of the blockchain without this transaction. As a result, the attacker changed the blockchain to a new reality in which this coin has not been spent, allowing the attacker to spend this coin again.

To ensure this immutability, blockchains make use of a consensus mechanism, an algorithm the different entities in the network use to agree upon the newest state of the blockchain. Every consensus mechanism is designed in such a way that no one malicious entity can break the system unless this entity has some sort of majority in the network. How this majority is defined, depends on the consensus mechanism.

Proof of work To suggest a new block for a blockchain with a proof of work consensus mechanism, an entity needs to solve a random computational problem. This problem is designed as such that this entity, called the miner, has a chance of p% to solve this problem if he controls p% of the computing power currently in the network. If the miner solves the problem, the miner will be rewards with the relevant cryptocurrency. Consequently, miners are incentivized to provide as much computing power as possible until it is not longer economically interesting regarding the reward. As the Bitfury Group correctly states [8]: "[the] security of the network is supported by physically scarce resources: specialized hardware needed to run computations, and electricity spent to power the hardware." (p.2)

If a malicious entity wants to manipulate a proof of work blockchain, this entity would need to control 50% of the computing power on the network. Therefore, the security of a proof of work blockchain depends on two main factors. First is the total computing power in the network, also called the hash rate. A higher the hash rate means it is more difficult for a new entrant to suddenly take over the network. Second is the distribution of the computing power in the network. If a couple of entities hold a high percentage of the total computing power in the network, they could effectively work together achieve a majority share in the network.

The main advantage of a proof of work system is its security, as it is very costly to attack. However, this comes at a steep ecological cost as a lot of electricity and computing hardware is wasted on useless calculations. Therefore, other consensus mechanisms have been suggested.

Proof of stake The difficulty to create a new valid block for a blockchain with a proof of stake consensus mechanism will depend on the balance an entity holds of the relevant cryptocurrency. An entity with a higher balance will more easily find a valid block, without having to spend electricity and computing hardware required by the proof of work algorithm [8].

The main idea is that entities with a higher balance, are more invested in the currency and its success. If one entity would control more than 50% of money available, it would be against his own interest to attack the blockchain as he would be hit the hardest.

The advantage of this system is the absence of wasteful spending. Instead of having to spend $1000 on specialized mining hardware, you can invest this money in the cryptocurrency itself with the same effect. However, there are several problems with the proof of stake mechanism. It is out of scope of this paper to discuss these issues individually, but all result out of the fact that the "proof of stake consensus is not anchored in the physical world (cf. with hashing equipment in proof of work)." (Bitfury Group, 2015, p22).

The most valuable cryptocurrency using some variation of proof of stake mechanism is Bitshares, with a market capitalization of $300 million on July 20, 2017 [14], making it the 14th most valuable cryptocurrency at the time of writing. This shows how the market trusts the proof of work mechanism better. Furthermore, Bitshares' market capitalization was only $20 million in April 2017 and reached a record of $1.2 billion on June 10, 2017. This illustrates further the extreme volatility of cryptocurrencies.

Other consensus mechanisms Several other consensus mechanisms have been proposed, besides proof of work and proof of stake. An example is proof of storage, in which an entity will be able to create new blocks depending on how many megabytes of storage he can provide the network [21]. However, as most of these other consensus mechanisms have not been implemented in successful blockchains, it is very hard to evaluate their viability.

Access to the blockchain data

As explained by the Bitfury Group [10], there are three levels of access to a blockchain. Firstly, there is read access to the data stored on the blockchain. This is the lowest level of access to a blockchain. Secondly, on a higher level, there is access to propose new transactions to the blockchain. Finally, on the highest level, there is access to create new blocks of transactions and add these blocks to the blockchain.

Depending on the design of a blockchain, different entities can have different levels of access to this blockchain.

Read access and and transaction submission: private and public blockchains The first two access levels are intertwined and will be discussed together.

As mentioned by the Bitfury Group [10], there are two options on these access levels. The first option is a public blockchain. This means that there are no restrictions an reading the data on the blockchain or on submitting transactions. The other option is a private blockchain. Only a predefined list of entities can then directly access the data or submit transactions.

Although a private blockchain might seem more secure because of the controlled access, the Bitfury Group [10] shows that several of the advantages of using a blockchain are lost. First of all, several clients will not have direct access to the blockchain data, they will have to rely on a node with direct access. This works against the decentralized aspect of a blockchain as there are only a limited amount of possible nodes to connect to. Next, as clients need to connect to a node with full access, not only do they have to trust that node, the interaction with that node becomes a vulnerability as well. For example, a 'man in the middle attack' is possible. This means that when a user tries to interact with the blockchain, a malicious party could intercept this communication and reply with false information [11]. Finally, only a limited set of computers has access to the transactions on the blockchain, creating the possibility of a human factor intervening in the blockchain operation. This circumvents the reliance of the algorithmically enforced rules of a blockchain.

On the other hand, data on public blockchains can be properly encrypted. This will provide sufficient security and privacy in most use cases.

Access to transaction processing: permissioned and permissionless For the third level, the developer of the blockchain has to chose who is allowed to process transactions that get incorporated into the blockchain. The first option is a permissionless blockchain. On a permissionless blockchain, there are no restrictions on the identities of the transaction processors. The opposite is a permissioned blockchain. Only a predefined list of entities can process transactions.

The use of a permissioned, public blockchain can be a necessary compromise in situations where compliance is an issue, such as in the financial sector.

2.2.3 Examples of blockchains

Bitcoin

The best known implementation of blockchain technology is in the bitcoin cryptocurrency. Bitcoin is a cryptocurrency created by Satoshi Nakamoto in 2008 [15]. In June 2017, bitcoin hit a record market capitalization of $48.8 billion [13]. It is by far the most popular and best established cryptocurrency currently. The drawback of this popularity is the difficulty to change flaws in the system, resulting in crises such as the current block size crisis [16].

The bitcoin blockchain is an public, permissionless blockchain with a proof of work consensus mechanism. This means that everyone has read access to the data on this blockchain, everyone can propose transactions to be added to the blockchain and everyone can start a node to help accepting proposed transactions.

The bitcoin blockchain uses a basic scripting language to power its transaction processing, only advanced enough to make a transactions and store short pieces of data [1]. On one side, the simplicity of this language could be seen as a limitation of its potential. On the other hand, is it a feature of the bitcoin blockchain making it more secure.

The bitcoin blockchain is interesting because of its extreme security. In July 2017, the bitcoin blockchain mining network has a hash rate of around 6.6 exa hashes per second [12]. This means that the bitcoin network is being supported by the equivalent computing power of 1.9 trillion Macbook Pro's from 2014. This shows the security of the bitcoin network. To break the system, one entity would need to suddenly control half of this computing power, which is quite unlikely. The only way would be to develop a supercomputer that is as a more powerful than all the computers ever made, such as a 50 qubits quantum computer [9].

Because the bitcoin mining is so strong, it is often used by other cryptocurrencies and other blockchains to help secure their own network. Namecoin is such an example [17].

Ethereum

Ethereum is another cryptocurrency powered by a blockchain, and the second most valuable cryptocurrency with a market capitalization of $36.2 billion in June 2017. While bitcoin focuses on creating a secure, non-nonsense currency, ethereum tries to leverage the entire potential of the blockchain. This creates endless possibilities but makes ethereum also more complex than bitcoin.

Just like the bitcoin blockchain, the ethereum blockchain is currently public and permissionless with a proof of work consensus mechanism. However, there are plans to move to a proof of stake approach in the future [20].

As explained in the Ethereum white paper [18], the programming language powering the ethereum blockchain, called Solidity, is Turing complete. Practically, this means that any conceivable program could be written in this language. This creates the possibility for smart contracts. These are transactions that will be executed automatically only if a certain condition is met. An example is a crowdfunding campaign that only transmits the raised budget to the beneficiary once the threshold is met before a certain deadline.

Because of its wide potential, the source code of the ethereum blockchain is often copied (called 'forked'), adjusted and redeployed for custom blockchain solutions.

2.3 Distributed databases

2.3.1 Distributed, decentralized and centralized databases

To fully understand distributed databases, it is important to distinguish them from decentralized and centralized databases. Baran (1964) [22] explains the difference in the context of networks. A centralized network makes all nodes connect to one central node. A distributed network, often referred to as a mesh, is the opposite in which different Figure 2.2 visualizes the differences.

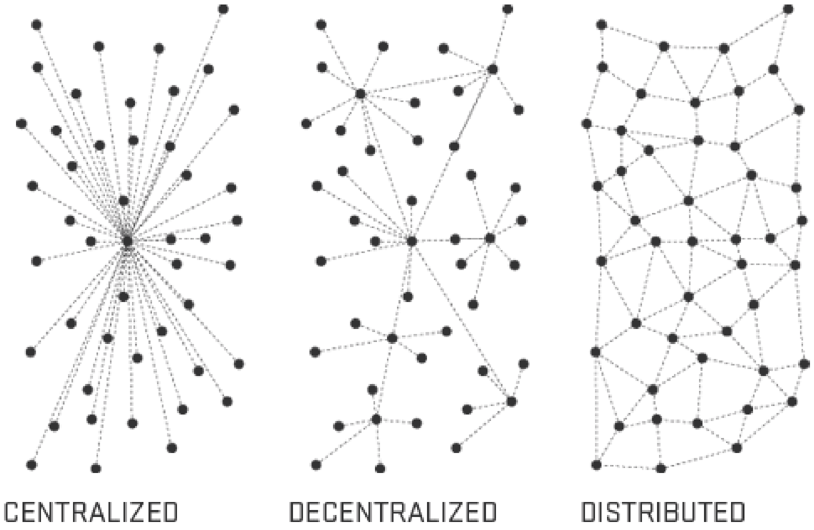

Figure 2.2: Baran, P. (1964) Centralized, decentralized and distributed networks.

2.3.2 Advantages of using a blockchain as a distributed database

2.3.3 Examples of distributed databases and networks

Git

BitTorrent

Non-persistent P2P networks

Chapter 3

Aim of research

Chapter 4

Proposed solutions

4.1 Storing onto the bitcoin blockchain

4.2 Blockchain distributed storage

4.3 Timestamping a database

4.4 Applied example: replacing notary will services

Chapter 5

Conclusion

Bibliography

[1] Antonopoulos, A. M. (2014). Mastering Bitcoin. Sebastopol, CA.

[2] Cipher, C., & Cipher, M. (2004). Introduction to cryptography. EEC, 484, 584.

[3] Rivest, R.L. and Shamir, A. and Adleman, L.M. US Patent 4,405,829 issued in 1983

[4] Calderbank, M. (2007). The RSA Cryptosystem: History, Algorithm, Primes.

[5] Bischoff, P. (2017). What is Blockchain? 10 experts attempt to explain it in 150 words or less. Comparitech. Retrieved 20 July 2017, from https://www.comparitech.com/blog/information-security/what-is-blockchain-experts-explain/

[6] What is data structure? - Definition from WhatIs.com. (February 2006). SearchSQLServer. Retrieved 20 July 2017, from http://searchsqlserver.techtarget.com/definition/data-structure

[7] Abelson, H. et al. (1996). Structure and Interpretation of Computer Programs. MIT Press.

[8] Bitfury Group (2015). Proof of Stake vs Proof of Work

[9] Bauer, M. R. (2017, April 14). Quantum Computing is going commercial with the potential to disrupt everything. Retrieved July 19, 2017, from http://www.newsweek.com/2017/04/21/quantum-computing-ibm-580751.html

[10] Bitfury Group (2015). Public versus Private blockchains. Part 1: Permissioned Blockchains.

[11] Desmedt, Y. (2011). Man-in-the-middle attack. In Encyclopedia of cryptography and security (pp. 759-759). Springer US.

[12] Hash Rate - Blockchain. (n.d.). Retrieved July 19, 2017, from https://blockchain.info/charts/hash-rate

[13] Bitcoin (BTC) price, charts, market cap, and other metrics — CoinMarketCap. (n.d.). Retrieved July 19, 2017, from https://coinmarketcap.com/currencies/bitcoin

[14] CryptoCurrency Market Capitalizations. (2017). Coinmarketcap.com. Retrieved 20 July 2017, from https://coinmarketcap.com/currencies/

[15] Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.

[16] Chen, L. Y., & Nakamura, Y. (2017, July 10). Bitcoin Is Having a Civil War Right as It Enters a Critical Month. Retrieved July 19, 2017, from https://www.bloomberg.com/news/articles/2017-07-10/bitcoin-risks-splintering-as-civil-war-enters-critical-month

[17] Loibl, A. (2014). Namecoin. namecoin. info.

[18] Buterin, V. (2013). Ethereum white paper.

[19] Kepser, S. (2004, August). A Simple Proof for the Turing-Completeness of XSLT and XQuery. In Extreme Markup Languages. Chicago

[20] Buterin, V. et al. (2017). Proof of Stake FAQ. URL https://github.com/ethereum/wiki/wiki/Proof-of-Stake-FAQ

[21] Alternatives for Proof of Work, Part 2: Proof of Activity, Proof of Burn, Proof of Capacity, and Byzantine Generals Bytecoin Blog. (2017). Bytecoin: private secure financial system. Retrieved 19 July 2017, from https://bytecoin.org/blog/proof-of-activity-proof-of-burn-proof-of-capacity/

[22] Baran, P. (1964). On distributed communications networks. IEEE transactions on Communications Systems, 12(1), 1-9.

Appendix A

A.1 Appendix

FACULTY OF BUSINESS AND ECONOMICS
Naamsestraat 69 bus 3500
3000 LEUVEN, BELGIË
tel. + 32 16 32 66 12
fax + 32 16 32 67 91
info@econ.kuleuven.be
www.econ.kuleuven.be