UNIVERSITÀ DEGLI STUDI DI TRENTO

Department of Information Engineering And Computer Science

Bachelor degree in
Computer Science

Final Dissertation

Titolo

Sottotitolo (alcune volte lungo - opzionale)

Supervisor
Agostinelli Claudio

Candidate/Student
Vasile Adrian Bogdan Pop

Accademic Year 2018/2019

Ringraziamenti

...thanks to...

Indice

Abstract

The thesis tries to explain what a blockchain is, how it works and why it seems to have so much potential. In particular, this debate focuses on the different purposes of this technology with its functionalities and its implementations by analyzing two specific blockchains: Bitcoin (BTC) and Bitcoin Cash (BCH). The thesis also exposes which are the requirements a blockchain should have when it is used as a coin, in order to be globally handled as the leading currency. Hence, it goes deeper into the differences between the BTC and BCH analyzing the distinct approaches to satisfy the demands of a cryptocurrency. Especially, It illustrates the main reasons why the Bitcoin hard fork was reached on November 15, 2018, giving life to the Bitcoin Cash blockchain. Once the behavior of this new technology is clear, the thesis goes further with the integration of BCH in the programming language "R". In this way, it provides a mean to work with the data stored into the Bitcoin Cash blockchain and retrieve interesting statistics out of it.

In other words, the steps to achieve the goal for this thesis will be:

- Understanding and analyzing blockchain technologies
- Studying the two most important Bitcoin blockchains BTC - Bitcoin and BCH - Bitcoin Cash
- Design a new Package in R for data reading and manipulation of the Bitcoin Cash blockchain
- Implementation of the Package in R with a description of all the possible commands developed
- Sample of the Package functionalities and its potential

1 Blockchain

The application of the blockchain in multiple fields is gaining more and more traction today, until to be considered by some as the new Internet. But actually, the main concept behind this revolutionary idea was described as early as 1991 in an accademic paper[1] published by Stuart Haver and W. Scott Stornetta. With their solution, the two scientists aim to Time-Stamp documents, in order to certify when a document was created of last changed. The system used a cryptographically secured chain of blocks to store the time-stamped documents and in 1992 Merkle trees were incorporated to the design, making it more efficient by allowing several documents to be collected into one block.[2] However, this is not enought. Indeed, to have a fully working blockchain we need to add more components.

1.1 Distributed ledger

As said in a report from the UK Government, "A distributed ledger is essentially an asset database that can be shared across a network of multiple sites, geographies or institutions. All participants within a network can have their own identical copy of the ledger"[7]. Due to this technology, there aren't either central administrators or centralized data storages. To have this mechanism working, we need both a Peer-to-Peer Network[2] and a Consensus Algorithm 1.2.

1.2 Consensus Algorithms

To clarify the benefits of this type of algorithms we must start with a dilemma as known as "The Byzantine Generals Problem", in whom, each general has its own army and that each group is situated in different locations around the city they intend to attack. The generals need to agree on either attacking or retreating. It does not matter whether they attack or retreat, as long as all generals reach consensus, i.e., agree on a common decision in order to execute it in coordination. Therefore, we may consider the following requirements:

- Each general has to decide: attack or retreat (yes or no)
- After the decision is made, it cannot be changed
- All generals have to agree on the same decision and execute it in a synchronized manner

The aforementioned communication problems are related to the fact that one general is only able to communicate with another through messages, which are forwarded by a courier. Consequently, the central challenge of the Byzantine Generals' Problem is that the messages can get somehow delayed, destroyed or lost.

In addition, even if a message is successfully delivered, one or more generals may choose (for whatever reason) to act maliciously and send a fraudulent message to confuse the other generals, leading to a total failure.

If we apply the dilemma to the context of blockchains, each general represents a network node, and the nodes need to reach consensus on the current state of the system. Putting in another way, the majority of participants within a distributed network have to agree and execute the same action in order to avoid complete failure.[8][2] There are various implemetation of the mechanism through which a blockchain network reach consensus.

[1] `https://www.anf.es/pdf/Haber_Stornetta.pdf`

[2] the peers are computer systems connected to each other via the Internet, the data can be shared directly between systems on the network.

1.2.1 Proof of Work

Due to Bitcoin 2, the most known is the Proof of Work (PoW) algorithm, in whom is utilized for block generation. The process of generating correct proofs in order to add a block to the blockchain is known as “mining” and the individuals that participate in the mining process are known as “miners”.[1]

The PoW mining involves numerous hashing attempts, so more computational power means more trials per second. In other words, miners with a high hash rate have better chances to find a valid solution for the next block. The PoW consensus algorithm makes sure that miners are only able to validate a new block and add it to the blockchain if the distributed nodes of the network reach consensus and agree that the block hash provided by the miner is a valid proof of work.[2]

1.2.2 Proof of Stake

The Proof of Stake (PoS) consensus algorthm is relatively different and a new way to generate and append blocks in a blockchain. It was developed in 2011 as alternative to PoW, which requirements a massive amout of energy to make it work. In fact, according to the bitcoin energy consumption tracker at Digiconomist[3], bitcoin currently consumes 66.7 terawatt-hours per year. That's comparable to the total energy consumption of the Czech Republic, a country of 10.6 million people.

Basically, the mining process is replaced with a mechanism where blocks are validated according to the stake of the participants, also known as "validadors". The criteria used to choose the validator depends on the proof of stake system but mainly, the choiche is based on the economic stake in the network of each node.[2][1]

1.2.3 Proof of Burn

Also the Proof of Burn (PoB) algortigm is an approach to avoid the massive waste of energy used for hashing. Indeed , with this technique, the mining process is replaced with a greener one. Where, the "miners" of the PoB coins will send coins to an uspendable address, known as "eater address". Even this transactions are recorded on the blockchain ensuring that the coin cannot be spent again. The principle behind this consensus algorithm is that the user burning the cryptocurrency is showing long-term commitment to the coin by burnin it. This is because they are taking a short-term loss in exchange for a long-term gain.[1]

1.3 Advantages and Disadvantages

As a result for its complexity, blockchain's potential as a distributed form of record-keeping is almost without limit. But as every other technology, even the blockchain have its pros and cons.

1.3.1 Advantages

Trustless system and small fees In most traditional systems, a consumer pays a bank to verify a transaction, a notary to sign a documents, or a minister to perform a marriage. Using blockchain technologies, this is no longer necessary because the distributed nodes verify the transactions through consensus algorithms.
Therefore, blockchain systems negate the risk of trusting a single person or organization and reduces the overall costs and transaction fees by cutting third-parties. For this reason, blockchain is often referred to as a "trustless" system.[2][5]

Improved accuracy by removing human involvement in verification

Transactions are secure, private and efficient Also in this case, by eliminating the human factor, we obtain a 24 hours a day working system, instead of institutions operating only during business hours. As a result, transactions can be completed in about ten minutes and can be considered secure after just a few hours. Actually, once a transaction is recorded, its authenticity must be verified by the blockchain network. Where, millions of nodes rush to confirm that the details of the transaction are correct.

Moreover, every block of the chain has the hash of the block before it along with its own hash, wich is computed with the data contained in the block. So, when an information in a block is edited in

[3] `https://digiconomist.net/bitcoin-energy-consumption`

any way, its own hash will change, but the hash in the block after it would not. This contrast makes extremely hard to change data inside the blockchain without beeing noticed.

Considering that the majority of blockchains are pubblic, it means that anyone can access the blockchain and read the data inside it with all the transaction history. So, if anyone can read see all the transactions, why is people say that blockchain transactions are anonymous?
That's why anonymity is confused with confidentiality. Indeed, every user makes public transactions with their unique code called "public key"[4], which is recorded on the blockchain, rather than their personal information.[5]

Transparency A big benefit to this technology is given by his open source nature. That means that the users of the blockchain network can modify the code as they see fit, so long as they have the approval from the majority of the network's nodes.

Distributed Blockchain does not store any of its information in a unique node or central location. Instead, data is stored in thousands of devices on a distributed network. Whenever a block is added to the chain, every node updates its blockchain to reflect the change. Because of this, there is no single point of failure and so becomes more difficult to tamper with.[2][5]

Stability Once a block is confirmed, its extremely difficult to remove or change the data inside it. This makes blockchain a great technology for storing financial records or any other data where an audit trail is required because every change is tracked and permanently recorded on a distributed and public ledger.[2]

1.3.2 Disadvantages

While there are significant upsides to the blockchain, there are also serious challenges to its adoption. The most ones are political and regulatory, but there are also some technical challages to take in account.

Energy inefficiency Some blockchains, especially those using Proof of Work consensus algorthms 1.2.1, are higly inefficient. Since mining is higly competitive and there is only one winner every ten minutes, the work of every other miner is wasted. As miners are continually trying to increase their computational power, so they have a greater chance of finding a valid block hash, the resources used by the Bitcoin network has increased significantly in the last few years, and it currently consumes more energy than many countries, such as Denmark, Ireland, and Nigeria.[2]

Scalability The Bitcoin is a perfect case study for the low scalability of blockchains. In fact, to add a new block on the blockchain, it takes about ten minutes and the maximum size of each block is of 2 MegaBytes. This entails that there is a limited number of transaction per block. It is estimated that the Bitcoin's blockchain can manage only seven transactions per second (TPS). Meanwhile, the Visa's actual infrastructure can process 24.000 TPS.

51% Aattacks The consensus algorithms that protetc the blockchain has proven to be very efficient over the years (thanks to). However, there are a few potential attacks that can be performed against blockchain networks. The most discussed is the 51% attack. Such an attack may happen if one entity manages to control more than 50% of the network, which would allow them to edit or modify the order of the transactions. Altought being theoretically possible, there was never a succesful 51% attack on the Bitcoin blockchain, additionaly, as the network gets larger, the security increases.[2]

Private keys Blockchain uses public-key cryptography to give users ownership over their blockchain data. Each public-key correspons to a private-key which should be kept secret. Users need their private-key to access their data. If a user loses their private key, the money is effectively lost, and there is nothing they can do about it.[2]

[4]Public-key cryptography, is a cryptographic system that uses pairs of keys: public keys which may be disseminated widely, and private keys which are known only to the owner.

Storage Distributed ledgers can increase very much over time. For example the Bitcoin blockchain currently requires 239[5] GigaByte of storage. Even if in the Bitcoin paper its said that thanks to "Moore's Law"[9] the storage should not be a problem, the current growth in blockchain size appears to be outstripping the growth in hard drives and the network risks losing nodes if the ledger becomes too large for individuals to download and store.[2]

1.4 Use Cases

The merit to the great attention given to the blockchain technology is to be attributed to Bitcoin. Indeed, thanks to this digital currencies (cryptocurrencies) the blockchain-based solutions have grown. So learning how this innovative technology can be applied to different scenarious is very important.

1.4.1 Digital Identity

The identity component in a blockchain is fulfilled through the use of cryptographic keys. Combining a public and private key creates a strong digital identity reference based on possession of informations or certifications. The public key, is distributed publicly and it is what identifies the identity, the private key is kept secret by the identity and it is used to express consent to digital interactions and to sign the data to record in the block.[4]

1.4.2 Charity

The number of charitable organizations which addopted cryptocurrencies as a donation method is increasing more and more. The reason is because crypto's transparency, in fact, thanks to this characteristic, every donation can be tracked and verified. So, the higher level of transparency and public accountability can ease donors' minds and encourage them to give while also reinforcing the charity's reputation for integrity. Moreover, blockchain technology reduces the fees and taxes to manage every donation by reducing the number of required intermediaries.[2]

1.4.3 Healthcare

Blockchain may offer significant benefits to hospitals in terms of security, interoperability and transparency. Unlike traditional databases that rely on a centralized server, the use of a distributed system allows for data exchange with higher levels of security. In addition, distributed ledgers increase the interoperability among clinics, hospitals, and other health service providers by allowing all theese parties to interface with a unique storage system.

Furthermore, blockchain systems may also give higher levels of accesibility and transparency over their own health information. But also, allows pharmaceutical organizations to increase the efficiency of theri infrastructure by cutting down on the widespread problem of drug counterfeiting.[2]

1.4.4 Governance

The governance can be greatly improven in various sectors by the blockchain technologies. A short example can be the efficiency of taxes, ideed, because governments rely on taxpayer funds, it's important that they use their budgets wisely. Blockchain systems and smart contracts can be employed to automate tasks and workflows, which would greatly reduce time and money spent on bureaucratic processes.[2]

Another application can be the election process. With the support of a workign digital identity system, and the high level of immutability given by the blockchain, this technology could be an excellent solution for ensuring that votes cannot be tampered with. Creating the possibility to transform the secure online voting into a reality.

1.4.5 Payments

Thanks to Bitcoin, this is the most widespread use case of the blockchain. It starts in 2008 when Satoshi Nakamoto published the Bitcoin's white paper, where, for the first time, all the above components of a blockchain were used to solve the double-spendign problem.

[5]Last update: 10/9/2019

2 Bitcoin

In a few words, Bitcoin is a digital form of the actual cash. But in contrast to the traditional fiat currency, which relies on a central bank controlling it, this new form of currency relies on blockchain technology 1 and cryptography[1]. As a result, this new digital cash is mainly called "cryptocurrency".

2.1 Infrastructure

Bitcoin, as the majority of cryptos, is based on a public permissionless blockchain architecture. Which means that anyone can join, read, write and commit, plus it is hosted on public servers, is anonymous and high resilient.
The Bitcoin's mechanism to reach consensus is the Proof of Work algorithm 1.2.1. Due to the implementation described by Satoshi Nakamoto in the white paper "The proof-of-work involves scanning for a value that when hashed, such as with SHA-256, the hash begins with a number of zero bits. The average work required is exponential in the number of zero bits required and can be verified by executing a single hash. For our timestamp network, we implement the proof-of-work by incrementing a nonce in the block until a value is found that gives the block's hash the required zero bits. Once the CPU effort has been expended to make it satisfy the proof-of-work, the block cannot be changed without redoing the work. As later blocks are chained after it, the work to change the block would include redoing all the blocks after it"[9]. It follows that the difficulty required to close a block is flexible. In fact, it changes every 2016 blocks suiting the computational power of the network. Considerng that, at the desired rate of one block each 10 minutes, 2016 blocks mean that the difficulty is adjusted every two weeks.

2.1.1 Incentive

To have a working PoW, miners have to spend their CPU time and electricity. So, why should they allocate their resources in Bitcoin?

The answer is: the incentive. Indeed, the first transaction of every block is a special transaction which gives a new amount of currency to the creator of the block. This, not only is a good incentive to support the network but also provides a way to initially distribute coins into circulation, since there is no central authority to issue them. However, this process is destinated to end, because there is a predetermined number of coins (21 million) which can spread. To ensure that, every 210.000 blocks (corresponding to 4 years), the rewarding given to the creator of the block is halved. This leads to the second type of incentive in the Bitcoin network: transaction fees. In fact, for every transaction there is also a fee amount, which is given to the creator of the block as a reward for its effort.

2.2 Transactions

To figure out how a transaction works, before it is necessary to understend the concept owning a coin. Because, unlike the fiat currency, where you own a physical coin to spend, owning a bitcoin means that you own the key to unlock the coins from a past transaction and use them in the new transaction. To get a better sense of a the transaction mechanism, it is recomended to read a brief extract of the Bitcoin paper:

"We define an electronic coin as a chain of digital signatures. Each owner transfers the coin to the next by digitally signing a hash of the previous transaction and the public key of the next owner and adding these to the end of the coin. A payee can verify the signatures to verify the chain of ownership"[9]

[1] In computer science, cryptography refers to secure information and communication techniques derived from mathematical concepts and a set of rule-based calculations called algorithms to transform messages in ways that are hard to decipher.

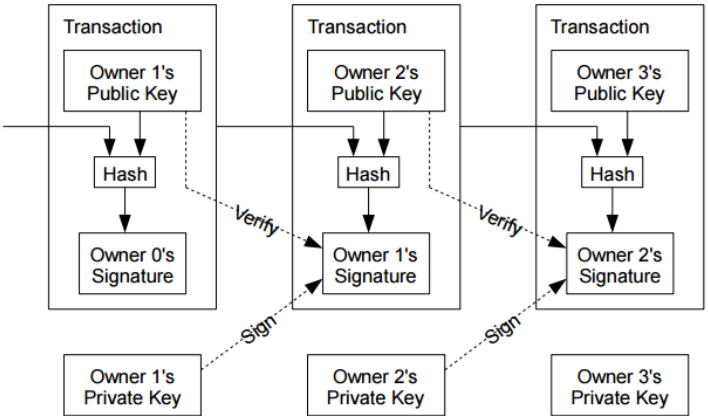

Figura 2.1: Transaction schema. From [9]

To go deeper into the design of a transaction, it is usfull to imaginate a double-entry ledger. In few words, every transaction have one or more "inputs"[2] and one or more "outputs"[3]. Usually outputs add up to slightly less than inputs, because the difference represents an implied "transaction fee" 2.1.1. Higher the fee, higher the probability for the transaction to be included in the next block to mine.

Each output, before is spend as an input in a new transaction, waits as an "Unspent Transaction Output" (UTXO). The sum of all the owned UTXOs, compose the total balance of a wallet 2.3.
But, not every transaction in the blockchain produces an UTXO.
Indeed, it depends on the scriptPubKey[4] used into the specific transaction.

Pay To Public Key Hash (P2PKH) Is the most common form of scriptPubKey used to send a transaction to one or multiple addresses. In this type of scriptPubKey the coins sended can be unlocked by a single private key corresponding to the public key hashed in the transaction itself. It is possible to recognize it by looking at the first character in the address used for the output. In fact, if it starts with the number "1" it is P2PKH.[3]

Pay To Script Hash (P2SH) Was implemented in 2012 to add more flexibility to transactions by adding a hash of a second script, the redeem script[5]. It can be used to process multisignature transactions, Open Assets Protocol[6] and to store textual data on the blockchain. The P2SH can be recognize by the number "3" at the begin of the addresses.[3]

Multisig Special addresses where to spend a transaction, multiple private keys signature are required. Although P2SH multisig is now generally used for multisig transactions, this base script can be used to require multiple signatures before a UTXO can be spent. In multisig pubkey scripts, called m-of-n, m is the minimum number of signatures which must match a public key; n is the number of public keys being provided.[3]

Null Data This type ads arbitrary data to a probably unspendable scriptPubKey that nodes don't have to store in their UTXOs database. Consesus rules allow null data outputs up to the maximum allowed scriptPubKey size of 10.000 bytes provided.[3]

[2] debits against a bitcoin account.
[3] credits added to a bitcoin account.
[4] A script included in outputs which sets the conditions that must be fulfilled for those satoshis to be spent. Data for fulfilling the conditions can be provided in a signature script (another type of data generated by the spender).
[5] A script similar in function to a scriptPubKey. One copy of it is hashed to create a P2SH address (used in an actual scriptPubKey) and another copy is placed in the spending signature script to enforce its conditions.
[6] Used to create new currencies upon the Bitcoin blockchain. Called colored coins

2.3 Wallets

In simple works, a wallet is the mean through wich is possible to manage all the owned cryptocurrency. But it can be splitted in two components, programs and files. Wallet programs create the publick keys to receive the Bitcoin and uses the corresponding private key to spend them.
Wallet files, instead, stores the private keys in a file or physically on pieces of paper. Furthermore, it also gives the process used for the public-private key creation and usage, and an approach to a deterministic hierarchy key creation process in an unlinkable keys manner.

3 Bitcoin Cash

"Imagine you have a dollar in your wallet, and you put it aside for a while and then realise that one coin has split in two. Sounds weird? Not in the cryptocurrency world." That is an analogy with the fiat currency reported in an article of Shouth China Morning Post[6].
Bitcoin is like a software, but, due to its distributed nature, unlike all the software we know, there isn't a single entity who determines how it should be updated.

3.1 Fork

3.1.1 SegWit

3.1.2 Addresses

3.1.3 Block Size

3.1.4 Block mining

4 Package implementation in R

4.1 Environment

4.2 Feasibility study

4.3 Implementation

4.3.1 Commands

4.4 Usage samples

Bibliografia

[1] https://medium.com/coinbundle/consensus-algorithms-dfa4f355259d.

[2] Binance Academy. https://www.binance.vision/blockchain.

[3] bitcoin. https://bitcoin.org/en/.

[4] Coindesk. https://www.coindesk.com/information.

[5] Investopedia. https://www.investopedia.com/terms/b/blockchain.asp.

[6] Shouth China Morning Post. https://www.scmp.com/tech/blockchain/article/2173389/bitcoin-cash-hard-fork-everything-you-need-know-about-latest.

[7] UK Government Office for Science. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/492972/gs-16-1-distributed-ledger-technology.pdf.

[8] Marshall Pease Leslie Lamport, Robert Shostak. *Distributed Systems: concepts and Design.* ACM New York, NY, USA, 1982.

[9] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. *www.bitcoin.org*, 2008.