Università degli Studi di Cagliari

**PHD DEGREE**

Electronic and Computer Engineering

Cycle XXX

# TITLE OF THE PHD THESIS

## The blockchain technology. Analysis and applications.

Scientific Disciplinary Sector(s)

ING-INF/05


PhD Student: Andrea Pinna

Coordinator of the PhD Programme Fabio Roli

Supervisor Giorgio Giacinto

Co-Supervisors Michele Marchesi



Final exam. Academic Year 2016 – 2017

Thesis defence: February-March 2018 Session

*Ph.D. in Electronic and Computer Engineering*
*Dept. of Electrical and Electronic Engineering*
*University of Cagliari*

# The blockchain technology. Analysis and applications.

Andrea Pinna

*Advisor*: Giorgio Giacinto
*Co-Advisors*: Michele Marchesi, Roberto Tonelli
*Curriculum*: ING-INF/05

Cycle XXX
March 2018

# Contents

# List of Figures

# List of Tables

# Preface

Throughout the history of humanity, learning a new technique or using a new technology has often led to profound social changes. In cultures without writing, knowledge is transmitted orally, with the disadvantage that the amount of transferable information is limited, and error propagation is inexorably high. The technique of writing has made it possible to accumulate knowledge and to untie it from time, giving everyone who reads the opportunity to access it, without the need to acquire knowledge directly. The writing, as well as pass on thoughts and works, has allowed to codify the laws regulating the life of a community and to establish property ownership. Finally, technology's and computer science progress, have elevated and dematerialized the writing, codifying the concept of information. Technology allows to do something new or to do differently and better than has been done before. Every novelty involves, among other things, social changes and different ways of interacting, that is, technology changes relationships between people.

The blockchain technology, the subject of this thesis, may turn some concepts that seemed now consolidated in computing. This is a promising technology, currently under development, which in many ways can be considered a milestone in the history of communication. The invention dates back to 2008, when a technical proposal was presented on the Internet in November, under the title "Bitcoin: A Peer-to-Peer Electronic Cash System". The author, the mysterious Satoshi Nakamoto, described, in nine pages, all the features of a new electronic payment system, Bitcoin, which does not need a central authority. The great revolution is precisely this: to create a decentralized, horizontal system, in which users do not have the classic guarantor or intermediate figure above the parties, but this is replaced by the blockchain technology.

Bitcoin is a free and open source decentralized system, and it is, above all, a trusting system. In fact, it allows two people who do not know each other to make a financial transaction without any guarantor and without requiring mutual trust. Unlike other digital payment systems (bank transfer, credit card, or other centralized systems, such as Paypal), Bitcoin does not have an owner. In addition, it is a transparent system in which all data is of public domain

and is stored in the data structure called blockchain, replicated in thousands of identical copies distributed all over the world. It is also true that no personal data revealing the identity of the users is recorded; all transactions are public and written in the blockchain, but users remain anonymous. Creating an anonymous account is overall very simple: just install a program that automatically allows to get an electronic address (or many addresses) in which to credit the bitcoin.

After a fairly long period of about five years, during which very few people were interested in this new technology, the number of users took off and Bitcoin proved to work very well. Users have created an ever-populated and active community. With the formation of discussion groups, many have become true experts. Some users have focused on enhancing anonymity, others in creating advanced services and opening electronic markets. Others have also aimed at capitalizing, buying and selling Bitcoin as a stock market.

New electronics companies were also born to produce the most efficient machines that are required to run the algorithm. Those who buy and operate machines to perform the controls and calculations needed to validate transactions are rewarded with a bitcoin prize. But the total number of bitcoins will not grow forever. The system has an upper limit. As the limit gets closer, new quantities will be getting harder to get. So, like for gold or diamonds, Bitcoin is a rare resource and can be the subject of trading deals that create and float its unitary value. In addition, based on the average operating costs, those who transact in Bitcoin may pay a voluntary bid. In a sense, people's free decisions take the place of politics of a central authority.

Looking at the effects of this technology from a broader perspective, it is possible to see the evolution towards a non-hierarchical society in which power is dematerialized and distributed among system's users. In fact, blockhain technology overthrows the concept of central bank and currency control: Bitcoin can be seen as an anti-state system, a sort of anarchy without geographical boundaries. Bitcoin shows that it is not always necessary to check everything from above to ensure "peaceful coexistence". But it is enough that certain rules are imperative. In this regard you may think that doing without guarantors and bureaucracies is socially advantageous. This prospective makes more difficult to leverage human weaknesses. In fact it obstacles the occurrence of corruption to obtain unjust privileges.

The freedom offered by the Bitcoin system has also a negative side. As mentioned, payments in Bitcoin are anonymous, as are payments with traditional cash. Malicious and cunning users succeed in hiding their tracks in the blockchain buying and selling illegal and criminal assets and services undisturbed. For example, the sale of drugs, computer attacks or other criminal acts, and so on can occur. All in all, this has always happened and still happens with traditional centralized money systems. Sometimes there have been thefts and

cheating, but in no case these are due to the inefficiencies of the algorithm. In fact, to "steal" bitcoin, you must have the secret codes (private keys) of the victims and this can only be done if someone does not properly guard them.

On the other hand, with the advent of Bitcoin and blockchain technology, the computer science and the economy overlook in a world yet to discover. Over the last five years, driven by widespread enthusiasm, several stakeholder groups have built their own currency and implemented their own blockchain. To date, there are almost nine hundred blockchain-based electronic coin implementations, and some of them represent an important evolution of Bitcoin. The case of Ethereum stands out. It is a system that, in addition to the electronic payment capabilities, allows the writing and execution of computer programs in a decentralized way. These programs are commonly called "smart contracts", and represent a digital and immutable bond that, upon the occurrence of the established conditions, perform a certain action. Many applications can be developed with the blockchain and smart contracts, such as fundraising, transparency of acts, certification mechanisms and collaborative consumption. These are automated and reliable and possible on a global scale.

The hope is that research on blockchain technology can bring benefits to humanity in the future, improving the quality of life of people from the less fortunate, and which can also help to valorise local realities, offering new services and improving existing ones. The blockchain technology may inspire us to follow the path to a horizontal society, where is no longer necessary to base society on social hierarchies. Where is natural and automatic to give everyone the same opportunities and rights for progress and common well being.

# Chapter 1

# Introduction

This thesis concerns the study of the blockchain-based cryptocurrencies, and can be framed in the field of the software engineering. In particular, it involves the study of the blockchain technology, the engineering aspects related to blockchain based software and services development, and several application scenarios development.

In the last few years, cryptocurrencies and the blockchain technology have interested the scientific community, that approached the topic from various angles. For example, there are studies on the cryptocurrency market evolution [65, 24, 123], on the forms of anonymity [41, 50, 63], on implementation of Smart Contracts [23, 72, 7, 36], on operative costs of Bitcoin [130, 129, 95], on the blockchain analysis [115, 70, 62],and on non-financial applications [96, 69, 23].

Blockchain based Cryptocurrencies are payment systems that do not need any centralized control to verify the money transfer reliability. In fact, transactions can be accepted if and only if they are consistent with the data inside the blockchain. The first implementation and to date the most important blockchain-based cryptocurrency is Bitcoin [89]. In principle, a blockchain is a shared database containing the historical data of transactions, and is available through a peer-to-peer Internet network. The blockchain evolves block by block. Some network participants execute the activity called Mining and generate blocks. They are called Miners.

A consensus algorithm, based on the super-majority principle, characterizes Mining. The algorithm allows only one new valid block generation by time. The majority of cryptocurrencies makes use of a mechanism based on the "proof-of-work", that is a computational effort that privileges those who have more hardware resources. The mechanism is associated with the systematic gathering and validation of pending transactions. The name Mining recalls the classic mining of resources, because with the Mining, the system generates and gives an amount of cryptocurrency to the participant who creates a new block. That

operation consists of the computation of a digital code of 256 bit length, produced using the Secure Hashing Algorithm (SHA). The digital code takes the name of Hash of the block and it has the typical properties of a hash code. It is deterministic, is preimage attack resistant (it is very hard obtaining the message starting from the code), and it's extremely sensitive to changes. The computation of a new block is a competitive operation and the number of attempts is proportional to the total computation power of the nodes of the network.

So the blockchain is a data structure that collects all the transactions held since the cryptourrency creation. Data are unchangeable because protected by a chain of consistence proofs. Information written in each block is unchangeable too, thanks the presence of a cryptographic code called Merkle Root, computed as a function of the transaction recorded in the block. Changing a single bit of the block data leads the loss of consistence and the loss of validity of the entire block. In reality, if a malevolent user takes the possession of the majority of network nodes, he could force the block validation process. This is very hard for two reasons. The first concerns the size of the network: the investment required would be very high. The second concerns the nature of the trust system: if a person or a pool of people can change the data to their liking, the system would lose credibility and the Bitcoin would lose its value.

But the blockchain is also the enabling technology which allows new business opportunities and the realization of several typologies of new services. Regarding this aspect, the release of the Ethereum, during the 2014, represents an important turning point. Ethereum is defined as blockchain based decentralized platform in which is possible to create and run programs called smart contracts. Smart contracts [125, 27] allow the realization of several intelligent transaction management systems, that includes found deposits, auctions, tradeable tokens and decentralized organizations. The creation of a Smart contract consists in sending to the blockchain the program code, written in a specific supported language.

To address all the facets and implications of blockchain technology, the research activity focuses on many aspects about it, and attempts to cover the various issues which are currently under the magnifying glass of the scientific community, especially in the field of the software engineering. These include metrics and analysing models of the data inside the blokchain, the sustainability evaluation (economic, social, and environmental) of cryptocurrency systems and blockchain-based software [67, 34], tools and practices to drive the blockchain software design and development[136, 48], the real development of blockchain software, and finnaly social aspects[124], the relationships between blockchain users, the networks of developers and their sentiment [86, 99, 38, 98].

The thesis stems from the exigence of having a broad understanding of this new technology, starting from the low level and arriving to the definition of

properly designed applications. It follows the path of a comprehensive study of the blockchain technology, and offers an overview of scientific contributions produced during the doctoral research activity. In order to become familiar with the world of the blockchain, the thesis explore some of the concepts about the blockchain technology, starting with the concept of cryptocurrency, and arriving to the application of smart contracts in real research projects. The first part of the thesis analyses the blockchain technology, proposing empirical methods and studies which provide a wide range of results. Studying the blockchain means studying the evolution of its dimension and of the distribution of wealth. And includes both an investigation of users' activity and an overview of efficiency problems in blockchain based cryptocurrency.

The second part of the thesis put the hands on engineering problems. Developing blockchain-based software is a challenge which hides several critical issues. The thesis faces the problem of defining practices to well design and develop blockchain-oriented software, providing issues to be considered in the future. In addition, in order to provide interesting cases of study, two blockchain-based systems are proposed.

## Thesis overview

The thesis is organized as follows.

Chapter 2 deals with the empirical study of Bitcoin transaction data. Studying data stored inside the blockchain means studying the network dimension evolution, the users' interactions and the richness distribution . This Chapter developes the results we have presented in [104, 106], and concerns the modelling and the analytic study of the entire transaction set, during the period between the first Bitcoin years (2009-2012). The modelling process passes through the creation of a bipartite graph, specifically a Petri net, to exploit the properties of a well structured algebraic formalism. Two sets of nodes describe the addresses set and the transaction set . The elements of the incidence matrices describe node interconnections. Results show interesting properties of the blockchain data. Addresses usage statistics and the number of interconnections follow power law distributions [29, 40, 97] . The model allows a higher level representation, that takes into account the real usage of addresses and transactions.

In addition, this chapter presents an intelligent algorithm for the deanonymization that implements an expert system, able to recognize if a user has owned disposable addresses (i.e. addresses that appears only one time in the blockchain) and the complete chain of address changes it did Results show the number of disposable addresses found, and how many times users change address.

Chapter 3 directs the research focus on the study of blockchain challenges

and opportunities if used in place of centralized banking systems. The chapter report the research results we have presented in [26] and discusses the necessity of a sustainable development, that can pass through the global financial infrastructure optimization, using more efficient systems than at present. Many banks are currently focusing on blockchain technology to promote economic growth and accelerate the green technologies development. Furthermore, the chapter describes the real performances of the Bitcoin system, in terms of its efficiency. After collecting data about mining costs and computing specific regression functions, it shows the results in terms of its efficiency, defining three quantities: "economic efficiency", "operational efficiency", and "efficient service". The obtained results show that by overcoming the disadvantages of the Bitcoin system, and of blockchain technology, we could be able to handle financial processes in a more efficient way than under the current system.

Chapter 4 opens the second part and discusses the need of a new software engineering branch, specific for the blockchain technology. It develops the results we have published in [107]. The chapter describes the key elements that characterize blockchain-based software, and define the Blockchain-Oriented Software Engineering (BOSE) environments. Blockchain-oriented software projects can be distinguished from other software projects due to the nature of the technology used. They need special attention, especially in terms of security and reliability issues, architecture specifications, metrics and modeling language definition [37]. In fact, there are not enough guidelines that properly drive the use of the blockchain. In addition, the chapter highlights the open issues, such as the role and effects of collaborations in blockchain projects, the improvement of development and testing platforms[32], and the creation of advanced support tools for creating more complex and efficient smart-contract systems.

Chapter 5 offers two case studies of the blockchain technology in application scenarios. The first concerns the use of the blockchain to create a smart city system that allows citizens to give, through mobile smart objects, some environmental measurements (air pollution, noise, humidity, etc), and that creates an aviable geographical located dataset. The development phase follows the scrum method[28]. The blockchain allows citizens to collaborate without any centralized system and make the data always available, immutable and certified. In addition, the use of smart contracts allows intelligent management of collected data. We have presented this case study in [55]. The second concerns the definition of a blockchain-based energy market. This system aims to allow the decentralized trading of the electric energy produced by citizens through renewable sources. We have presented it in [74]. This system uses the blockchain to handle deals and record energy purchases. The purpose is to make the citizen

free to buy and sell electricity to the best bidder and to optimize his profits and save. In this application, a smart contract system manages the energy transactions thorough the blockchain.

The following paragraph of this introduction briefly describes the general structure of a transactions scheme in blockchain systems.

## 1.1 The blockchain technology

The *Bitcoin electronic cash system* was conceived in the 2008 by the scientist Satoshi Nakamoto [90] with the aim of producing digital coins whose control is distributed across the Internet, rather than owned by a central issuing authority, such as a government or a bank. It became fully operational on January 2009, when the first mining operation was completed, and since then it has constantly seen an increase in the number of users and miners.

At the beginning, the interest in the bitcoin digital currency was purely academic, and the exchanges in bitcoins were limited to a restricted elite of people more interested in the cryptography properties than in the real bitcoin value. Nowadays bitcoins are exchanged to buy and sell real goods and services as happens with traditional currencies.

The main distinctive feature introduced by the Bitcoin system is the Blockchain, that is a shared infrastructure where all bitcoins transfer are recorded. Value transfer is called *transaction* and is an operation between users. To send and receive bitcoins, a user needs an alphanumeric code, called *address*. It epresents the user account and each address has a private key associated with it. No personal information is usually recorded in a Blockchain and for this reason Bitcoin protocol offers pseudo-anonymity. Different blockchains have been implemented so far and the technology often seems to work properly, even if most of them suffer from a lack of software engineering principles application in their development and deployment [107].

To date blockchain is the technology underlying Bitcoin, but is also the technology underlying other cryptocurrencies, such as Ethereum, Litecoin and hundreds of other cryptocurrencies. By analyzing this technology we can obtain many statistical properties of its associated cryptocurrency network, as well as the typical behavior of users, for example how users move bitcoins between their various accounts in order to preserve and reinforce their privacy.

The surge of interest regarding Bitcoin led scientists to face several other topics, in addition to the Blockchain analysis, Cocco et al. in [25] presented an agent-based artificial cryptocurrency market in which heterogeneous agents buy or sell cryptocurrencies, in particular Bitcoins. The model proposed is able

to reproduce some of the real statistical properties of the price returns observed in the Bitcoin real market. In [67] the same authors proposed an agent-based artificial cryptocurrency market in order to model the economy of the mining process. Starting from the GPU (Graphics Processing Unit) generation they reproduce some "stylized facts" found in real-time price series and some core aspects of the mining business.

Other works focus on security and privacy issues [17], cryptographic problems [52], social aspects of the Bitcoin users behavior [119, 131, 132].

### 1.1.1 The Bitcoin Cash System: an overview.

The Blockchain is a distributed and global database where all information about bitcoins' transactions are stored, but the term can also be used to denote the technology behind. It works as a public ledger which is composed of an ordered sequence of blocks. Blocks are validated and inserted into the chain and each block contains data about a variable number of validated transactions.

Bitcoin transactions originally represented value transfer of a cryptocurrency but they can be used to transfer any kind of information. Each transaction is composed by an input section and an output section, which report a list of addresses[1] and their associated values meaning bitcoins.

The information associated to each transaction in the Blockchain are characterized by:

- A list of inputs, each one containing one previous transaction;
- A non empty list of outputs (possibly coinciding with some inputs);
- The associated amounts to each output.

Users can own one or more addresses, and address creation is costless. Users' anonymity is preserved since the Blockchain stores only addresses, and neither user names nor other identity information are required to create an address.

Bitcoin clients (software which allow users to interact with the Bitcoin network) manage the addresses in *digital wallets*. Wallets store both public and private keys which are used to receive and to send payments.

Fig. 1.1 shows a simplified scheme of the interaction among transactions (called $\theta_i$) and addresses (called $\alpha_j$). In the figure seven transactions and six addresses are involved in the chains. The balance of bitcoins owned by users

[1] An alphanumeric string of 32 base-58 numbers which can begin only with "1" or "3", e.g.

$$1JQfVfzfxtfUb9kexSt7mHhcHxX6fyBJ5A.$$

;

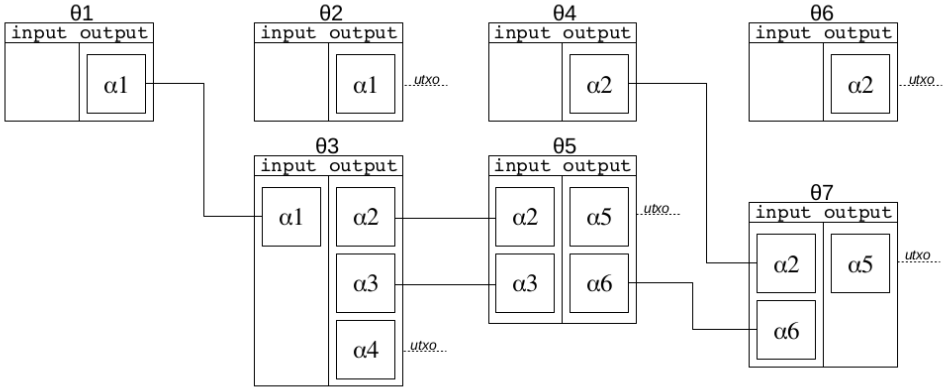


Figure 1.1: Simplified transaction schema.

is associated to their own address, and it is equivalent to the total value of the *unspent transaction outputs* (i.e., *UTXO*s) that the address has received and not spent yet. Each square in the input section represents a spent transaction. Each square in the output section that is not connected to the input of another transaction, represents an UTXO. For example, the addresses $\alpha_1, \alpha_2, \alpha_4$ and $\alpha_5$ have one or more UTXOs, so their balance is not null.

Each transfer of bitcoins among users implies changes on the balances associated to the respective addresses, similarly to what happens with a traditional bank account. Transaction requests wait in a "pending" status in the peer-to-peer network until they are validated by miners, in order both to prevent frauds and to avoid double spending.

Technical details about the network implementation can be found in [90]. Briefly, users interact with the Bitcoin network through clients which establish a Internet connection with some other client.

Each client become a node of the peer-to-peer network and, potentially, each node of the Bitcoin system has the same importance of any other one. Nodes listen for transaction requests arriving from other nodes. A transaction between addresses can be accepted only if it satisfies the following constraints:

- The transaction's inputs must correspond to the outputs of previous unspent transactions (UTXO) with same address and values;
- The transaction's output total value must be less or equal to the total value of the inputs, with a possible difference being the transaction *fee*.

The validation procedure, called *mining*, is carried out by *miners* and consists

in solving the (computationally hard) problem of determining an hash key starting with a given number of zeros (*nonce*) starting from a set of transactions requests as input. This hash key will be associated to the new *validated* block. In addition to transaction data, each new block contains several information such as the hash code of the previous block in the Blockchain, its height (its associated progressive number), and the IP address of the miner.

Mining the blocks is a competitive task which involves all the miners in the peer-to-peer network, which try to be the first to validate the next block. The first miner who is able to validate a new block receives a reward in bitcoins (presently 12.5 BTC). In case of multiple miners validate a new block, only one of them becomes a part of the Blockchain, the one with the greatest consensus. Mining and economic aspects will be discussed in Chapter 3.

The difficulty of this computational problem is automatically adjusted by the network, from time to time, in order to maintain constant, on a statistical base, the release rate of the new blocks (about a new one every ten minutes) and the consequent release of new Bitcoins.

In Fig. 1.1, we can identify the mining transactions. They are the transactions $\theta_1, \theta_2, \theta_4$ and $\theta_6$, which are the transactions having their input section empty. Nowadays, miners are gathered in pools to optimize the computational effort and to make constant the incoming of pool members. The whole Bitcoin system can be seen as a special typology of financial system in which, according with its technical specification, everyone can be a trader. Real time financial instruments, made possible by cloud and grid computing[8] could aid users in that operations. currently, the five greatest pool of miners have about the 70% of the total hash power. In order of importance, they are BTC.com, AntPool, ViaBTC, BTC.TOP and SlushPool. Each one has a computational power between the 11% and the 16% of the total.

The peer-to-peer network nodes control every new potential block, and elect it according with the consnsus algorithm. Nodes who store a full copy of the blockchain are called *full nodes* and are able to check new blocks for their validity, verifying that they respect the Bitcoin protocol. The number of full nodes is over ten thousand, distributed around the world. On November 2017, the United States had the highest number of nodes (28%), followed by the Germany (17%) and China (7%). [2]

[2] Source https://bitnodes.earn.com/

### 1.1.2 The Ethereum System: a brief introduction

Ethereum [3] is an open-source platform for running decentralized applications, based on a new generation blockchain. The idea was proposed by Vitalik Buterin and Gavin Wood in 2014, and has been described with a *white paper* [135]. The system includes a crypto-currency called Ether and is the second most important crypto-currency for capitalization after Bitcoin, worth over 31 billion (November 2017).

The Ethereum follow a development roadmap. The current version is called *Homestead* and shortly it will be updated to the *Metropolis* Version. Each update consists in a modification of the algorithm and pass through an operation called hard fork. It need the agreements of Ethereum nodes.

The Ethereum blockchain, similarly to the Bitcoin one, evolves by the creation of new blocks. They are generated by miners with a time interval of about one block every twelve seconds. Miners receive a fixed reward and equal to five Ether per mining block. Like the Bitcoin system, users of the Ethereum system can generate one or more addresses that can be used to receive and send Ether.

The main new feature introduced by Ethereum is the Ethereum Virtual Machine. The Ethereum Virtual Machine is the environment in which the smart contracts are executed and has a Turing-complete instruction set. A smart contract is a compiled script for the Ethereum Virtual Machine, developed by a user. The bytecode of the smart contract is included in the blockchain via a transaction. Smart contracts can be written in one of the programming languages supported by the Ethereum system. The most used language is called Solidity. After loading, the contract become equivalent to an Ethereum account, and will be able to receive and send transactions to and from any other contract. The contract code is executed each time it receives a transaction. The cost of executing Smart contracts is determined by the computational complexity. The cost of each instruction is measured in Gas. The value of Gas is expressed in Ether but is not fixed. They are system users who decide how much to pay for Gas. In fact, the Gas measures every type of operation (either a transaction or a calculation) within the Ethereum blockchain. Multiplying the value in Ether by the number of Gas you get the fee of the operation.

Ethereum clients exhibit a rich collection of RCP interfaces. This feature makes it possible to develop applications that use the blockchain, taking full control of all functionalities. In addition, it is possible to export the blockchain data in the Json format.

[3] https://www.ethereum.org

# Part I

# Blockchain analysis

# Chapter 2

# Blockchain Analysis

In this chapter we introduce a novel approach, based on a Petri Net to analyze the Blockchain. Using Petri Net we define a single useful model, a unique data structure, by which not only all main information about transactions and addresses are represented, as can be done using other approaches, but also the overall architecture and scheme of blockchain transactions are fully and natively implemented through a well known and powerful formalism.

We assume that each address corresponds to a place and each Bitcoin transaction corresponds to a transition in a Petri Net (also known as Place/Transition Net or P/T Net). The proposed model, called "Addresses Petri net", allows to quickly collect information on the identities owning Bitcoin addresses and to recover measures and statistics on the Bitcoin network. We reconstruct an Entities network associated to Block Chain transactions gathering together Bitcoin addresses into the single entity holding permits to manage Bitcoins held by those addresses. In other words, the use of Petri net formalism easily allows us to construct first the"Addresses Petri Net", and then the "Entities Petri Net". Even if we analyzed only a few features of the bitcoin blockchain, our model perfectly fits blockchain behavior and features and can potentially be used to exploit the full behavior of this new technology and to preform statistical simulations[127] on it.

There a number or advantages in using Petri net as a model to investigate the Bitcoin transactions. First of all, the well-defined algebraic model allows to manage straightforward algorithms to perform several structural analysis. Second, it allows to represent natively the Blockchain transactions, providing an alternative graphical representation of the Blockchain scheme. Finally, it opens up the possibility to perform dynamic simulations to forecast the future properties of the Bitcoin network. In fact the model allows the creation of higher level representations of the Bitcoin ledger, by grouping addresses in specific places and obtaining transition firing statistics.

## 2.1 Background

In these last years, the unique features of Blockchain have attracted more and more researchers and several are the works that examined this shared data collection. Even if several papers focused on heuristics and algorithms in order to analyze and cluster Bitcoin addresses identifying networks of users, no researcher focused on the analysis of the blockchain by modeling it within the framework of Petri Nets [104]. Consequently this section on related works will mainly focus on the works in literature which investigate blockchain technology, structure and properties from the point of view of dynamical networks.

Ron and Shamir [116] analyzed and measured the Blockchain up to the block number 180,000, from January 03th, 2009 to May 13th, 2012, by using a model called *transaction graph*. They analyzed the distribution of the number of transaction per address and introduced the concept of *entity* as a group of addresses of the same owner. They ran a variant of a Union-Find graph algorithm in order to find sets of addresses belonging to the same user. First, they constructed the transaction graph, the address graph, and then constructed the contracted transaction graph and the entity graph. Thanks to this entity graph, the authors determined various statistical properties of each entity, such as the distribution of the accumulated incoming bitcoins, the balance of bitcoins updated to May, 13th 2012, and the balance of the number of transactions per entity and per address. The authors obtained, for both the original and the clustered network (the entities network), some statistical properties which are typically encountered in complex networks [30, 93, 44, 127]. In addition they investigated the most active entities in the system.

In [116], the users' common practice to move bitcoins between their various accounts (addresses) is tracked as a good practice to preserve and reinforce user's anonymity.

Many other strategies adopted in order to preserve and reinforce users' anonymity have been analyzed in literature. Some of these strategies improve the privacy and anonymity including mixing protocols, and are discussed in CoinShuffle [118]. CoinJoin and CoinParty [140] investigated the use of anonymity networks obtained by using software like TOR. Biryukov et al. in [12] found countermeasures to block users who access in the Bitcoin network using Tor or other similar protocols. Reid and Harrigan [111] studied how an attacker could make a map of users' coins movement tracing their addresses and gathering information from others sources. They also focused in the topology of addresses network and transaction network, showing their properties of complex networks.

These results can be compared to those reported in [31] for clustering other software networks. Androulaki *et al* [6] analyzed how users try to reinforce theirs anonymity in the Bitcoin system. In particular, they studied the technique of changing address and how this makes more complex the network.

Meiklejom et al. [79] proposed an heuristic to recognize the changing addresses method, and to keep track of potential criminal users, thanks to information extracted from the Blockchain and from other sources, such as forums. They also tried to give a name to each address. Kondor et al. [62] focus on retrieving the Blockchain transaction network, studying its features over the time.

Recently, Lishke and Fabian [70] proposed an exploratory analysis of the Blockchain and of Bitcoin users. They studied the economy and main features of the Bitcoin cash system, but did not focus neither on the concept of "entity", nor on disposal addresses, as we do in this work. Their analysis revealed the major bitcoin businesses and markets, giving insights on the degree distribution (probability density function and complementary cumulative distribution function) of bitcoin transactions for several aggregations of time, businesses categories and country. These distributions revealed the existence of a scale-free network, and hence that Bitcoin network follows a power law distribution although not over the entire period. These results can be compared to those reported in [127] about the mechanism of power law distribution generation in similar technological networks and have also been replicated in our work, where we found that the distributions of several investigated quantities follow a power-law very closely.

## 2.2 The model: the Blockchain Petri Net

The proposed model is based on the Petri Net formalism. Using the Petri Net formalism obtained a lightweight but useful representation of the Blockchain that we call the Addresses Petri Net. Petri Net is an oriented graph, made of two types of nodes, place and transitions, where each node can be connected only with a node of the other type. Also the Bitcoin Blockchain can be modeled as an oriented graph, made of two types of nodes, addresses and transactions, where the latter activate transfers of tokens between the former, and thus can be natively modeled by using the Petri Net formalism for places and transitions, respectively.

### 2.2.1 Petri Nets: A brief introduction

A Petri Net [85, 47, 20] is a formalism to describe systems based on a bipartite graph with two kind of nodes called *places* and *transitions*. For this reason, Petri

nets are also called *Place Transition nets* (*P/T nets*). Connections between nodes are made by directed arcs. Each node can be only connected to nodes of the other type and there are two types of arcs: arcs ingoing into a transition, called *pre-arc*, and arcs outgoing from a transition, called *post-arc*.

One of the advantages of using Petri Nets is that they are also well described by an algebraic formalism. The formalism provides sets to define the nodes, and matrices to describe the arcs. A Petri Net $N$ is a quadruple defined as described below.

**Definition 2.1**

$$N = (P, T, Pre, Post) \tag{2.1}$$

*where*

- $P = \{p_1, p_2, ..., p_m\}$ is the set of $m$ places,
- $T = \{t_1, t_2, ...t_n\}$ is the set of $n$ transitions,
- $Pre : P \times T \rightarrow \mathbb{N}$ is the *Pre-incidence* function
- $Post : P \times T \rightarrow \mathbb{N}$ is the *Post-incidence* function.

$Pre$ and $Post$ incidence functions are usually defined by mean of matrices with dimension equal to $m \times n$. Each element of these matrices contains the number of arcs which connect places with transitions. The $Pre$ matrix contains the numbers of ingoing (to transitions) arcs for each place-transition pair. Vice versa, each element of $Post$ matrix is the number of outgoing arcs for each place-transition pair.

Petri nets are also a powerful formalism to describe discrete event systems, as is the case of blocks generation in the Blockchain. To model the state of a system, a marking $M$ (i.e., a vector which defines the distribution of tokens in places) is needed. Transitions are aimed at modifying the marking of the system. Transitions absorb tokens from places connected with Pre-arcs and produce tokens for the places connected with Post-arcs, an operation called *firing* of a transition. Petri net and the associated initial marking form the Network system defined as $\langle N, \mathbf{M}_0 \rangle$, where $\mathbf{M}_0$ is the initial marking. In this work we do not describe a specific state of the Blockchain so we do not need to define a marking.

### 2.2.2 Advantages of the Petri nets modeling

In this section we discuss the motivations for preferring the Petri nets formalism in modeling the bitcoin transactions on the blockchain (as well as other possible

transactions) and describe the intrinsic advantages carried by this formalism. Part of this discussion will include proposals for further research.

First, Petri nets formalism allows for the "non determinism criteria" in the system's dynamics. Such criteria accounts for respecting the locality principle in the system's evolution. In other words, Petri nets formalism natively includes independence between events generated by enabled transactions so that one enabled transaction can occur regardless the occurrence of other transactions for any given marking. Once an (or a set of) enabled transaction occurs he new marking has to be evaluated in order to understand which transactions are enabled in the new marking.

Such formalism perfectly fits into the blockchain transactions system where only transactions with non null UTXO are "enabled" and can occur, and their occurrence is independent from other transactions occurrence. A transaction occurrence is not deterministic and depends not only by the owner decision of sending bitcoin to another address, but also on the winning miners and on the probability that such transaction is included into the block validated by the hashing mechanism, which in turn has a different probability depending on the fee the owner accepts to pay. In such model enabled transactions natively correspond to UTXO and the marking corresponds to the set of all UTXO determined by the last block validated. The validation of a new block, where transactions are included in an independent fashion, determines a new "marking" of the bitcoin Petri nets net with a renewed set of UTXO. This enabling mechanism is hardly accounted for using a simple bipartite graph or a matrix representation for the bitcoin network and its transactions, even if many properties illustrated in this chapter can be recovered by using such representations. The advantage of the Petri nets formalism is that it natively includes such features.

The second aspect we discuss is related to simulation modeling which allows to analyze systems dynamics and which is a typical advantage provided by the Petri nets formalism. Differently from bipartite graphs, which account for a static analysis, PT nets formalism includes systems dynamics and allows for non deterministic system's dynamics modeling. In fact in Petri nets simultaneous transactions can occur provided they are not in conflict. Again this is a characterizing feature of bitcoin transactions dynamics where many non conflicting transfers of bitcoin between addresses can be included into the same validated block in the blockchain. Conflicting transactions, like for example double spending, are controlled and not allowed. Furthermore Petri nets can include into the dynamics modeling priorities between transactions and this can be used in a statistical modeling of the different probabilities the bitcoin transactions have to occur depending on the fee the owner accepts to pay.

A third feature natively included into the Petri nets formalism is the sequence of transactions: two transactions $t_1$ and $t_2$ are in a sequence if $t_1$ precedes $t_2$ with

$t_1$ enabled and $t_2$ not enabled for a given marking, and when the occurrence of $t_1$ enables $t_2$ in the new marking. The bitcoin transactions network natively contains sequences of transactions, like for example the sequences of UTXO generated into a single chain of disposable addresses monitored in our work and used for preserving bitcoin anonymity. Once again sequences of transactions can hardly accounted for using different representations, like bipartite graphs or matrices, without inserting ad hoc constraints into such representations.

Another important advantage is that Petri nets formalism includes the possibility to set state equations for the evolution dynamics. Given an initial marking the state equation allows the determination of the new marking according to the rules fixed for choosing the enabled transaction that effectively occur. The rules can be chosen with great freedom (respecting network constraints) and in particular a stochastic or probabilistic approach can be used in order to simulate the evolution of the blockchain from a statistic point of view. For example, one of the future improvements the authors are presently working on is to collect statistics on the bitcoin fluxes between addresses paying attention to address clustered in entities, to addresses corresponding to exchanges and to addresses owned by miners pools, in order to assign transitions probabilities for bitcoin flux exchanges between such addresses to be used for choosing the enabled transactions to choose into the corresponding Petri nets to make evolve its marking using a statistical approach. This will provide a set of possible future marking, each with its own probability of occurrence, which will correspond to future states of the blockchain. Such statistical modeling can provides hints on which addresses are going to get richer with a given probability, which pools of miners are going to exploit the future mining and at which rate and so on. The possibility of performing such a statistical simulation for the blockchain dynamics is straightforward within the Petri nets model whilst is hardly accounted for using different approaches which are mainly static.

Last but not least, the formalism, through the use of pre and post matrices, allows to recover many different and independent results following straightforwardly from standard computations over the pre and post matrices associated to the blockchain transaction network. For example, counting the number of rows and columns of matrices **PreA** or **PostA** it is straightforward to find the number of addresses and transactions, or we can find bitcoin addresses never used to spend looking at addresses with only zeros in **PreA** rows and at least one non null element in **PostA** rows, or we can recover disposable addresses looking at transitions that correspond to columns of **PreA** having only one non zero element and to columns of **PostA** having two non zero elements but in different rows.

### 2.2.3 Addresses Petri Net

In order to obtain the Petri net algebraic representation for the Blockchain we provide a set theory description of the two Blockchain elements involved, e.g., addresses and transactions.

We denote $\mathscr{A} = \{\alpha_1, \alpha_2, ..., \alpha_m\}$ the finite set of $m$ addresses $\alpha$ registered either as inputs or outputs in the Blockchain, and with $\Theta = \{\theta_1, \theta_2, ..., \theta_n\}$ the set of $n$ transactions $\theta$ validated by the Blockchain.

Let $N_\alpha = (P_\alpha, T, \mathbf{PreA}, \mathbf{PostA})$ be the network of addresses, where:

- $P_\alpha = \{p\alpha_1, p\alpha_2, ..., p\alpha_m\}$ is the set of $m$ places with each place $p\alpha$ associated to one and only one address $\alpha \in \mathscr{A}$;
- $T = \{t_1, t_2, ...t_n\}$ is the set of $n$ transitions where each transition $t$ is associated to one and only one transaction $\theta \in \Theta$;
- **PreA**: is the *pre-incidence* matrix;
- **PostA**: is the *post-incidence* matrix.

The sets $P_\alpha$ and $T$ can be recovered by browsing all the addresses and transactions validated in the Blockchain, which are publicly available, and inserting a new place every time a new address is found, and a new transition every time a new bitcoin transaction is encountered.

In order to build the matrices **PreA** and **PostA** let us consider one transaction $\theta$ in the Blockchain and the associated transition $t$. In the Blockchain, a transaction $\theta$ consists in a set of input and output addresses with the associated amounts in bitcoin. We denote by $In(\theta) \subseteq \mathscr{A}$ the set of input addresses, and by $Out(\theta) \subseteq \mathscr{A}$ the outputs set. For each address $\alpha \in In(\theta)$ we consider its associated place $p\alpha$ and we add a *pre-arc* leaving from $p\alpha$ and arriving to the transition $t$ associated to $\theta$. At the same time, for each address $\alpha \in Out(\theta)$ we add a *post-arc* leaving from transition $t$ associated to $\theta$ and arriving to the place $p\alpha$ associated to $\alpha$. For each couple $(p\alpha, t)$ to which a *pre-arc* has been added we set $\mathbf{PreA}(p\alpha, t) = 1$, while for each couple $(p\alpha, t)$ to which a *post-arc* has been added we set $\mathbf{PostA}(p\alpha, t) = 1$.

This model does not carry all the information available in the Blockchain (e.g. transactions amounts) and so it cannot completely represent Blockchain's behavior and properties. However, in contrast with the methodologies used in other works, in which different models were applied in order to analyze the Blockchain overloading the analysis, our approach natively represents the Blockchain structure and dynamics and includes into one single model and into one single data structure different features and properties of the Blockchain.

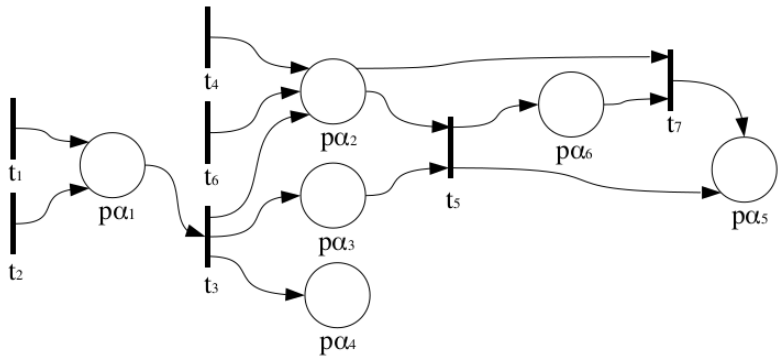


Figure 2.1: Addresses Petri Net equivalent to the simplified transaction chains in Fig. 1.1

$$\textbf{PreA}=\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{matrix} p\alpha_1 \\ p\alpha_2 \\ p\alpha_3 \\ p\alpha_4 \\ p\alpha_5 \\ p\alpha_6 \end{matrix}$$
$$\begin{matrix} t_1 & t_2 & t_3 & t_4 & t_5 & t_6 & t_7 \end{matrix}$$

Figure 2.2: Pre-incidence matrix of the Petri net for the example in Fig. 1.1.

Consider for instance the simplified transaction chains in Fig. 1.1. There are seven transaction and six places. The equivalent Address Petri Net is composed by six places and seven transitions. The graphical representation is shown in Fig. 2.1. This Net is defined by a set of places $P_\alpha = \{p\alpha_1, p\alpha_2, ..., p\alpha_6\}$, a set of transactions $T = \{t_1, t_2, ...t_7\}$ and by the *pre-* and *post-incidence* matrices **PreA** and **PostA**, shown in Fig. 2.2 and 2.3.

These matrices can be straightforwardly used to perform several analysis of the network. For example, we can compute the difference between *post* and *pre-incidence* matrices and consider one of its row. The number of not null elements in such row is equal to the number of UTXO contained in the address related to the place corresponding to the row. This number must be greater than or equal to zero, and if it is equal to zero the balance of the associated address is null.

In addition, we can easily compute the number of times that an address appears as input in a transaction. In fact, all the not-zero elements of the row $i$

$$\mathbf{PostA}=\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{matrix} p\alpha_1 \\ p\alpha_2 \\ p\alpha_3 \\ p\alpha_4 \\ p\alpha_5 \\ p\alpha_6 \end{matrix}$$
$$t_1 \quad t_2 \quad t_3 \quad t_4 \quad t_5 \quad t_6 \quad t_7$$

Figure 2.3: Post-incidence matrix of the Petri Net for the example in Fig. 1.1.

of the matrix **PreA** provide the number of times the address $\alpha$ corresponding to the place $p\alpha = i$ has been the input of a transaction.

As other example we consider the case of different transactions occurring in different moments which share the same input set and the same output set. Using our model, these transactions can be represented with only one transition, which is characterized by a firing clock. This feature, along with the creation of the entities net, can be useful to enable a dynamical and high level analysis of the Bitcoin system. We will show in the following that our model allows to easily detect such sets of transactions.

## 2.3 Deriving the Entities

It is quite common for Bitcoin users to hold more than one address in order to manage bitcoin exchanges and anonymity more easily. As in [116] we define an *entity* as the person, the organization, the group of people, or the firm that hold the control of the bitcoins associated to a set of addresses. All addresses appearing in an input section of a single transaction must be owned by the same entity. This is because, in order to activate the bitcoin transfers from those addresses, the same entity must hold all the private keys of all corresponding wallets

In order to build the Entities Petri Net $N_\epsilon$ we associated each entity to a collection of addresses, associating places $p\epsilon \in P_\epsilon$ in $N_\epsilon$ to a set of places $p\alpha$ of $N_\alpha$. We denote by $E = \{\epsilon_1, \epsilon_2, ..., \epsilon_k\}$ the set of *entities* where each *entity* $\epsilon \in E$ is a finite set of addresses such that $\epsilon \subseteq \mathscr{A}$.

The matrix **PreA** has $m$ rows, one for each place, and $n$ columns, one for each transition. Given a transition $t$ we consider the array **PreA**$(\cdot, t)$ which is the column of **PreA** with index $t$. Its non zero elements correspond to places $p\alpha$ with **PreA**$(p\alpha, t) = 1$, namely places with outgoing arcs *pre-arc* towards transition $t$. These places $p\alpha$ correspond to input addresses $\alpha \in In(\theta)$, for the transaction $\theta$

Let be $T^* = T$ the set of unexplored transitions and $E = \emptyset$ the set of entities.

- while $T^* \neq \emptyset$
  1. take a $t : t \in T^*$ and remove this form $T^*$
  2. let $e = \emptyset$
  3. for all $i : \mathbf{PreA}(p_i, t) \neq 0$ do $e = e \cup \{p_i\}$
  4. let $e^* = e$ the set of unexplored places
  5. while $e^* \neq \emptyset$
     (a) take a place $p \in e^*$
     (b) let $T' = \emptyset$
     (c) for all $j : \mathbf{PreA}(p, t_j) \neq 0$ do $T' = T' \cup \{t'\}$
     (d) for all $t' \in T'$
        i. let $e_{new} = \emptyset$
        ii. for all $h : \mathbf{PreA}(p_h, t') \neq 0$ do $e_{new} = e_{new} \cup \{p_h\}$ endfor
        iii. $e = e \cup e_{new}$ and $e^* = e^* \cup e_{new}$
        iv. $e^* = e^* \setminus p$
        v. $T^* = T^* \setminus t'$

        endfor

     endwhile
  6. $E = E \cup e$
- endwhile

Figure 2.4: Algorithm used to compute the set $E$ of entities.

corresponding to transition $t$. As a consequence, all these places belong to one single entity $e \in E$.

It is also possible that a given address appears in two or more input sections, together with other addresses. In this case, the entity must be composed by all the addresses in these input sections.

### 2.3.1 Building the Entities Petri Net

To build the Entities Petri Net, $E$, we applied the following algorithm.

We denote *unexplored place*, every place which is an element of the current entity, but is not yet processed. In fact, in order to find other places to be inserted

| Entity in $E$ | Places |
|---|---|
| $e_1$ | $\{p\alpha_1\}$ |
| $e_2$ | $\{p\alpha_2, p\alpha_3, p\alpha_6\}$ |
| $e_3$ | $\{p\alpha_4\}$ |
| $e_4$ | $\{p\alpha_5\}$ |

Table 2.1: Entity in the Entities Petri Net of the simplified transaction chains in Fig. 1.1.

into the current entity $e$, each unexplored place must be processed as in step 5. In this step, all the other places $p_h$ element of $e$ are found.

Each $e \in E$ is a set of places of the Addresses Petri Net or, equivalently, is the representation of a set of addresses that compose an entity.

The algorithm creates the set $E$ of entities. The correctness of the algorithm can be discussed analyzing the two requirements: the finite number of iterations and the correctness of the solution. Firstly, the number of iterations is limited by the number of transitions. In fact, the set of unexplored transitions will be emptied every time a transition will be examined. In particular, both in step 1 and in step 5.d.v. a transition is removed from $T^*$. Regarding the second point, because place determination occurs by evaluating the *pre-arcs* connected to each transition, entities are correctly created and populated. Furthermore, it is possible to check that the resulting entities form mutually disjoint sets and that the result of the entities' union contains all the places of the Addresses Petri net.

We can define $N_\epsilon$, the Entity Petri Net, as $N_\epsilon = (P_\epsilon, T, \mathbf{PreE}, \mathbf{PostE})$, where $P_\epsilon$ is the set of places that are associated one to one with elements of the entities set $E$.

The definition includes the set $T$ of transitions. This is the same that we have in the Addresses Petri Net.

In order to compute **PreE** and **PostE** rows, we take every entity $e \in E$. Given an entity $e$, we first extract from **PreA** and then from **PostE** the rows corresponding to every place $p_\alpha \in e$. Then, for each matrix, we sum these rows together. In this way, we obtain one new row for both **PreE** and **PostE**, corresponding to the entity $e$.

For instance, looking at the Address Petri Net in Fig. 2.1 and at the **PreA** matrix, we recognize that places $p\alpha_2, p\alpha_3$ and $p\alpha_6$ can be joined to an entity, and that hence their related addresses $\alpha_2, \alpha_3, \alpha_6$ are owned by the same person. In total four entities are recognized as described in Table 2.1.

To each entity, a place $p\epsilon \in P_\epsilon$ is then associated. In the following tables,

$$\mathbf{PreE}=\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{matrix} p\epsilon_1 \\ p\epsilon_2 \\ p\epsilon_3 \\ p\epsilon_4 \end{matrix}$$
$$t_1 \quad t_2 \quad t_3 \quad t_4 \quad t_5 \quad t_6 \quad t_7$$

Figure 2.5: Pre-incidence matrix of the Entities Petri Net for the simplified transaction chains in Fig. 1.1.

$$\mathbf{PostE}=\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{matrix} p\epsilon_1 \\ p\epsilon_2 \\ p\epsilon_3 \\ p\epsilon_4 \end{matrix}$$
$$t_1 \quad t_2 \quad t_3 \quad t_4 \quad t_5 \quad t_6 \quad t_7$$

Figure 2.6: Post-incidence matrix of the Entities Petri Net for the simplified transaction chains in Fig. 1.1.

**PreE** and **PostE** of the example resulting Entities Petri Net are shown in Fig. 2.5 and 2.6. In Fig 2.7, the graphic representation of the Entities Petri net is shown.

## 2.4 Analysis set-up and Results

Blockchain can be explored mainly using two approaches. The first consists in downloading all binary data from the peer-to-peer network, and in identifying transactions, addresses and other information by using protocol instructions. The second one consists in exploring specific websites where the decoded Blockchain is shown, and application interfaces or other utilities, are provided to explore it. We followed the second approach and downloaded blocks as formatted JSON files from the website *blockchain.info*.

We parsed the first 180,000 blocks in the Blockchain, corresponding to a period of about three and half years, from January 2009 to March 2012, in order to compare our results with those in [116].

The data processing performed in this work is carried on in steps as shown in Fig. 2.8. All implementations are made with R language and RStudio IDE.

The analyzed portion of Blockchain was processed without specific hardware resources and processing time to elaborate the first 180,000 blocks has been about 250 hours long. The average time required to compute a block is 5 seconds. The single block computation time depends on the number of addresses

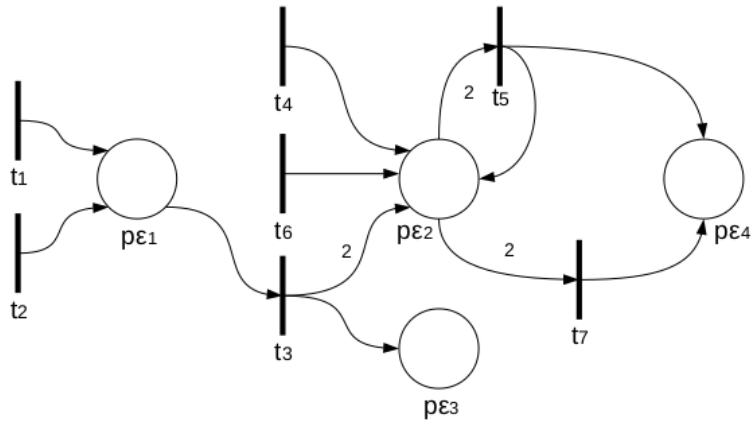


Figure 2.7: The Entities Petri Net of the simplified transaction chains in Fig.1.1

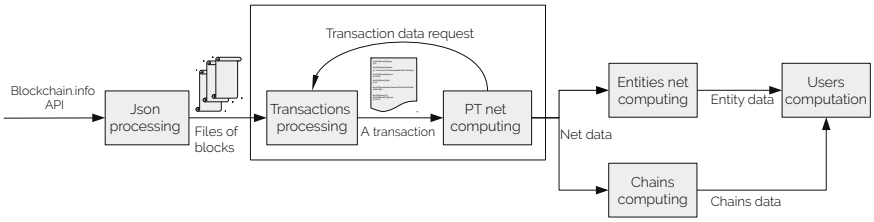


Figure 2.8: Diagram of the data processing path for the study of Blockchain

contained in it, considering every address in input or in output of a transaction. The procedure requires about ten seconds to elaborate eight addresses and does not increase significantly even when matrices become larger.

The situation has quite changed for blocks validated in subsequent periods. Currently a block contains about three thousand addresses and the time to compute it using our Petri Nets modeling is about six minutes long. Generally the time to elaborate an addresses is larger if the address has not yet been found and the search algorithm must add it into the matrices. The downloaded JSON files, elaborated and saved in a R structure, occupies 2.8GB and after the elaboration the addresses Petri net occupies about 800MB in RAM. Saving corresponding data in a Rdata file, it occupies about 40MB.

### 2.4.1 Investigating on the Addresses Petri net

We found 3,730,480 different addresses and 3,142,019 transactions, which in our model correspond to the number of rows and columns of matrices **PreA** or **PostA**. We associated the addresses to the corresponding places in the set $P_\alpha$ in the Petri Net $N\alpha$. From the analysis of the matrices **PreA** and **PostA**, simply counting the non zero elements, we found 4.575.888 *pre-arcs* and 7.352.494 *post-arcs* in total. The number of non zero elements $L(i)$ on the corresponding row of **PreA**$(p\alpha_i, \cdot)$ represents the number of transitions occurring from the place $p\alpha_i$ through a *pre-arc*. The number of non zero elements $L(i)$ in the row **PostA**$(p\alpha_i, \cdot)$ represents the number of transitions connected to the place $p\alpha_i$ through a *post-arc*. Using this formalism our model easily takes into account the total number of bitcoin transactions in input and output of each address.

Figures 2.9 and 2.10 report the Complementary Cumulative Distribution Functions (CCDF) defined as the probability $P$ that $P(L) > x$, where $L$ is defined as the number of non-zero elements in the matrices **PreA** and **PostA** respectively.

The figures show an uneven distribution of $in$ and $out$ transactions among addresses so that there are many addresses with few transactions and relatively few addresses with many transactions, displaying a typical power-law distribution. Such distribution has been straightforwardly recovered using the Petri Nets formalism.

In table 2.2 we report the ten most used addresses, found summing up the number of non zero elements in **PreA** rows to that of non zero elements in **PostA** rows.

Our analysis identifies also 609,295 addresses with only zeros in **PreA** rows, and at least one non null element in **PostA** rows, namely 609,295 addresses never used to spend (until the 180,000 block), but only to accumulate. Part of them are

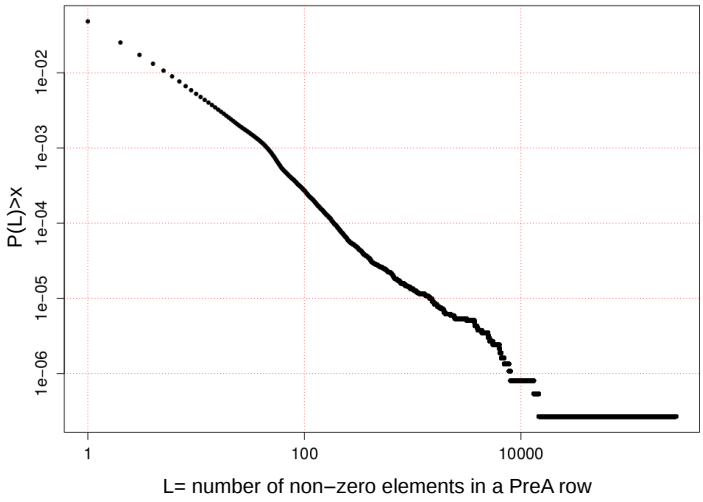


Figure 2.9: CCDF of the length L for $PreA$.

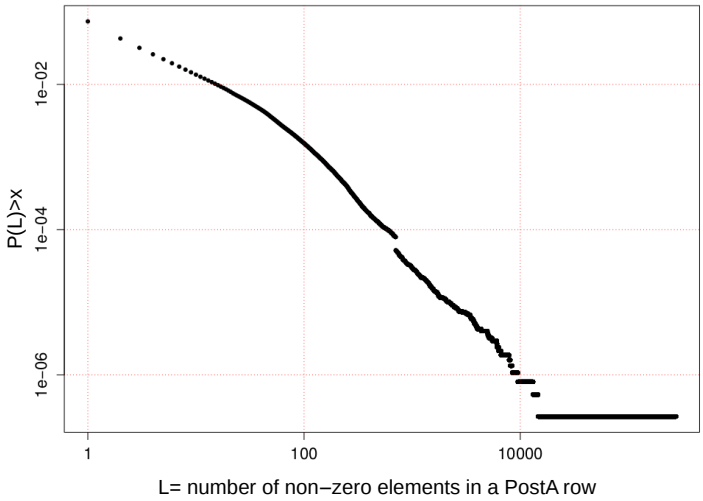


Figure 2.10: CCDF of the length L for $PostA$.

still unused up to today. Table 2.3 reports the first ten ranked by the number of *post-arcs*, namely the number of incoming transactions, and shows their current balances, as checked from blockchain.info: four of these (row 1,7,8 and 9 in the tab) are never used again, and can be called *dormant*. Their balance can be quite high, since they've been used like sort of bitcoin deposits.

| Address | L pre | L post | tags |
|---|---|---|---|
| 1VayNert3x1KzbpzMGt2qdqrAThiRovi8 | 270,204 | 275,398 | deepbit.net |
| 1dice8EMZmqKvrGE4Qc9bUFf9PX3xaYDp | 14,606 | 14,605 | SatoshiDICE 48% |
| 1dice97ECuByXAvqXpaYzSaQuPVvrtmz6 | 13,137 | 13,124 | SatoshiDICE 50% |
| 159FTr7Gjs2Qbj4Q5q29cvmchhqymQA7of | 8,016 | 8,425 | - spammer ? - |
| 1CDysWzQ5Z4hMLhsj4AKAEFwrgXRC8DqRN | 6,382 | 9,501 | Instawallet |
| 1E29AKE7Lh1xW4ujHotoT4JVDaDdRPJnWu | 7,761 | 8,079 | - unknow - |
| 15VjRaDX9zpbA8LVnbrCAFzrVzN7ixHNsC | 6,999 | 7,888 | faucet donation |
| 15ArtCgi3wmpQAAfYx4riaFmo4prJA4VsK | 6,578 | 6,622 | faucet donation |
| 1dice9wcMu5hLF4g81u8nioL5mmSHTApw | 6,318 | 6,306 | SatoshiDICE 73% |
| 1Bw1hpkUrTKRmrwJBGdZTenoFeX63zrq33 | 5,498 | 5,498 | - unknow - |

Table 2.2: Summary of first 10 most used addresses

| Address | L post | current balance BTC |
|---|---|---|
| 15S1TFTosxrgZxkqJR2n1AFJ22ZJE2rTCk | 3,853 | 120.85215349 |
| 1PtnGiNvhAKbuUQ6nZ7nF3CDKCKGfeMsCX | 1,199 | 0 |
| 129FTwWoi5H5ujasMZ6M6VjJzBJfsXVQGw | 1,138 | 0.78425567 |
| 1FN9kKsZA9XttrAwuDDgsXjs6CXUR2fzmt | 1,111 | 0 |
| 1DYvtKtZ2Ay9vTjzjb9BiRauMgXdjRDaD | 973 | 14.5601 |
| 1STRonGxnFTeJiA7pgyneKknR29AwBM77 | 949 | 1.79274504 |
| 1Q3nqtUzBp6jw7opi674Pyfgu4MUmVRdrk | 861 | 16.31551365 |
| 1Hh3eNNqR8MajEtDfvUF3hoxgf8CuUXVwY | 819 | 257.32881319 |
| 14sx4sFdUE9YDpJ9XbD6xAUEKPKvc8QHq2 | 811 | 59.56546509 |
| 17igtzSD39ZAapsut2DQTTKFyqSp7CToMq | 809 | 0 |

Table 2.3: Summary of first 10 most imbalanced addresses

We counted how many times users repeat the same transaction in terms of the same set of addresses in input section and the same set of addresses in output section. In our model identifying these repetitions is trivial. When two or more transactions involve identical sets of addresses in input and output, the corresponding transitions are connected with the same places both in pre and in post matrices.

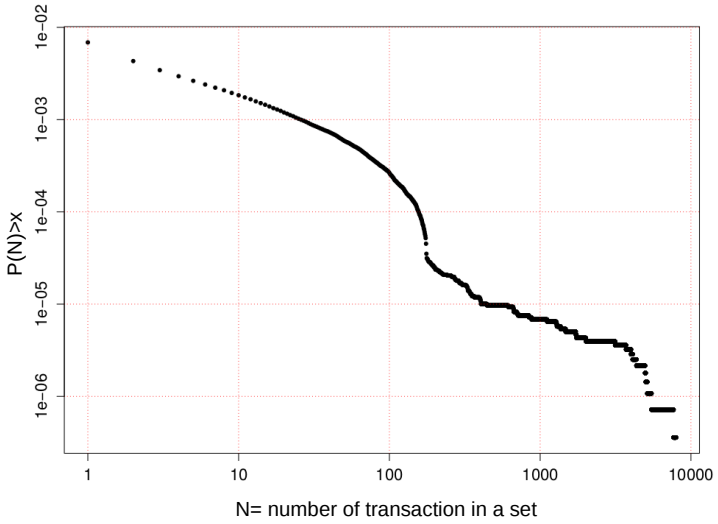


Figure 2.11: CCDF of the size L of grouped transaction set for the address net

In fact, taking the matrix **PA** = (**PreA**, **PostA**), in which the two matrices are concatenated in column, for each column $t_j$ it is possible to check the existence of other identical columns.

We found that about 11% of transactions are a repetition of another one. These represent repeated transfer of bitcoins from one group of addresses to another group of addresses where the two groups are always the same, revealing steady fluxes of bitcoins. Figure 2.11 reports the CCDF for the sizes of these groups of repeated transactions.

### 2.4.2 Investigating on the Entities Petri Net

The reducing algorithm discussed in section 2.3 is applied to the Addresses Petri Net in order to recover the corresponding Entities Petri Net. Among the owners, we found that 2,461,010 entities hold all the 3,730,480 addresses, and the distribution of addresses among entities is highly not uniform. Figure 2.12 shows that also such distribution follows a power-law very closely. This means that there are many entities holding a single address but also a few entities controlling very many addresses, and thus able to control a great fraction of the bitcoins flux transactions.

There are only 246,660 entities containing two or more addresses and these contains 1,516,130 addresses. The number of non null elements in the rows

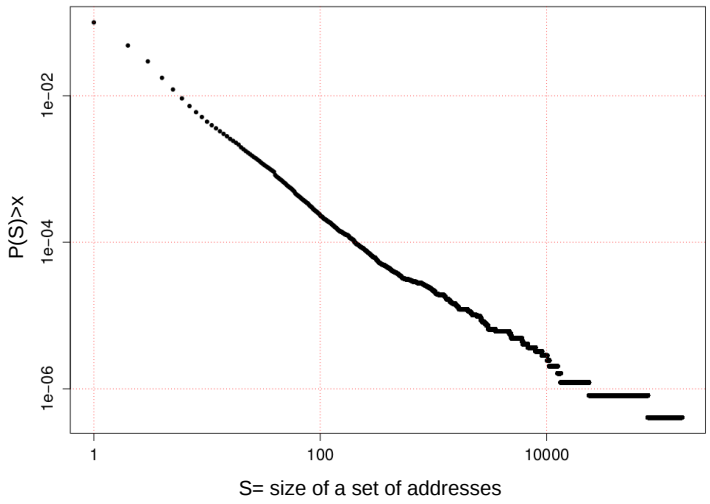


Figure 2.12: CCDF of the distribution of addresses across entities.

of matrices **PreE** and **PostE** for the Entities Petri Net is reported in Figure 2.13 and 2.14 respectively. This corresponds the number of transactions where the entities are involved. They clearly show a power-law distribution for transactions among the entities.

In Tab. 2.4 we report the ten most used entities, found summing up the number of non zero elements in **PreE** rows to that of non zero elements in **PostE** rows. Their balances can be computed summing the balances of all addresses belonging to the corresponding entity and are owned by a single user.

| Entity number | L pre | L post | size | tags |
|---|---|---|---|---|
| 95237 | 270,204 | 275,398 | 2 | deepbit.net |
| 2 | 102,186 | 283,973 | 156,725 | ilovethebtc |
| 37 | 51,228 | 147,712 | 78,251 | jmm5699 |
| 11 | 49,959 | 97,732 | 10,37 | - unknow - |
| 130 | 20,857 | 58,350 | 23,649 | Instawallet |
| 66437 | 14,219 | 60,868 | 13,289 | Rai, Dread88 |
| 42 | 9,268 | 31,147 | 10,561 | Quip, iosp and other |
| 37598 | 8,923 | 31,004 | 12,520 | generalfault, safetyvest.com |
| 220 | 11,133 | 27,487 | 9,093 | zephram |
| 1503 | 9,044 | 29,400 | 10,116 | folk.uio.no/vegardno |

Table 2.4: Summary of first 10 most active entities

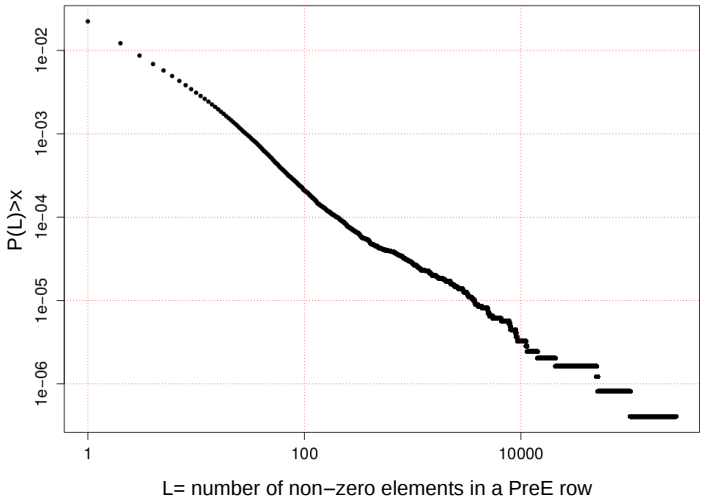


Figure 2.13: CCDF of the lenght L for $PreE$.


Figure 2.14: CCDF of the lenght L for $PostE$.

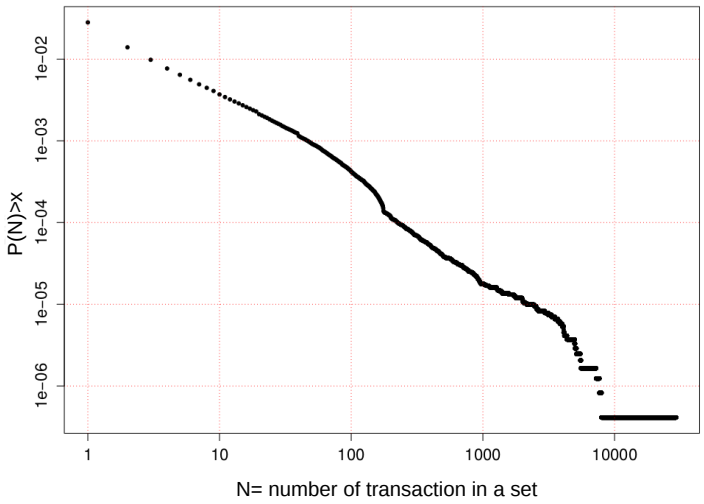


Figure 2.15: CCDF of the size L of grouped transaction set for the Entity Petri Net net

Like for the Addresses Petri Net, we computed groups of repeated transitions for the Entities Petri Net. We found that about 22.6% of transactions are a repetition of another one occurred among the same entities in input and in output. This information allows to identify steady fluxes of bitcoins at the owners level. Figure 2.15 reports the CCDF for the sizes of these groups of repeated transactions.

## 2.5 Deanonimization: disposable addresses

Most Bitcoin users are very careful about keeping their anonymity. This paragraph investigates the use of disposable addresses, a very common method adopted by users to protect their identity and proposes a method to recognize these addresses. This is applied on the first 180,000 blocks of the Bitcoin Blockchain. Results highlighted that a large part of Bitcoin transactions involves disposable addresses. Further, they showed that many of these transactions form chains whose length is characterized by a power-law distribution.

In order to preserve and reinforce the anonymity of the Bitcoin users, many strategies have been proposed. Some of these strategies improve the privacy and anonymity including mixing protocols (eg. CoinShuffle, CoinJoin and CoinParty [140]), and others are based on the TOR network. One of the most known strategies to preserve and reinforce the anonymity is the massive use of *disposable*

*addresses*. This strategy consists in using an address only one time. In other words, a user uses a new Bitcoin address each time it receives a new payment or executes a new payment. Despite the adoption of these strategies, de-anonymize users' identity can be possible. Recent researches try to de-anonymize the users' identity by using external data [12, 64] and others propose clustering heuristic to form user networks [79, 116].

In the following an original strategy based on a Petri net formalism is proposed, with the aim to recognize the disposable addresses, and hence the chains of transactions in which these addresses are involved.

### 2.5.1 Chain of disposable addresses

After having built of the Petri net $N$, we focused our attention on the chains of disposable addresses, , that is the addresses that appear only two times in the Blockchain. The first time when they receive Bitcoin and the second time when they are completely emptied. and hence on the transactions having only one address, $\alpha_a$, in the input section and only two addresses, $\alpha_b$ and $\alpha_c$, in the output section. In more detail, the address $\alpha_a$ in the input section is used by a user $u_1$ to send bitcoins to one of the addresses in the output section, $\alpha_b$, belonging to a user $u_2$. The other address, $\alpha_c$, in the output section is created by the user $u_1$ to collect the change. The set of potentially disposable addresses $A_d$ can be created starting from the set $A$ of the addresses $\alpha$ and from the set $\Theta$ of the transactions $\theta$ in the Blockchain.

Let $\Theta_d$ be the set of transaction $\theta_d$ such that:

$$\Theta_d \subseteq \Theta = \{\theta_d : |IN(\theta_d)| = 1, |OUT(\theta_d)| = 2, IN(\theta_d) \in A_d,$$
$$\exists\, \alpha \in OUT(\theta_d) : \alpha \in A_d, \forall \theta_d \in \Theta_d\}.$$

In order to build a chain, for each $\theta_d$ we need to know the previous transactions $\theta_{dp} = PREV(\theta_d)$. Using $Pre$ and $Post$ matrices, it is very easy to look for these previous transactions. We call $\Theta_{ds} \subseteq \Theta_d$ the set of transaction $\theta_{ds}$ that could be considered the starting point of a chain because it does not have a previous transaction inside $\Theta_d$ . We denote with $\alpha_{ds}$ the address in input to a transaction $\theta_{ds}$. Finally, we call $NEXT(\theta_d)$ the transaction $\theta_{d'}$ which has, in the input section, the disposable address that is contained in the output section of the transaction $\theta_d$. To find the chains $c$ of disposable addresses, we defined and implemented the following algorithm:

1. Let $C = \emptyset$ be a set of empty chains, $c$,

2. for each $\theta_{ds} \in \Theta_{ds}$:

| Description | Value |
|---|---|
| Potential disposable address $\alpha_d$ | 2,897,577 |
| Involved transaction $\theta_d$ | 1,350,010 |
| Number of chains $c$ | 122,155 |

Table 2.5: The dimension of the sets of potentially disposable addresses, the number of involved transactions and the number of the chains.

(a) take a empty chain, $c$,

(b) insert $\theta_{ds}$ in $c$

3. for each $c \in C$

(a) take the last element inserted in $c$, $\theta_d$,

(b) while $\exists\, \theta_{d'} = NEXT(\theta_d)$

i. insert $\theta_{d'}$ in c,

The algorithm returns a set $C$ of chains $c$. Each chain $c$ contains the transactions ordered by execution order.

### 2.5.2 Results

The computation of the chains is a three steps process. The first step is to identify the potential disposable addresses. The second step is to recognize transactions in which are involved disposable addresses. The third step is to build chains. The result of the performed analysis are illustrated in Tab 2.5 e 2.6. The analysis allow us to compute the dimensions of sets of potential disposable addresses and the transactions, which are involved in the computation, and the number of found chains ( see Tab. 2.5). We found that over one third of addresses inside the Blockchain are actually disposable addresses. The chains length is highly variable. The found longest chain contains 3,658 transactions and involves an equal number of disposable addresses. The first five chains ordered by length, the number of blocks where each chain appears (calculated as the difference between the ending block number and starting block number) and the rate of transaction execution per block (calculated as the ratio between the length of the chain and the number of blocks) are summarized in Tab. 2.16. It is interesting to note that some long chains were executed in the time of few tens of blocks (in temporal terms, in few hours). We computed the Complementary Cumulative Distribution Function (CCDF) of the lengths. The graph of the distribution, in

| Statistics | Min | Max | Mean | Median | Variance |
|---|---|---|---|---|---|
| Value | 2.00 | 3,658.00 | 11.05 | 3.00 | 1230.629 |

Table 2.6: Statistics of the chains lengths

| Chain | Length | Blocks | Rate |
|---|---|---|---|
| 121877 | 3658 | 132 | 27.71 |
| 120862 | 2502 | 42 | 59.57 |
| 1918 | 2454 | 724 | 3.31 |
| 120871 | 2169 | 19 | 114.16 |
| 28719 | 2000 | 1387 | 1.44 |

Figure 2.16: table
Top five chains ordered by length.
*Blocks* is the number of blocks tha contain each chain from the beginning to the end and *Rate* is the average number of transaction per block for the chain

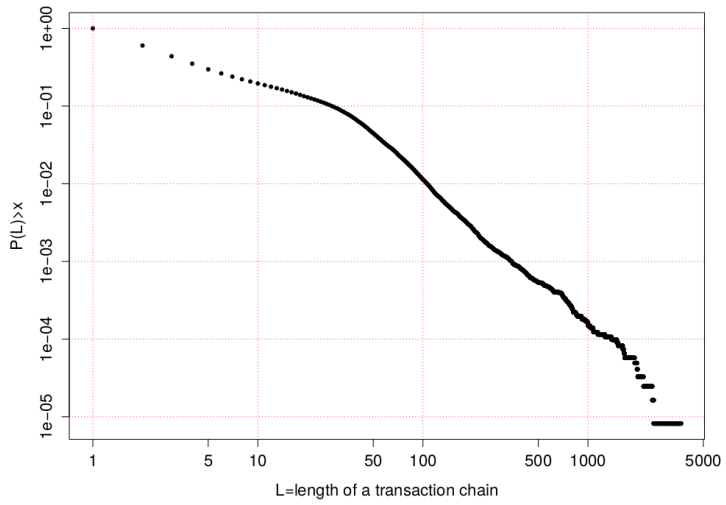


Figure 2.17: CCDF of chains lenghts.

Log-Log scale, is showed in Fig. 2.17. This distribution follows a Power-law in the tail (starting from a length about 20).

To recap, Fig. 2.18 summarizes the results for the address number found for each type of address defined in the chapter.

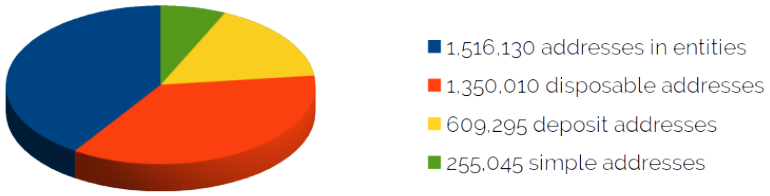


Figure 2.18: Addresses per typology across the first 180.000 blockchain blocks

## 2.6 Discussion

The Petri Net formalism can be natively used to infer many information and features about Bitcoin users and the Blockchain. We used the Petri Net model to gather together group of addresses (as entities or groups of disposable addresses) trying to associate an identity to each group. We also estimated how many users have been actually involved in the first three years and half of Bitcoin activity.

Analyzing the entities we found that 1,516,130 addresses are controlled by 246,660 owners at most. With our model we were able to trace transactions chains whenever disposable addresses are involved. Each chain holds addresses belonging to one owner, but one owner may control more than one chain. So, according to our results, the 1,350,010 addresses involved are owned at most by 122,155 owners. Using $pre$ and $post$ matrices we found that 609,295 addresses are used only as output of bitcoin transactions and are not involved in entities or chains. Fig 2.18 shows the addresses

These three facts, enable us to estimate a threshold for the number of different Owners or users in the Bitcoin system. We compute that there were 368,815 *engaged* (or expert) owners that adopted disposable addresses practice or used two or more addresses in their operations. The addresses of such owners are involved in the 72,6% of transactions. We suppose that the 609,295 addresses that appear only in output are used by some *engaged* owners for the purpose of bitcoin depositing. Finally, the 255,045 remaining addresses are owned by occasional users. Furthermore we tried to associate an identity to addresses showed in Tab. 2.2 and to entities showed in Tab. 2.4.

Information about these addresses can be found on the Internet, in particular on blockchain explorer websites like *blockchain.info*, or specialized forums like *bitcointalk.org*. Some of theme are made more easily recognizable by attributing a *tag*. Take the case of the most used address we found, which appears 270,204 times as the input of a transaction. We were able to recognize that it belongs to a (now closed) Bitcoin *pool* which was called *DeepBit*.

Another example regards the most used entity in output that includes 156,725

addresses and regroup the 4.2% of the total number of addresses. Searching on the Internet some of its addresses, we found out that who manages this entity has used varied tags, such as *ilovethebtc*, *mikeo*, *FredericBastiat*, *edgeworth*, etc. [1] Several addresses in that entity have no tag.

From all the reported CCDFs it is clear that all the distributions are characterized by a strongly uneven amount of transactions across the addresses, either for pre and post transactions. This means that there are many addresses where the bitcoins are hardly exchanged, and few addresses where the rate of bitcoin exchange is particularly high. This analysis can be helpful for identifying addresses which are used by pool of miners. In fact, when miners join together in a pool to share computational facilities for mining operations, they need to define a common address where the mining rewards is accounted to. Then they need to redistribute the amount of gained bitcoins among all the pool users. As a consequence the address will be affected by a number of transitions in the corresponding Petri Net as large as the pool's size.

Finally, since we analyzed a limited window of 180,000 blocks, the amount of transitions found in the matrices are also a signature of the average rate of Bitcoin transferred between different entities and such rate can be used to infer information on the organizations which can manage massive Bitcoin transfers.

[1] It is possible to find a portion of the addresses included in that entity, in the input section of a Bitcoin transaction, available on *blockchain.info* and reachable from this short link `http://tinyurl.com/ilovethebtc`. Some of theme have a tag.

# Chapter 3

# Sustainability of blockchain-based banking systems

This Chapter discusses the challenges and opportunities of implementing blockchain technology across banking. The blockchain technology can optimize the global financial infrastructure, achieving sustainable development, using more efficient systems than at present.

One of the great challenges of mankind regards the sustainable development, reducing the greenhouse effect and avoiding climate change. Many organizations, such as financial institutions, are looking for reduce their carbon footprint, also trying to save money.

Information and Communication Technology (ICT) has an essential role in tackling these challenges. However, if, on the one hand, ICT can help to reduce energy and resource consumption, on the other hand, its ever-increasing usage induces rising demands for energy and resources. The cost of running an IT Infrastructure goes well beyond the cost of acquisition and manpower. It comprises the cost of powering the whole system, and depends heavily on computer software and software process models. There are three main impacts of ICT on the environment (see work by [91]):

> *"First-order impacts are environmental effects that result from production and use of ICT, i.e., resource use and pollution from mining, hardware production, power consumption during usage, and disposal of electronic equipment waste."*
>
> *"Second-order impacts are effects that result indirectly from using ICT, like energy and resource conservation by process optimization (dematerialization effects), or resource conservation by substitution of material products with their immaterial counterparts (substitution effects)."*

> *"Third-order impacts are long-term indirect effects on the environment that result from ICT usage, like changing life styles that promote faster economic growth and, at worst, outweigh the formerly achieved savings (rebound effects). These effects do not appear sequentially and disconnected. In reality they are nested, which means that second-order effects can only emerge on the basis of first-order effects and third-order effects can only appear as ramifications of second-order effects."*

Effective sustainability initiatives are needed as soon as possible, and environmental sustainability shall play a key role in doing business responsibly and successfully.

Up to now, much effort has been spent to address the environmental aspects of sustainability of computer hardware, but there is much to do in the field of computer software, and software process models. A software product, "Green and Sustainable", should have an economic, societal, and ecological impact, and an impact on human beings as small as possible over its whole life cycle. However, such a software product can be achieved only if all the various stakeholders recognize these impacts, and the whole developing organization is aware of negative and positive impacts that the usage of the software product will likely cause over its whole life cycle.

## 3.1 Background

In past years, many organizations have launched sustainability market initiative to improve environmental performance and environmental management. In addition, many banks are experimenting with the blockchain technology, betting on its ability to promote economic growth by freeing up trade, in order to speed up the rate of technological innovation, and its ability to lead to faster development of green technologies (see work by [78]).

The introduction of blockchain technology may provide substantial energy savings if it may take the place of some of the energy consumptive systems, services and locations that support the fiat currency [16]. Blockchain technology seems to have the potentiality to optimize the global financial infrastructure, dealing with global issues, such as sustainable development, or with asset transfers much more efficiently than current financial systems.

The financial sector incurs in many operative costs in order to efficiently run the whole system. These costs include time and money required to invest heavily in infrastructure, electricity costs to operate and from automated teller machines (ATMs), gas and water consumed by employees and waste produced.

In addition, no fiat currency can be created without costs. Periodically in order to guarantee the quality standards for the banknotes in circulation, the worn banknotes are shredded, so, to all operative costs just mentioned, the cost of production of coins and notes and those for the shredding systems have to be added to get an overview of the total cost of the actual financial system.

In contrast, systems based on blockchain technology have only to connect to the network and do not incur electricity costs such as those from ATMs, costs from gas consumed by employees or waste including, for example, paper and toner for printers. Furthermore, in these systems, the production cost of the cryptocurrency is included in the cost of mining activity that comprises also the costs of transaction validation, and, in turn, the distribution costs of new cryptocurrency. This, of course, implies substantial savings with respect to the traditional financial system. The mining activity is the process by which new bitcoins, or, in general, new crypto coins, are generated and new transactions are verified and added to the blockchain, the public ledger which stores the entire transaction history. Anyone who is connected to the Bitcoin network and owns suitable hardware can participate in mining and is called a "miner". In order to secure the network, by adding to the blockchain only the valid transactions, the participants have to solve a computationally difficult puzzle. Specifically, they have to find the so called "Proof of Work" (PoW) burning computational power on useless calculations. Whoever first solves the puzzle gets a reward in Bitcoins, and eventually gets the transaction fees associated with the transactions compiled in the block validated by her.

Two recent articles by [73] and by [35] explore the energy efficiency of Bitcoin. Malmo wrote: *"adopting Bitcoin as a major currency in the next few decades would just exacerbate anthropogenic climate change by needlessly increasing electricity consumption until it is too late"* (Ref. [73]).

Deetman is less pessimistic and categorical than Malmo. He discussed how hashing is related to mining hardware and hence to energy consumption, providing noteworthy "optimistic" and "pessimistic" energy forecasts. He makes some interesting plots comparing efficiency and product ship dates, and discusses mining trends and scalability. He ended his articles stating that *"Personally, I haven't given up on the idea of distributed network transactions, but a radical rethinking of how these may be secured would be beneficial, be it at least for the environment. Perhaps a system where all miners are rewarded for their pledged surplus in CPU processing power, but the actual hashing is performed only by a few thousand randomly selected and continuously changing CPUs, would be a solution."* (Ref. [35]).

In agreement with this last claim, in this work, we investigate the potentiality of the blockchain technology and the leading role that it could have in addressing the environmental aspects of sustainability. In our opinion, a future evolution

and deployment of this technology could revolutionize the banking system.

Blockchain technology, the shared ledger technology based on an open source distributed database, although still in its relative infancy has already triggered much interest, and a lively debate is ongoing about its future progression and the important benefits that could bring in the context of the transfer of assets within business networks.

In traditional business networks, the processes to underpin asset ownership and asset transfer are often inefficient, expensive and vulnerable. The blockchain technology could have, in a not too distant future, a transformative impact on central bank, financial institutions and technology firms.

It could follow in the footsteps of Internet technology, in which the government, industry and academia, beginning with early research on packet switching and the ARPANET (the first network to use the Internet Protocol), contributed to the evolution and deployment of the technology that revolutionized the computer and communication world like nothing before.

Blockchains have the potential to bring great value to several financial service activities, from trade finance to payments, securities settlement, and regulatory compliance. In addition, they could contribute to overcome some traditional banking inefficiencies, such as the foreign exchange (FX) transfer costs and times, to augment existing business networks, and to provide increased discoverability and trust working in cooperation with the banking payment and messaging systems.

A key prerequisite to reach such an interconnected system is achieving a standard way of implementing this technology. In future, we may have multiple ledgers, such as a foreign exchange network and a bond network, which need interoperability to function, just as the internet and intranets share the same technology.

Using blockchain in conjunction with actual banking systems will augment the power operating between counterparties. We potentially may have a common, ubiquitous blockchain, able to reduce the need for intermediaries to validate financial transactions and the friction created in financial networks due to different intermediaries, which often use different technology infrastructures. In theory, such an interconnected infrastructure has the potential to generate significant efficiency gains, reducing duplicative record keeping, eliminating reconciliation, minimising error rates and facilitating faster settlements (see work by [78]).

In addition, such an infrastructure would also be critical to underpinning a future “internet of things”. Every device connected to the internet becomes a potential user of banking services, and this infrastructure may enable offering services at much lower cost. Of course, blockchain mining protocols, at least as they look today, are not able to achieve the millisecond response times needed

by transactions on the internet of things. A future blockchain with millisecond latency could give devices autonomy, and allow them, for example, to transfer ownership of physical goods without the need to refer to a central management system.

Of course, there are many concerns about scalability, costs, and security to be overcome before blockchain technology moves to widespread usage. There is much concern about whether this technology will be able to achieve the processing speed of an automated clearing house, about the more computational power required to each participating block of a blockchain, and about the actual ability to lower costs compared to traditional payment systems when larger transaction volumes will be involved (see works by [78, 68]).

In this chapter, we focus on the role of financial and cryptocurrency markets in sustainable development, examining recent trends in banking sector, and possible future events that could shape the role of the blockchain technology in the sustainable development of an integrated financial and cryptocurrency market. We give significant insights about the efficiency of the actual Bitcoin system, showing that the efficiency of the Bitcoin system could increase only by overcoming some of its main limitations, such as the low number of transactions, the block size limit, and the high computational power.

Many works provide food for thought about the potentialities of blockchain technology that if exploited and advanced in an adequate way could bring valid support to the actual financial system, such as the works quoted above [78, 68] and a recent report by [3] that estimated that blockchain based systems could bring high potential cost savings. However, to the best of our knowledge, no work focuses on the efficiency of the actual Bitcoin system and on the limitations that hinder a widespread usage of the Blockchain technology, providing an empirical study of the economic and energetic footprint of the Bitcoin system, as we do.

For example, the two works quoted above [73, 35] explore the energy efficiency of Bitcoin. The former discuss about the unsustainability of the Bitcoin system claiming that the energy cost of a single Bitcoin transaction could power 1.5 American homes for a day. The latter discusses how hashing is related to mining hardware and hence to energy consumption, discusses mining trends and the scalability. However, both works do not discuss about the limitations that hinder a widespread usage of the blockchain technology in the banking system. In addition, we can cite the work by Vranken [130], who focused on the estimation of the power usage of the Bitcoin network, considering four families of mining hardware. He concluded that the order of magnitude of the energy power is 100 MW. Moreover, we can cite the work by Urquhart [129], who evaluated the economic performance of the Bitcoin system inferring that bitcoin returns are insufficient to cover the energy expenditure of mining operations. Previously, Hayes [51] describe the cost of production of one bitcoin, and O'Dwyer and

Malone [95] analysed the bitcoin production cost until 2014 .

All of these works do not provide an analysis of the economic and energetic footprint of the Bitcoin system focusing on the technological limits of the system that hinder the spread of the blockchain technology in the banking sector.

## 3.2 Banking Sustainability in the Fiat World

In banking sector, sustainability market initiatives operate in two key directions [43]:

> *"The pursuit of environmental and social responsibility in a bank's operations through environmental initiatives (such as recycling programs or improvements in energy efficiency) and socially responsible initiatives (such as support for cultural events, improved human resource practices and charitable donations)"*

> *"The integration of sustainability into a bank's core businesses through the integration of environmental and social considerations into product design, mission policy and strategies. Examples include the integration of environmental criteria into lending and investment strategy, and the development of new products that provide environmental businesses with easier access to capital"*

Sustainability strategies try to minimize impact on the environment, starting from making people more efficient, improved recording of environmental key performance indicators, efficient building technology, green travel to sustainable purchasing, and from end-to-end management of resources and waste.

A key concern for banking institutions is climate change and environmental protection to lower the total $CO_2$ emissions. By working together, banks, their employees, but also service providers and suppliers, can implement sustainability plans more efficiently. The main goals are to get the highest energy efficiency of buildings, employees paper consumption, business travel, but also the running of the cafeterias, where using local products, offering eco-friendly dishes, and working on ways to reduce water consumption should become a common practice. Moreover, in order to minimize environmental footprint, organic waste has to be recycled and converted into a clean source of energy, and renewable energies, to reduce direct and indirect $CO_2$ emissions, should be more extensively used.

In recent years, there have been many bank endeavours to improve environmental performance and environmental management, launching sustainability market initiatives focused on working together on key matters, such as the

development of a joint climate change strategy. In addition, many banks are experimenting with and implementing the blockchain technology, believing in and betting on its ability to lead to faster development of green technologies, in addition its ability to promote economic growth (see work by [78]). The most attention is undoubtedly focused on one of the most interesting aspects of blockchain, "the concept of smart contracts". Smart contracts, encoded in a programming language, are embedded in the blockchain and are executed with the transactions. They may be used, for example, to define the conditions under which the transfer of a bond occurs, giving rise to bond networks.

Thanks to all its potentialities, blockchain technology has triggered much interest and has given rise to several initiatives to advance it, such as the Linux Foundations Hyperledger project [45], the innovation hub blockchain and distributed ledger solutions by Hong Kong's central bank [114], the applications to move money across borders in real-time money using blockchain technology by several banks for examples Santander, UniCredit, Goldman Sachs and Barclays [137], and the initiative by BNP Paribas (Paris, France), the multinational bank, that is working on a blockchain platform in order to enable retail investors to lend money to businesses via an instrument known as a mini-bond[113]. A recent report by [3] estimated that blockchain based systems could bring a potential cost savings of 70% on central finance reporting due to the more streamlined and optimized data quality, transparency and internal controls, of 50% on business operations, such as trade support, clearance and settlement, due to a more efficient and effective clearance and settlement process, of 30–50% on compliance thanks to transparency and auditability of financial transactions, and of 50% on centralized operations due to more robust digital identities and mutualization of client data among participants.

Let us give some insights about the power consumption and carbon footprint, looking at one large financial service german provider, DZ Bank AG (Frankfurt, Germany) and at the whole US banking system.

DZ Bank AG is one of Germany's largest financial service providers. It employs approximately 30,000 people worldwide, of whom 27,800 work in Germany, has more than 1000 cooperative banks and 12,260 branches and over 30 million customers that attest to its importance (see work by [117]).

Since 2013, environmental data for all German offices (see report by [117]) have been collected. In 2015, data highlighted an electricity consumption of 25,520,138 kilowatt hours (kWh), and a heating consumption of 13,152,631 kWh in 2015. In 2015, a reduction in electricity consumption, leading to the drop in $CO_2$ emission, was registered thanks to the much better management of the electricity generated by hydroelectric plants.

As regards the total $CO_2$ emissions from electricity and heating, the water consumption, and the volume of waste, including printer and copier paper

consumption, envelopes, greeting cards, sympathy cards, toilet paper, electrical and toner lighting, and so on, −243,444 kilograms (kg), 91,109 cubic meters ($m^3$), and 534,907 kg were generated, respectively.

Concerning the carbon footprint and costs of the whole US banking system, let us cite an article by [77] entitled "Under the Microscope: The True Costs of Banking". This article describes the results of an analysis about the environmental impact of the world financial access points. The analysis, developed by the CoolClimate Network at the University of California, Berkeley estimated an impact expressed in million tonnes of $CO_2$/year equal to 383.1 for bank branches, and equal to 3.2 for ATMs, and an energy use expressed in GJ equal to 2.3 billion for bank branches, and equal to 18.9 million for ATMs. This article concludes by making a comparison with the Bitcoin system

> *"At 0.75 million tonnes of $CO_2$ produced per year, Bitcoin has 99.8% fewer emissions than the banking system".*

## 3.3 The blockchain and cryptocurrency world

Nowadays, many are the cryptocurrencies and their underlining blockchain technology present in the web, but undoubtedly the most popular are bitcoin and ether. Bitcoin system was created in 2009 by a computer scientist known as Satoshi Nakamoto whose real identity is not known (see work [88]). The Ethereum system was created very recently. It was initially described by Vitalik Buterin in late 2013 and was formally announced by him, in January 2014, at the The North American Bitcoin Conference in Miami, FL, USA [27].

Cryptocurrencies are based on distributed databases for their transactions, and hence on public or shared ledgers, which store the entire transaction history. These ledgers are called the blockchain, because transactions are bundled into blocks. Each block references a previous block, but the first block is called the genesis block.

Blockchain technology is designed as a decentralized peer-to-peer network and does not rely on a single central authority. It uses a broadcast network to propagate transactions and blocks. It broadcasts messages across a network using nodes. Each node has its own copy of the blockchain, which is synchronized with other nodes. No node knows a priori which version of the ledger is valid, and to secure the blockchain against attacks, the cryptocurrency network relies on precise algorithms, consensus mechanisms, such as the PoW in the Bitcoin network and the proof of stake (PoS) in the Nxt network. For a brief overview about the two main consensus mechanisms, PoS and PoW, see work by [15]).

Blockchain technology has triggered much interest around its future progression and the important benefits that it could bring in the context of the transfer

of assets within business networks. However, there are many concerns around the blockchain technology, such as its possible and future ability to achieve the processing speed needed for an automated clearing house, to lower costs compared to traditional payment system, and to contain the increase of wasted mining resources when larger transaction volumes will be involved (see works by [68, 78, 5]).

Looking at PoW as a general consensus mechanism of mining activity (see the next section for its detailed definition), we note many flaws that question its sustainability. The peril of 51% attacks, the ASIC (Application Specific Integrated Circuit) dominance and the high energy inefficiency are the most prominent concerns that could undermine the sustainability of Bitcoin system.

In blockchain technology, the transactions are almost instantaneous but their confirmation needs to be performed by miners, and the average time for the confirmation of a block depends on the consensus mechanism. Bitcoin validates one block every ten minutes, Nxt validates one block every few seconds and Ethereum one every minute [19]. This influences the maximum number of transactions per second (tps) achievable. Today, in Bitcoin system, there are on average 7 tps. In contrast, payment systems like Visa, Mastercard, and Paypal, can afford several thousand tps. For example, VISA handles on average around 2000 tps, and PayPal handles on average around 115 tps [14].

However, looking at the time to complete a transfer in a traditional international bank settlement network, such as Swift and SEPA (Single Euro Payments Area), it depends on the currencies involved, the payment method as well as bank holidays and weekends, and is within 1–4 working days. Banks do settlements between each other only once a day, not including weekends and holidays. In contrast, blockchain technology allows settlements between any different banks in 10 min around the clock, and seven days out of a week. To be fair, a Bitcoin user has to wait about one hour before he can consider its transaction confirmed. In fact, a new transaction can be considered confirmed only after at least five or six block are added in the blockchain. This because block generation process can provoke the creation of a short chain composed by orphan blocks. Orphan blocks are blocks added to the blockchain by a few nodes but that the majority of nodes do not take into consideration. For this reason, orphan blocks are quickly discarded from the blockchain also in nodes which at first considered them as new blocks. For this reason, the time required to have the certainty of confirmation is long at about one hour.

To secure blockchains against attacks, every cryptocurrency network relies on precise algorithms, such as the PoW in the Bitcoin network and the PoS in the Nxt network, and on specific mining hardware.

In the Bitcoin network, each node participating in mining is called a “miner” and has to solve a computationally difficult problem in order to confirm the

validity of newly mined blocks. The first node that solves the problem is rewarded with bitcoins. The probability of winning the reward and creating a block is proportional to the total computational power owned. Consequently, an attack against the blockchain is possible only if the attacker owns significant computational resources. The security of the network is supported by the cost of physically scarce resources, and this makes the network inefficient from a resource point of view. Specifically, specialized hardware is needed to run computations, and spending money on electricity is needed to power the hardware.

To increase the probability of winning the reward and creating a block, miners have to participate in an arms race (see work by [67] for more details), that makes prohibitively high the cost of a possible attack, but that makes at the same time the Bitcoin protocol ecologically unfriendly. As a result, alternative mechanisms of block mining that are much less resource intensive have been proposed. Even if the debate is lively and still ongoing, many are convinced that the introduction of the PoS as the consensus mechanism, in place of the PoW, would guarantee a long-term sustainability.

In the PoS algorithms, the probability of winning the reward and creating a block is proportional to a node's ownership stake in the network. The security of the network is guaranteed because, on the one hand, nodes with the highest stake have the most interest to keep the network secure, and, on the other hand, to mount a successful attack, one needs to acquire most of the currency, but this is prohibitively expensive.

PoS offers many advantages with respect to PoW as a mining method. Firstly, it is much more environmentally friendly than PoW. In fact, in order to secure the network, it does not require miners to burn computational power on useless calculations. Secondly, there are no centralization concerns. Indeed, in contrast with PoW, where mining has been essentially dominated by specialized hardware, and there is a large risk that a single large miner will take over and de-facto monopolize the market, PoS is CPU friendly in the long term [19, 18].

However, there are also some disadvantages in PoS. For example, the so called "nothing at stake" problem. Miners have nothing to lose by voting for multiple blockchain-histories. This is because, unlike PoW, the cost of working on several chains is small, and miners can attempt to double-spend (in case of blockchain reorganization) "for free" [19, 18].

Many have attempted to solve these problems. Peercoin uses centrally broadcasted checkpoints and no blockchain reorganization is allowed deeper than the last known checkpoints (see work by [60]). This system uses a combination of PoW and PoS. It was the first proof-of-stake based coin and was released by Sunny King in 2012. In the PeerCoin system, the PoS is based on a notion of coin age. Coin age of an unspent transaction output is its value multiplied by the

time period after it was created. A transaction spending a previously unspent output consumes, or destroys, its coin age (Ref. work by [15])).

Nxt system only allows to reorganize the last 720 blocks. Work by [94] presents a detailed description of Nxt, a 100% proof-of-stake cryptocurrency. Nxt system offers some interesting advantages with respect to the Bitcoin system, such as the potential for reliable instant transactions, increased security, and significant energy and cost efficiency improvements (see work by [34]). In addition, it allows for the processing of up to 367,200 transactions per day. Nxt is resistant to so-called nothing at stake attacks, and since the full token supply was distributed in the genesis block, when an account successfully creates a block, the transaction fees are awarded to that account.

Ethereum developers proposed Slasher protocol that allows users to "punish" the cheater, who mines on the top of more than one blockchain branch [19, 18].

Note that Ethereum was designed as a system based on a proof-of-work algorithm named Ethash, and Slasher was never adopted [19, 18].

Ref. [11] presented a hybrid mining protocol, based on a consensus mechanism called Proof of Activity (PoA), that relies both on PoW and PoS, and, as a result, takes advantage of the best properties of both consensus mechanisms, giving rise to a better system. Recently, Ref. [10] proposed the Chains of Activity (CoA) system, a pure PoS protocol based on the core element of PoA [11], which aims to overcome the problem of rational forks, caused by the network fragility if the nodes are more rational than altruistic.

Ref. [103] proposed SpaceMint, a cryptocurrency based on proofs of space (PoS), designed to lower setup and overhead costs with respect to the wasteful PoW and to have a fairer reward structure for all miners. The name of this proof stems from the fact that miners dedicate disk space rather than computation power.

In addition to the concern of the Bitcoin protocol being ecologically unfriendly, another concern regards the number of tps in the Bitcoin system today. As already mentioned, this system can do on average around 7 tps. In contrast, payment systems like Visa, Mastercard, Paypal can do several thousand tps. A possible solution, in order to overcome the scalability limitations and the speed of Bitcoin, and to experiment with new working models is that of adding one or more chains, called "sidechains", alongside the Bitcoin blockchain. Sidechains are an innovation proposed and developed by the startup Blockstream, that in early 2015 proposed its prototype sidechain, called "Elements". Such chains allow the creation of new blockchains "pegged" to the Bitcoin blockchain, and their protocols allow value transfer between sidechains and the Bitcoin blockchain that is automatically secured by the Bitcoin mining network. This allows a lower time to validate a block and a different consensus mechanism than those of the Bitcoin protocol. In addition, they allow for managing a more advanced

programming environment.

From this possible solution, many projects, such as Segregated Witness and Lightning Network, were generated. Segregated Witness is based on the general concept to separate transaction and signature data. In principle, this could introduce incompatibilities that would change the structure of blocks, causing a split in the Bitcoin network between upgraded nodes and non-upgraded nodes, and hence a hard fork. To avoid this problem, Segregated Witness was implemented by using a clever hack and was rolled out as a soft fork. Specifically, this clever hack marks the transactions as "anyone-can-spend" transactions for non-upgraded nodes, whereas upgraded nodes are redirected to an "add-on block" with signature data (for more details, see the articles by [134, 110]). On 15 November 2016, Bitcoin Core version 0.13.1 was released. It is the official introduction of Segregated Witness, which, if activated, enables a number of new features on the Bitcoin network, as well as an effective block size limit increase [133].

Concerning the Lightning Network, it is a decentralized system that allows payments to be securely routed across multiple peer-to-peer payment channels, solving some problems of the Bitcoin network. Specifically, it is a system for instant and high-volume micropayments that today are inconsistently confirmed and the fees render such transactions unviable on the Bitcoin network [56].

## 3.4 Bitcoin protocol performance

In this section, the actual performance of the Bitcoin system is discussed, with particular attention to its main limitations, such as its ecologically unfriendly consensus protocol, and hence the high computational power required to run the system, which implies high mining hardware expenses, and the low number of transactions, and then the block size limit.

### 3.4.1 Ecologically Unfriendly and Friendly Protocols: PoW vs. PoS

As already mentioned, in the Bitcoin network, miners have to run their mining hardware continually, proving that they are spending a substantial amount of money in order to secure the network. In exchange, they gain newly minted bitcoins. Contrary to Bitcoin, in a PoS system, such as Nxt, every one who owns stakes can be chosen to protect the network. The bigger the stake they own, the more often they are chosen to protect the network. With this mechanism, only one or possibly a few computers at a given time run on full power, processing

transactions and using energy for validating the transactions, and not for the sake of proving they exist, and spend a lot of money to secure the network.

Ref. [34] presented an interesting comparison between energy and cost efficiency of the Nxt and Bitcoin network. He computed the electricity and the hardware expenses of the Bitcoin network in May 2014. Instead, for the Nxt network, he computed these expenses considering a hypotethical Nxt size equal to that of the Bitcoin network in May 2014.

He analysed the energy and cost efficiency of the Nxt network under the hypothesis that 2500 Cubietrucks will power the network when it gets to Bitcoin's size. A Cubietruck is a forging machine that offers a 1.2 to 1.6 GHz dual core processor. Its value of power consumption, when idle, is equal to 3 W, whereas, at full power, is equal to 18 W. Correctly, he assumed that these forgers use up to 18 W while forging, and consequently that a number of network users equal to 2497 uses only 3 W while idling. As a result, in [34], the total power at a specific time in order to secure the network is equal to 7545 W, and hence equal to 181 kWh per day about 66 MWh per year. Assuming an average rate of 12 cents per kilowatt hour, Czarnek computed a cost of electricity to power the network per day equal to $7937 per year. He also computed the cost of hardware per year, and assuming a cost equal to $100 and a 5-year lifetime for Cubietrucks, he found a cost of hardware per year equal to $50,000.

Regarding the Bitcoin system, he computed the total power consumption on 24 May 2014, starting from the value of the hash rate at that date, equal to approximately 99,300,000 GH/s, and hypothesizing that all miners used on that date the best machines available on the market. Specifically, he picked the Cointerra TerraMiner II, which runs at 1000 GH/s and costs $3500. Assuming hence that 99,300 Bitcoin miners powered the network, he found a power consumption per year equal to about 520,000 MWh, a cost of electricity to power the network per year equal to $62,400,000, and a cost of hardware per year equal to $69,510,000 considering, as in the previous computations, an average rate of 12 cents per kilowatt hour and a 5-year lifetime for mining hardware.

Although the author considered the most energy and cost efficient machines in the market, he highlighted a difference of four orders of magnitude between the Bitcoin and Nxt system, highlighting hence a much higher efficiency of the system using PoS than that using PoW.

Results similar to those by Czarnek for the Bitcoin system emerge also from other works, such as that by [67], who simulate an artificial Bitcoin market, and that by [33], who wrote:

> *"In April 2013 it was estimated that Bitcoin miners already used about 982 Megawatt hours every day. At that time the hash rate was about 60 Tera Hash/s."* See article by [76]).

Adopting the same approach used by [67], based on the fitting curve of the hash rate per US$[H/(s*$)], $R(t)$ and on that of the power consumption [W/H/s] $P(t)$ (defined in the next section), in this work, we compute the electricity and hardware expenditures supported by the Bitcoin mining network over time, from 30 September 2010 to 31 December 2016. We estimated these expenditures dividing the real total hash rate in the network by $R(t)$. The real total hash rate data was recovered from the blockchain Web site.

Figure 3.1 shows these expenditures over time in a logarithmic scale. It highlights hardware expenditures increasing over time until 4 October 2014. Then, this increasing trend ends and the hardware expenditures range between $100 million and $382 million, this last value being the highest value reached exactly on 30 October 2014.

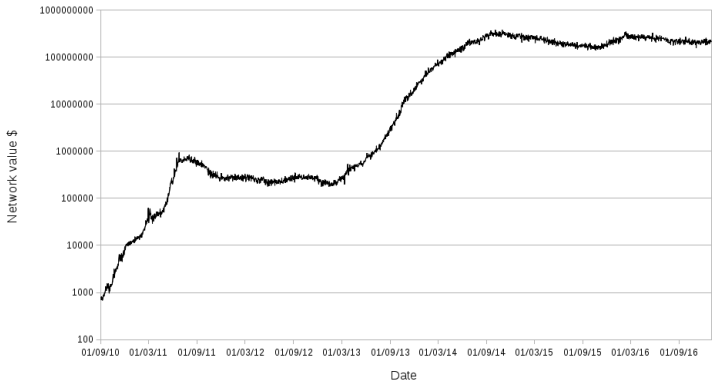


Figure 3.1: Total estimated investment to create the Bitcoin network .

The electricity expenditures incurred by the network to fuel the Bitcoin system, and hence the power consumption attributable to the Bitcoin system, shows a very similar trend. This is because these expenditures are also computed dividing the real total hash rate in the network by the power consumption $P(t)$.

Using the fitting curve of the power consumption $P(t)$, we estimated the power consumption of Bitcoin system from 1 September 2010 to 31 December 2016. Figure 3.2 shows this power consumption in a logarithmic scale, and Figure 3.3 expands the $x$-axis to highlight the power consumption from 1 October 2015 to 31 December 2016.

Figure 3.2b shows a power consumption increasing over time until 4 October 2014. On this date, the estimated power consumption was equal to 355.46 MW. Starting from September 2014, this increasing trend ends, and the power consumption ranges between 100 MW and 200 MW (see Figure 3.3).

Figure 3.4 shows the annual energy consumption expressed in kWh, from 2011 to 2016. It shows the decreasing trend of the energy consumption in the

last three years. This is in agreement with the introduction on the market of mining hardware more and more efficient.

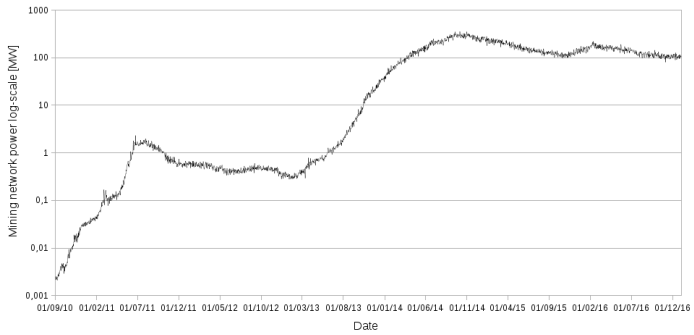


Figure 3.2: Estimated power consumption of Bitcoin system .

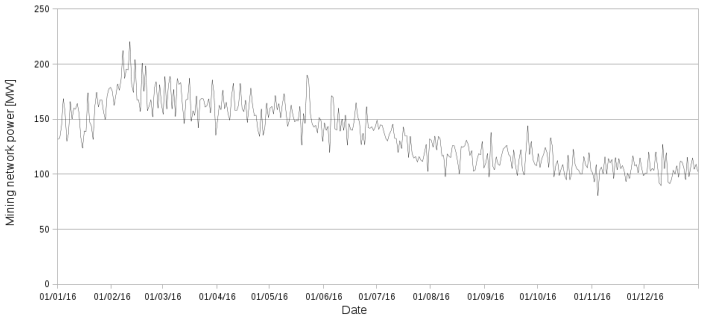


Figure 3.3: Estimated power consumption of Bitcoin system, from 1 October 2015 to 31 December 2016.

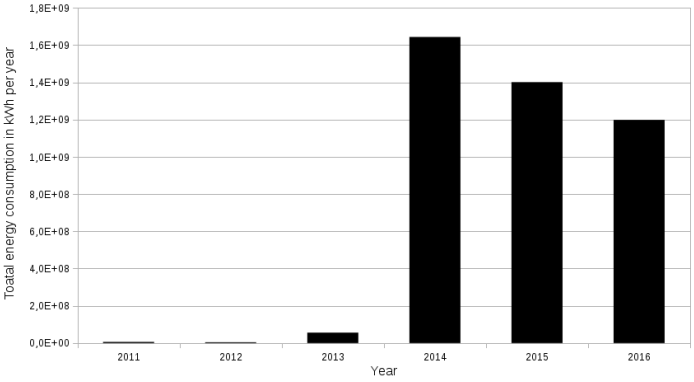


Figure 3.4: Total energy consumption per year.

All figures just described highlight that the Bitcoin system, as every system

using PoW, an ecologically unfriendly consensus mechanism, incurs high electricity and hardware expenses in order to increase the probability of mining bitcoins by buying hardware more and more powerful.

Despite of this, as already mentioned in the Section 3.1, all systems based on blockchain technology, both those using PoW and those using PoS, only have to connect to the network and do not incur such electricity costs from ATMs, in costs from gas consumed by employees or in waste including for example paper and toner for printers. Furthermore, in these systems, the production cost of the cryptocurrency is included in the cost of mining activity that comprises also the costs of transaction validation and in turn the distribution costs of the cryptocurrency.

Regarding the carbon footprint of Bitcoin system, let us cite an article entitled "Does Bitcoin Have an Energy Problem?" by [22], that gives an estimation of the million tons of $CO_2$ created by the Bitcoin system when all Bitcoin will be mined.

> *"The total circulation of bitcoin is capped at 21 million, at which point there will be no more mining. Currently, there are just over 14.7 million in circulation. That leaves 6.3 million to be mined. At a cost of $150 a coin and 1.5 tons of $CO_2$, it will cost nearly a billion dollars and create over 9 million tons of $CO_2$ just to produce the remaining bitcoins. If we assume that all bitcoins were mined as cheaply as $150 a coin, then it cost $3.1 billion to pay the electricity costs to put all those coins in circulation. It would also have created 31.5 million tons of $CO_2$".*

This article refers to an article by [109].

The actual Bitcoin system, in agreement with this article, has an impact expressed in million tonnes $CO_2$/year equal on average to 1, considering a period of 31 years, from 2009 to 2040, 2009 being the year in which the Bitcoin system appeared and 2040 the year in which the system will reach the Bitcoin's cap set at 21 million coins.

### 3.4.2 Efficiency

In order to evaluate the efficiency of the actual Bitcoin system, we defined three quantities, "economic efficiency" (EE), "operational efficiency" (OE), and "service efficiency" (SE), starting from the general definition of OE in a business context, defined as the ratio between the input to run a business operation and the output gained from the business. All of these efficiency measures are heavily affected by the features of the Bitcoin protocol, and specifically from the consensus mechanism unfriendly ecologically, and by the block size limit.

Consequently, only future advances in the Bitcoin system, and in general in blockchain technology, will be able to yield a higher efficiency, allowing us to create efficient blockchain based systems. In order to be able compute these measures, we started by gathering information about the mining hardware that entered the market over time, as in work by [67]. As already mentioned, the people who confirm transactions of bitcoins and store them in the blockchain are called "miners". The first miner who finds a proper hash (he finds the "proof-of-work"), gets a reward in bitcoins, and the successful hash is stored with the block of the validated transactions in the blockchain. Producing a single hash is computationally very easy. Consequently, in order to regulate the generation of bitcoins, the Bitcoin protocol makes the computational complexity of the process needed to find the proof-of-work more and more difficult over time. As a result, we have witnessed the succession of four generations of hardware, i.e., CPU's, GPU's (Graphics Processing Unit), FPGA's (Field Programmable Gate Array) and ASIC's generation, each of them characterized by a specific hash rate (measured in H/s) and power consumption. Over time, the different mining hardware available was characterized by an increasing hash rate, a decreasing power consumption per hash, and increasing costs.

Starting from the gathered information about mining hardware, we computed the average of Hash Rate and of Power Consumption over time (see Table 3.1).

| **Date** | **Simulation Step** | **Average of Hash Rate** $\frac{GH}{s*\$}$ | **Average of Power Consumption** $\frac{W}{GH/s}$ |
|---|---|---|---|
| 1 September 2010 | 1 | 0.0017 | 454.87 |
| 29 September 2011 | 394 | 0.0014 | 19.8 |
| 2 December 2011 | 458 | 0.00175 | 34.4 |
| 28 December 2011 | 484 | 0.0017 | 72.575 |
| 1 May 2012 | 608 | 0.0029 | 72.575 |
| 17 December 2012 | 835 | 0.03565 | 1 |
| 10 April 2013 | 953 | 0.0194 | 6 |
| 31 May 2013 | 1004 | 0.0201 | 6 |
| 15 October 2013 | 1141 | 0.1351 | 3.84 |
| 10 December 2013 | 1197 | 0.0595 | 3.84 |
| 22 January 2014 | 1240 | 0.245 | 2 |
| 4 July 2014 | 1403 | 0.583 | 1.1 |
| 23 October 2014 | 1513 | 1.6 | 0.69 |
| 30 August 2015 | 1824 | 2.756 | 0.51 |
| 1 December 2015 | 1918 | 2.666 | 0.249 |
| 1 May 2016 | 2070 | 4.746 | 0.273 |
| September 30, 2016 | 2221 | 8.465 | 0.099 |

Table 3.1: Average of Hash Rate and of Power Consumption over time.

We fitted a "best hash rate per $" and a "best power consumption function" and called the fitting curves $R(t)$ and $P(t)$, respectively.

We used a general exponential model to fit the curve of the hash rate, $R(t)$. It

is defined as:

$$R(t) = a * e^{(b*t)}, \tag{3.1}$$

where $a = 6.712 \times 10^6$ and $b = 0.003204$.

We used a similar curve also to fit the curve of the power consumption $P(t)$. It is defined as:

$$P(t) = a * e^{(b*t)}, \tag{3.2}$$

where $a = 4.636 \times 10^{-7}$ and $b = -0.004005$.

Note that the values of the coefficients $a$ and $b$ stem from the computation of the best exponential fitting curve of the hash rate for Equation (3.1) and of the average power consumption for Equation (3.2).

### Economic Efficiency

We defined the "economic efficiency" (EE), as the ratio between the value of bitcoins expressed in US$ mined by the power consumption of 1 kWh.

In order to compute this quantity, data about the number of bitcoins generated over time and the bitcoin price were recovered from the "`blockchain.info`" web site. Data about the power consumption are computed by using the fitting curve defined in Equation (3.2), and data about the real hash rate are recovered from the "`blockchain.info`" web site.

Figure 3.5 shows the trend of economic efficiency over time. We can observe that EE reached the highest values in the period between April and August 2013. In particular, the economic efficiency reached its highest value, exactly equal to US$63.47 per kWh, on 9 April 2013.

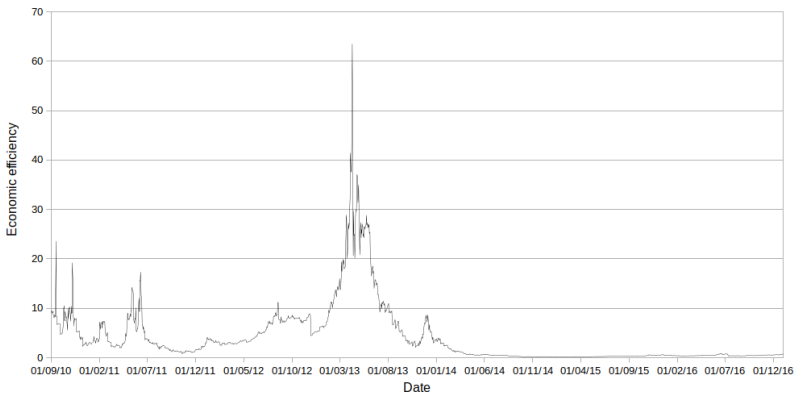


Figure 3.5: Economic efficiency expressed in US$ per kWh from 1 September 2010 to 31 December 2016.

The trend of EE is strictly linked to the rapidly growing interest in the Bitcoin system that steadily drove the prices higher and higher. Bitcoin price, starting from negligible values, reached values of about $140 in April 2013 and about $1000 in November 2013. This price increase caused the growing trend of EE until 10 April 2013 and also its fall from 10 April onwards. In fact, such a huge interest brought an arms race for acquiring efficient and specialized mining hardware. From 10 April 2013 onwards, EE started to decrease due to the dramatic increase of the total hash rate and power consumption also, in conjunction with the halving of the bitcoin mining prize, which halved two times in this period. Specifically, Bitcoin's block reward was halved the first time, from 50 to 25 bitcoins in November of 2012, and the second time, from 25 to 12.5 bitcoins on 9 July 2016. Figure 3.6 shows the trend of the EE, expanding the $x$-axis to highlight its values from 1 October 2015 to 30 September 2016. During this period, the EE ranges between 0.3 and US$0.85 per kWh. In this figure, it is also possible to observe the effect of the last halving on the trend of EE, which falls sharply on that day.


Figure 3.6: Economic efficiency expressed in US$ per kWh from 1 October 2015 to 31 December 2016.

Note that the EE is also strictly linked to the energy cost, and indeed the majority of hashing power of Bitcoin network is concentrated among a handful of Chinese mining pools [46], given that China is one of the countries where there are the lowest energy costs. Thus, if we take into account the variable component of the energy cost for Chinese industrial consumers, which ranges between 0.0525 and 0.0825 US$/kWh (0.35 a 0.55 Y/kWh). Nnote that, in China, the variable component of the energy cost has to be added to the fixed component, which depends on the stipulated contract. and compute the profit per kWh, we obtain a value that ranges between $0.2475 and $0.7675. Compared to China, in Italy (which is the European country with the highest electricity price), the

average energy cost for non-domestic users is equal to 0.2119 US$/kWh, and, as a result, the profit per kWh is much lower, exactly between $0.0881 and $0.6381.

### Operational Efficiency

We defined the "operational efficiency" (OE), as the ratio between the value of voluntary fees and the energy cost of a transaction. In general terms, it is defined as the ratio between the output gained by a business and the input to run a business operation. We defined this efficiency as the ratio between the value of the voluntary transaction fees and the energy cost of a transaction. The transaction fees are the fees paid to the miner who validates the block that includes that transaction. They are voluntary and are an incentive for miners in order to include a transaction into the next block. However, a miner can accept a transaction and include it in the new block also without any reward in return. Thus, a person posting a bitcoin transaction can include any fee, or none at all, in the transaction.

We computed this efficiency using the monthly average of total daily transaction fees, the daily energy consumption obtained through Equation (3.2) and the real data about the hash rate. Figure 3.7 shows the trend of OE.

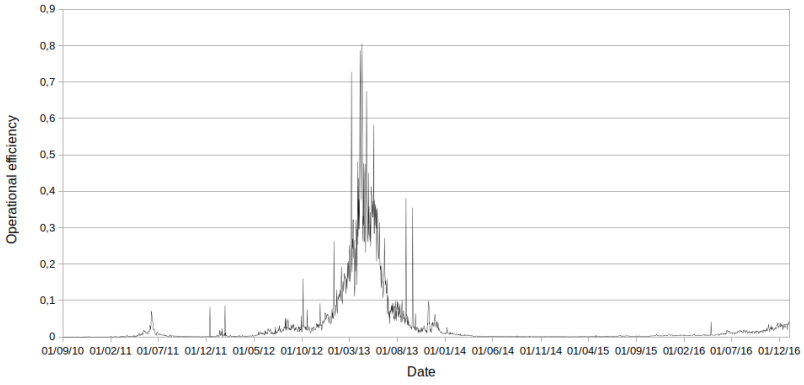


Figure 3.7: Operational efficiency from 1 September 2010 to 31 December 2016.

Its value increased until 10 April 2013, when OE reached its highest value, 80.6 US$ cents per kWh, but then it started to decrease.

Figure 3.8 shows the trend of the OE, expanding the $x$-axis to highlight the OE from 1 October 2015 to 31 December 2016. Note that, after the period in which OE decreases, from about July 2015 onwards, the OE started to slowly increase.

This increasing trend seems to follow the importance that the fees have over time. When the number of bitcoin generated will approach the value of

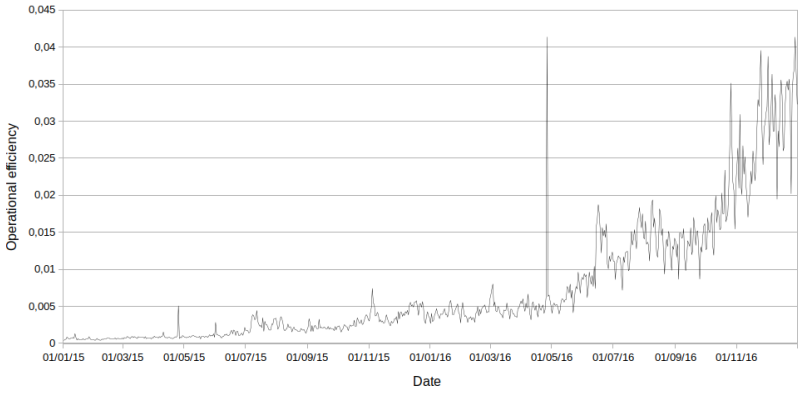


Figure 3.8: Operational efficiency from 1 October 2015 to 31 December 2016.

21 million, there will be no fixed mining reward anymore, and the only mining reward will be that associated with the transaction fees. As a result, only a growing OE trend will guarantee the survival of the Bitcoin system.

### Service Efficiency

We defined the "service efficiency" (SE), as the ratio between the number of transactions validated by the power consumption of 1kWh. In order to compute the SE, data about the number of transactions validated over time were recovered from the "`blockchain.info`" Web site, and data about the power consumption were computed as described for EE in Section 3.4.2.

Figure 3.9 shows the SE over time and Figure 3.10 shows its value limiting the max $y$-axis to 10 transaction per kWh. Until September 2013, the SE ranged between one and 10 transactions per kWh. Then, SE drastically decreased, keeping its values always under one transaction per kWh. The worst estimated SE dates back to 4 October 2014, when it had a value equal to 0.0098 transaction per kWh. From 1 October 2015 to 31 December 2016 (see Figure 3.11), SE ranges from 0.04 to 0.14 transaction per kWh.

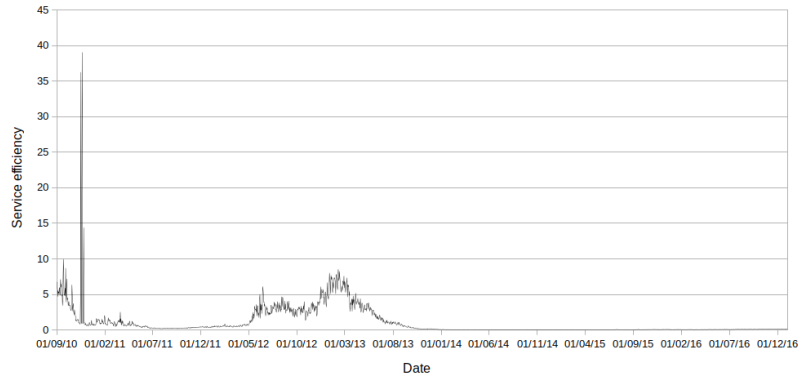


Figure 3.9: Service efficiency expressed in number of transactions per 1 kWh from 1 September 2010 to 31 December 2016.

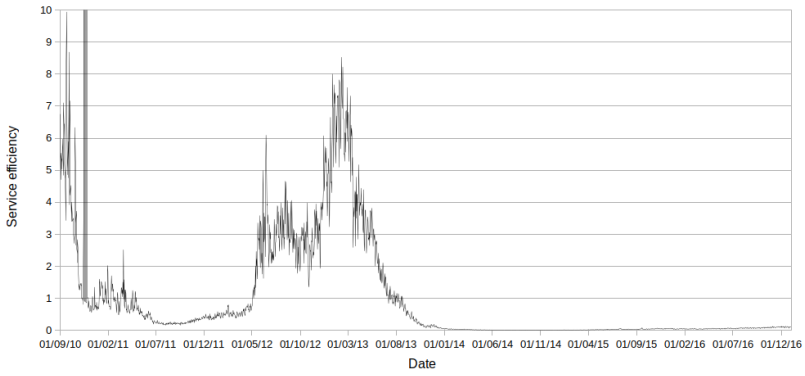


Figure 3.10: $y$-axis zoom of the service efficiency expressed in number of transactions per 1 kWh from 1 September 2010 to 31 December 2016

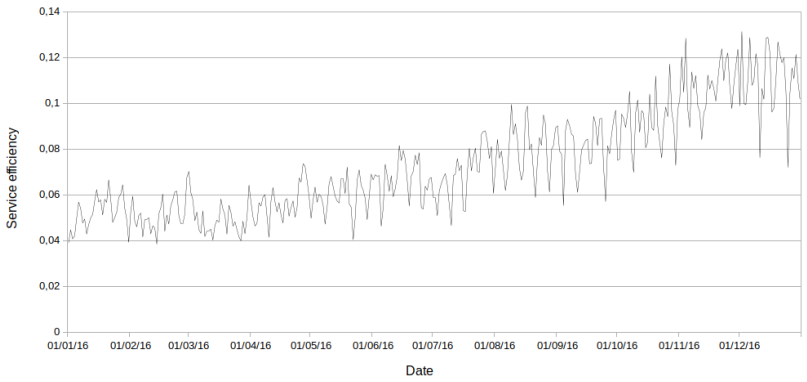


Figure 3.11: Service efficiency from 1 October 2015 to 31 December 2016.

Note that the block size limit and the time interval between blocks are two limitations that compromise the service efficiency of the Bitcoin system. When we reach the block size limit, the number of transaction per block cannot increase anymore. As a result, the energy consumption per transaction will increase whenever a new miner machine will be added to the network, and hence it will increase when the hashing capability of the network increases.

Figure 3.12 shows the number of transactions per block, which is steadily increasing. The increasing trend ended started from about September 2015, when the number of transactions approached the imposed block size limit, equal to 1 MB.

In agreement with the considerations made in the beginning of this chapter, the performed analysis confirms that the efficiency of the Bitcoin system, and hence the proposed efficiency measures could increase only by overcoming some of the Bitcoin system's main limitations, such as the low number of transactions, and then the block size limit, and the high computational power. They do not aim to demonstrate that the actual Bitcoin system is more efficient than the actual financial system but only to provide food for thought about the potentialities of blockchain technology that, if exploited and advanced in an adequate way, could bring a valid support to the actual financial system. Consequently, the research activity should move in this direction, increasing the number of transactions per block and decreasing the computational power required to run the system. Only overcoming these limitations can the introduction of the Bitcoin system, and, precisely, the introduction of the blockchain technology into the actual financial infrastructures, allow us to deal with global issues much more efficiently than current financial systems.

In a nutshell, all of our results show that the overall efficiency of the Bitcoin system can increase only after overcoming its main limitations: the low number of transactions per block and the too high computational power that it

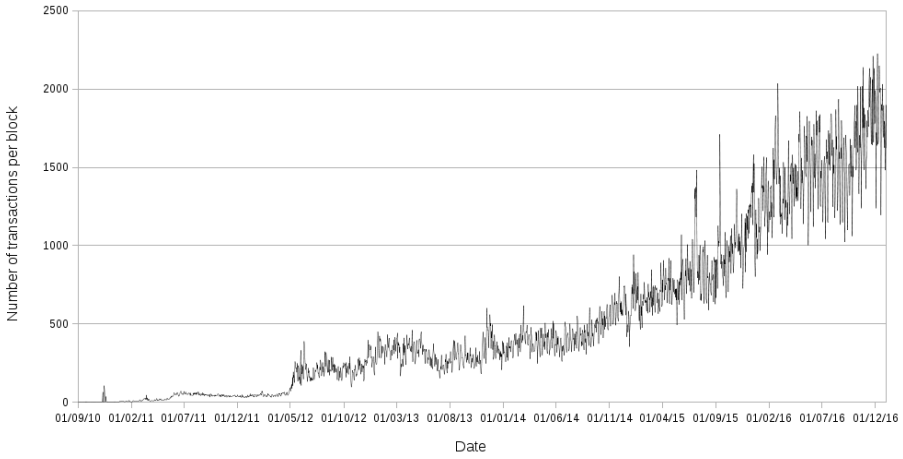


Figure 3.12: Daily number of transactions per block.

currently needs. In conclusion, our work provides a reflection on the potential of blockchain technology that could bring good support to the financial system. By sorting out the highlighted problems, the introduction of the Bitcoin system and, more in general, the introduction of blockchain technology to financial infrastructures could allow for addressing financial issues much more efficiently than current financial systems.

# Part II

# Blockchain engineering and applications

# Chapter 4

# Blockchain-oriented Software Engineering

This chapter concerns the need for software engineers to devise specialized tools and techniques for blockchain-oriented software development. Examining history and technical information about blockchains and related software systems, we could see that this technology is spreading in several application field. The blockchain seems to have what it takes to bring a disruptive innovation in the development of new software systems. In fact, all the software development revolving around the Blockchain technology is growing at a staggering rate, like the ones we will present in the next chapter,

To start with, we should take in consideration that in the past years, a lot of attention has been paid to the emerging concepts of Blockchain and Smart Contract. Organizations such as banking and financial institutions, and public and regulatory bodies, started to explicitly talk of the importance of these new technologies. According with several observers, blockchain represents the dawn of a new era [124, 128].

As discussed in the previous chapter, the majority of interest is given to the economic and financial perspective. The overall capitalization of digital currencies is over 410 Billions USD, as of March 2018, 42% of which is the capitalization of bitcoins alone[1]. Venture capital investments in blockchain startups has been steadily increasing, from $93.8 Million in 2013, to $315 Million in 2014, to $490 Million in 2015. In particular, during the 2017 we assisted to the explosion of the ICO (Initial Coin Offering) phenomenon. An ICO is a new and more democratic opportunity to found a startup. The total raised amount during this year was over 3.3 billion dollars[2]. By comparison, in 2016 ICOs raised a total of 106 million

[1] https://coinmarketcap.com/

[2] according to icowatchlist.com data

dollars[3]. An overview on this phenomenon, realized in collaboration with the author, will be presented in the XP2018 conference [54].

Furthermore, we can observe that the blockchain technology has a disruptive nature, such to remind the Internet of about 1993, when huge amounts of venture capital started to flow into Internet startups, leading to the emergence of companies such as Cisco, Yahoo, Google, Amazon and others. And, as we will discuss in the following chapter, this technology open the way to the creation of new typologies of services, based on the decentralization, the transparency, and to the trust in the (peer-to-peer) system.

But this great interest has also attracted malicious people. In fact, ever since digital currencies started to represent a real monetary value, also hacks and attacks started. To date, the most important victims have been the "exchanges", Web sites allowing to store digital currencies and to trade them against other currencies. The biggest was the MtGox attack that occurred at the beginning of 2014, leading to a declared loss of $600 Million. Most recent is the Bitfinex attack, happened during the August 2016, and featured $65 Million theft. Considering the world of smart contract, we can consider the attack sustained by the DAO, a decentralized organization hosted in the Ethereum blockchain, in June 2016, with an illicit withdrawal of Ether funds worth $50-60 Million [121]. In this case, thanks to a global agreement, the Ethereum community solved the problem performing a hard fork of the Ethereum system, that forcibly recovered the stolen Ethers and gave them back to the original owners. But in general, several are the security vulnerability of smart contracts[7]. Seeing that the increased interest in blockchain technologies lead to software projects fastly born and developed, vulnerability to attacks can be attributed to a poor software design and development practices affecting the final software quality.

This chapter discusses and traces motivation and the new directions that characterize the Blockchain Oriented Software Engineering, which is born from the exigency to take into account the need for novel specialized software engineering practices for blockchain-based software projects. We discussed results presentd in this chapert during the 39th International Conference on Software Engineering conference (ICSE2017) [107].

## 4.1 Blockchain-oriented Software Engineering: Challenges

A blockchain-oriented software is a software that includes the blockchain technology in its architecture. So, a blockchain-oriented software works, directly or

[3]https://www.coindesk.com/2016-ico-blockchain-replace-traditional-vc/

not, with an implementation of a blockchain.

Considering the distinctive features of the blockchain technology, software engineers could benefit from the application of blockchain specific engineering practices. For the purposes of this Chapter, a blockchain is a data structure characterized by the following key elements:

- data redundancy (each node has a copy of the Blockchain);
- check of transaction requirements before validation;
- recording of transactions in sequentially ordered blocks, whose creation is ruled by a consensus algorithm;
- transactions based on public-key cryptography;
- possibly, a transaction scripting language.

The most relevant BOSE challenges can be described in five issues. These are the security and reliability issues, the definition of a new architecture, the exigence of specific modeling languages, the study of software metrics, and the need of new professional roles. For most challenges, to properly frame the related issues, excerpts from the SWEBOK[4] [2] are provided.

**Security and reliability**. *"Software Security Guidelines span every phase of the software development lifecycle"* and *"Software Reliability Engineered Testing is a testing method encompassing the whole development process"*

A Blockchain must guarantee data integrity and uniqueness to ensure Blockchain-based systems are trustworthy. Ensuring security and reliability in BOS development might require specific methodologies such as Cleanroom Software Engineering [83] or thorough software reviews. Furthermore, mathematically sound analysis techniques could help enforcing reliability and security-related properties in blockchain-oriented applications.

Testing techniques can also enhance system security and reliability. In this regard, IBM recently expressed the need for continuous testing techniques to ensure blockchain software quality[5].

In addition, testing techniques should be based on the nature of the application which, in the case of BOS, is that of security-critical systems. In particular, there is a need for testing suites for BOS. These suites should include:

- Smart Contract Testing (SCT), namely specific tests for checking that smart contracts i) satisfy the contractors' specifications, ii) comply with the laws of the legal systems involved, and iii) do not include unfair contract terms.

[4] SWEBOK 2004 version

[5] https://twitter.com/ibmsoftware/status/776605297037172736

- Blockchain Transaction Testing (BTT), such as tests against double spending and to ensure status integrity (e.g. UTXO[6]).

**Software architecture**. Specific design notations, macroarchitecture patterns, or meta-models may be defined for BOS development. To this purpose, software engineers should define criteria for selecting the most appropriate blockchain implementation, evaluating the adoption of sidechain [9] or cross-chain [92] technology, or the implementation of an ad-hoc blockchain. For example, Ethereum[7] has adopted a key-value store, which is a very simplistic database. By adopting a higher level data representation such as an Object Graph, it would be possible to speed up many operations which would otherwise be expensive using a key-value store [66].

**Modeling languages**. Blockchain-oriented systems may require specialized graphic models for representation. More specifically, existing models might also be adapted to BOS. UML diagrams might be modified or even created anew to account for the BOS specificities. For example, diagrams such as the Use Case Diagram, Activity Diagram, and State Diagram could not effectively represent the BOS environment.

**Metrics**. BOSE may benefit from the introduction of specific metrics. To this purpose, it could be useful to refer to the Goal/Question/Metric (GQM) method, that was originally intended for establishing measurement activities, but it can also be used to guide analysis and improvement of software processes [2, 39, 38].

Due to the distributed nature of the Blockchain, specific metrics are required to measure complexity, communication capability, resource consumption (e.g. the so-called gas in the Ethereum system), and overall performance of BOS systems.

**New professional roles**. *"A recognized profession entails specialized skill development and continuing professional education"*

Due to the business-critical nature of the Blockchain, finance and legal subjects have shown increasing interest toward BOS. At the same time, bootcamps for Blockchain developers are flourishing. The Blockchain sector will need professional figures with a well-defined skills portfolio comprising finance, law, and technology expertise. An example of a new role could be that of an intermediary between business-focused contractors with low technology expertise and IT professionals.

[6] Unspent Transaction Output
[7] https://www.ethereum.org/

## 4.2 Blockchain-oriented Software Repositories

In order to define new research directions for the BOSE on the basis of the state-of-practice of blockchain-oriented software, we conducted an exploratory study on a corpus comprising 1184 GitHub software repositories, which were identified with the use of the Moody's Blockchain Report [21] and the CoinMarketCap website. First of all, the most relevant projects and players related to the blockchain technology were identified. Then, metadata on the corresponding BOS repositories were collected. Information from the corresponding issue tracking systems were also considered.[8]. In the remainder of this paragraph, details about the methodology used to build the BOS corpus, and the preliminary obtained results are provided.

### 4.2.1 Building a Dataset of Blockchain-oriented Software

We define a BOS project as a software project which contributes to the realization of a blockchain based system. This definition includes both blockchain platforms, such as Bitcoin and Ethereum, and other typology of blockchain based software [53].

To identify BOS repositories we start from the corresponding blockchain projects. Moody's Investor Services recently identified more than 120 publicly announced blockchain projects in an in-depth report of the blockchain sector [21]. The projects list covers rated issuers across financial institutions, nonfinancial corporates and the official sector, and can be considered as a comprehensive list of blockchain projects going on in the world. The Moody's list is not a list of software projects; nevertheless, BOS projects stem from the blockchain projects on the list.

In addition to the Moody's list, we searched for the software associated to the currencies and assets with the highest capitalization, as reported by CoinMarketCap. Since we took the Moody's list as a baseline, we did not include currencies and assets with a lower capitalization than Stellar, the least capitalized cryptocurrency in the Moody's list for which we found a related software repository. Being Stellar on the 17th position at CoinMarketCap, we focused on the first 17 most capitalized currencies and assets.

Finding the software corresponding to a project in the Moody's list is not straightforward. When the project name is within a list entry (e.g., The Hyperledger Project), the software can be easily found by searching for the specified project name. When not specified, we searched for the involved blockchain startups, which often choose to publish their software on code-hosting platforms

[8] We focused on Github-hosted projects, which come with the integrated GitHub issue tracking system

(e.g. GitHub, Bitbucket). As for the currencies and assets found on CoinMarket-Cap, this process was easier since each list entry is linked to the official website.

We focused on freely accessible, open-source software hosted on GitHub, a platform hosting the vast majority of the detected blockchain-oriented software projects. We decided to only consider software hosted on GitHub repositories because GitHub provides homogeneous metadata, which, in turn, allow us to compare projects on the basis of standard features. For instance, it is possible to evaluate project popularity by relying on the amount of stars given to a project by GitHub users, or on the number of forks stemming from it.

### 4.2.2 Dataset Analysis

At the end of the selection process, we identified 52 GitHub accounts, which comprise 1184 repositories. We extracted information on popularity (*Stargazers*), programming languages, community involvement (*Contributors*, *Open Issues*, *Watchers*, *Forks*), and age (time elapsed since creation).

To focus on the most relevant repositories, we only considered those that i) are base repositories (not a fork from a previously existing repository), ii) had been updated in the previous 30 days, and iii) were created more than 30 days before (i.e. have been modified at least once since creation)[9]. By using these criteria, we retained 193 repositories out of the initial 1184.

### 4.2.3 Analysis of the BOS repositories

The blockchain-oriented software are written in several programming languages. Figure 4.1 reports the most used programming languages among the 193 retained repositories. JavaScript, Python, Go, C++, and Ruby are the top 5 languages, with BOS JavaScript repositories accounting for more than 30% of the total. It is interesting to note the presence of Python and Go in the podium, especially in comparison with the number of Java repositories that does not reach the 4% of the total.

Analyzing the popularity of the repositories, we can observe that generally, the most popular projects are related to the most used blockchain-oriented systems. Table 4.1 shows that among the top 10 most popular repositories (i.e. those with the highest number of *Stargazers*) `ethereum/mist` and `coinbase/toshi` were created less than 1 year and roughly 2 years ago respectively. As expected, Bitcoin is the most popular project, neatly distinguished as for stargazers (9966) and contributors (396). Ethereum is also very popular, with three associated repositories in the top 10; in particular, the one written in Go is just behind the

[9] All data were retrieved on September 23, 2016

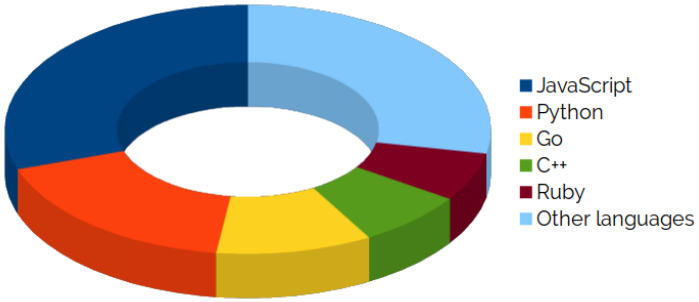


Figure 4.1: Languages across 193 repositories

| GitHub Repository | stargazers | contributors | open issues | age (days) | watchers | forks | first language |
|---|---|---|---|---|---|---|---|
| bitcoin/bitcoin | 9966 | 396 | 547 | 2105 | 1211 | 4266 | C++ |
| ethereum/go-ethereum | 2160 | 78 | 285 | 1002 | 367 | 695 | Go |
| ledger/ledger | 1813 | 108 | 14 | 3055 | 103 | 255 | C++ |
| digitalbazaar/forge | 1584 | 41 | 137 | 2260 | 103 | 241 | JavaScript |
| ripple/ripple-client | 1244 | 51 | 21 | 1437 | 968 | 486 | JavaScript |
| ethereum/mist | 1168 | 35 | 198 | 471 | 210 | 299 | JavaScript |
| dogecoin/dogecoin | 1153 | 300 | 52 | 1022 | 149 | 505 | C++ |
| ripple/rippled | 1144 | 53 | 118 | 1782 | 246 | 338 | C++ |
| coinbase/toshi | 839 | 18 | 97 | 749 | 98 | 187 | Ruby |
| ethereum/cpp-ethereum | 723 | 89 | 212 | 1001 | 196 | 270 | C++ |

Table 4.1: Extracted statistics across the top 10 BOS Repositories

main Bitcoin repository. The top 10 BOS repositories were created around 4 years ago on average, and most of them have a considerable number of open issues. The statistics about forks are staggering, topping at 4266 for the main bitcoin repository, followed by the Go repository from Ethereum with 695 forks.

## 4.3 Blockchain-oriented Software Engineering: New Research Directions

Analysing the results of the section 4.2, we identified some research directions that BOSE should take in order to fill the gaps.

**Testing.** A recent study on over 50000 GitHub projects [61] has proved that a bigger team size leads to a higher number of test cases, whereas the number of test cases per developer decreases with an increase in the team size. It would be interesting to investigate whether the same can be said about

BOS, considering that the most popular repositories have an unusually high number of contributors, even for open-source projects. For instance, almost 400 GitHub users are contributing to the `bitcoin/bitcoin` repository, as reported in Table 4.1.

**Collaboration.** The high number of voluntary contributors testifies to the attractiveness of BOS in the open source landscape. A large base of voluntary contributing members has been shown to be a pivotal success factor in OSS evolution [87]. To achieve sustainable development and improve software quality, specific practices to enhance the synergy between the system and the community would be highly beneficial to BOSE [1].

**Enhancement of testing and debugging for specific languages.** Figure 4.1 shows that a number of programming languages such as Go, Python, and Ruby are gaining increasing popularity among BOS projects. This arises the need for enhanced testing and debugging suites, tailored upon the most popular BOS languages. Indeed, Java testing suites have undergone much more testing than Go. In addition, as BOS projects work with the Blockchain, which is distributed by definition, testing in isolation would require properly mocking objects capable of effectively simulate the Blockchain.

**Creation of software tools for smart contract languages.** The implementation of Smart Contract Development Environments (SCDEs)–the blockchain-oriented declination of IDEs–might be pivotal for the building and diffusion of BOS expertise. Such environments could streamline smart contract creation through specialized languages (e.g. Solidity, a language designed for writing contracts in Ethereum).

It is interesting to note that during the last year, the scientific community has organized itself to respond to these challenges. In particular, scientific workshops were organized on the topic. To give a noteworthy example, the 1st International Workshop on Blockchain Oriented Software Engineering[10] (held on 20 March 2018 - Campobasso, Italy in conjunction with the SANER conference) and the 1st International Workshop on Emerging Trends in Software Engineering for Blockchain[11] (scheduled for 27 May 2018 - Gothenburg, Sweden in conjunction with ICSE).

[10] http://saner.unimol.it/blockchainOrientedSoftwareEngineering
[11] https://www.icse2018.org/track/icse-2018-Workshops

# Chapter 5

# Blockchain applications for people's participation in smart systems

As discussed in the previous chapter, the blockchain is not only the backbone of payment systems but is also the enabling technology of new typology of software systems, and allows the creation of new typology of services. In this chapter we discuss two case studies we realized to study applications in which the blockchain represent the infrastructural layer of advanced services. In particular, in these applications, the blockchain takes a primary role in the democratization of typically centralized systems, passing thorough the involvement of people in the acquisition of information and in the decision process. The first case of study is an application in the context of the smart city paradigm and involves the Internet of Things (IoT). The blockchain allows the creation of a decentralized repository of environmental data. Also the second case of study implements a smart system. This system uses the blokchain to implement a decentralized energy market in which producers and consumers can exchange energy in a decentralized way, exploiting the services offered by smart grids of energy. Given the novelty of the technology and the necessity to involve users in the development process, the two cases of study are conceived to be implemented using the Agile methodology. We are currently studying other case studies concerning the management of social systems [105, 75].

## 5.1 Blockchain and smart systems: CitySense

This section presents the case of study of the application of the blockchain in the development of a smart city system. The system is called CitySense and we presented it during the XP2017 workshop entitled "Generating Innovations for the Internet of the Things: Agility and Speed"[55].

CitySense aims to create new ways for citizens' public life participation. In facts, in order to enhance the continuous relationship between local government and people, CitySense promote the active role of citizens [138]. From a physical point of view, the CitySense can be seen as a network of mobile sensors, and can be framed as a Internet of Things system. Sensors produce digital measurements, useful for investigating and studying the life quality of a city.

In order to maximize social impact of the system, environmental data is designed to be available and shared with citizens. In this system, is important that people trust in the shared data and these data should be unmodifiable. For this reason, we propose to solve the problem of environmental data storage and management by using the blockchain technology. The blockchain satisfies the demand of availability and not-modifiability of data and, thanks to the potentiality of smart contracts, allows the management of sensors data by means the implementation of a decentralized control logic.

Analyzing the project specifics we propose to design and implement this system using an Agile methodology. In particular, the SCRUM methodology is a flexible, adaptive and iterative methodology that fits very well to the purpose of the system and to the use of the blockchain we plan to do. For these reasons, we propose to drive the development and the maintenance using the SCRUM methodology.

### 5.1.1 Background

Smart cities are systems that aim to provide new services and the better ground coverage of existent ones to citizens. This typology of systems include the interconnection of a huge number of devices, among which the computational effort is distributed. Some of these devices are simultaneously both as client and server. Modern smartphones have a key role in smart city systems. These devices are up, in performance, to some of the mainframe servers deployed 20 years ago in the ISP industry [101]. Tailored software and simple computation tasks can be written to be executed by such small devices with good performances. Across Europe, several cities have been engaged in environment safeguard plans, starting from the 2008 SETIS Plan to reduce CO2 emission that widespread over 12 countries and 200 mid-sized towns, up to single town's projects to reduce power consumption and invest on new energy markets while the industrial ones that relies on technologies and sources from the last century are slowly fading away. Emphasizing the social approach of a smart city plan would be the key for raising the interest of the citizens who will became the main actors in this project, for contributing to the mass effect of self knowledge of the environment they live in and for keeping it as wealthy as possible. This social approach can be interpreted correctly by the blockchain technology. The feasibility of a blockchain-oriented

smart city is an in-progress study. Looking in the web, is easy to find related initiatives, like that of the Dubai government which is programming to create a blockchain based smart city [1], and debates about potentiality of this application [2]. Sharma et al. [120] studied an application of blockchain technology to build a vehicle network which takes in account several problematics that are the mobility of nodes (which represent vehicles), the confidentiality and security of data. Security aspects are also discussed by Biswas et al. in their work [13]. We take this work into particular consideration since it work discusses the need of a specific security framework. Such framework is composed of four layers: the physical layer (which includes sensors), the communication layer (in which is considered the blockchain), the database layer and the interface layer (that considers all applications). Considering the IoT a key element for the smart cities development, we briefly discuss IoT related applications of the blockchain technology. In facts, it is an enabling technology for empowering the potentiality of the IoT. The work of Quaddah et al.[102] provides a well-defined framework named FairAccess to enable the communication between nodes by means of some blockchain based mechanisms (i.e smart contracts and transactions). Thanks to the blockchain, this framework provides a stronger and transparent access control tool. E-business aspects of IoT technology are discussed by Zhang [139]. He studied a blockchain application which he implements a seller-buyer model describing business operations between two or more devices. Christidis and Devetsikiotis [23] provide a discussion based on the literature, proving that smart contracts and blockchain applied to the IoT can be pretty powerful. They also provide an interesting section about the blockchain taxonomy. IoT (RFID based) and blockchain can also enable products traceability in a supply chain, as discussed by Tian [126], and can also enable the control of remote robots, as discussed by Ferrer [42]. Recent studies [71] have shown that a smart city needs to have a smart local council and a smart methodology in order to develop an efficient software. A good way to achieve this goal is the use of SCRUM process with some changes with respect to the original approach. Several kinds of data, collected by sensors in a IoT system [58], can be used to increase the openness of public government and political choices, to improve the people's awareness about well-being of the city and to encourage the involvement of citizens in the drive for sustainable development.

[1] `http://www.coindesk.com/dubais-museum-future-sees-blockchain-smart-cities/`
[2] `https://dcebrief.com/blockchain-powers-new-smart-city-initiative/`

### 5.1.2 The CitySense system

The CitySense aims to create a data collection mechanism that works combining the measurements acquired by specific mobile devices with the validation process made by the blockchain algorithm which processes the measurement data. Considering a geographic area, such as a town or a county, we can imagine a full coverage of ground environmental measures, made by portable devices that collect data and send them back to the blockchain. Some environmental phenomena are linked to pollution issues. They have a direct impact on people's lives and affects the public opinion. The measurable phenomena includes noise, temperature, humidity, light, and the concentration of hydrogen, methane, carbon monoxide, and micro-particulates in the air. Thanks to the wide diffusion of small programmable embedded devices, the number of actors eligible to be components of the network is very high, and small personal "Smart Objects" (namely smart phones, smart watches, wristbands and so on) allow the creation of an interconnected ecosystem of devices.

Smart phones are the best devices to be used for this task as they provide a quite good computational capability, moderate battery life, wireless connectivity to other near devices, internet access and they are easily programmable to run generic purpose software on top. To connect to hardware-level machine, like electronic devices and sensors, a HW/SW connection layer should be deployed. As today, a lot of such programmable tiny operational boards equipped with AVR processing units could be used for this task.

The number of sensors deployed could be larger as the software elaborates the results in a more sophisticated way. Computational power on this task in not a real big issue as digital multiplexing on data is quite easy for this scenario. In order to develop the CitySense system, we plan to use the Ethereum platform to record measurements arriving from the network of sensors. In our system, sensors are IoT devices, programmed to be connected with the blockchain and able to send messages. As in [13], CitySense is structured by layers as is shown in Fig. 5.1.

Layers of the system will be implemented in parallel. Each layer has specific and distinctive requirements and will be implemented in an iterative process, with the advantages of the SCRUM methodology. The set of sensors composes the physical layer. In our system, sensors are carried around the city by their owners. For this reason, each measurement must be associated with the geographical position, in addition to the typology of measurement and to the timestamp.

Considering the network layer, in our system sensors are programmed to send measurements to the blockchain through the peer-to-peer network by means of a light version of existent clients such as geth. Actually, sensors can

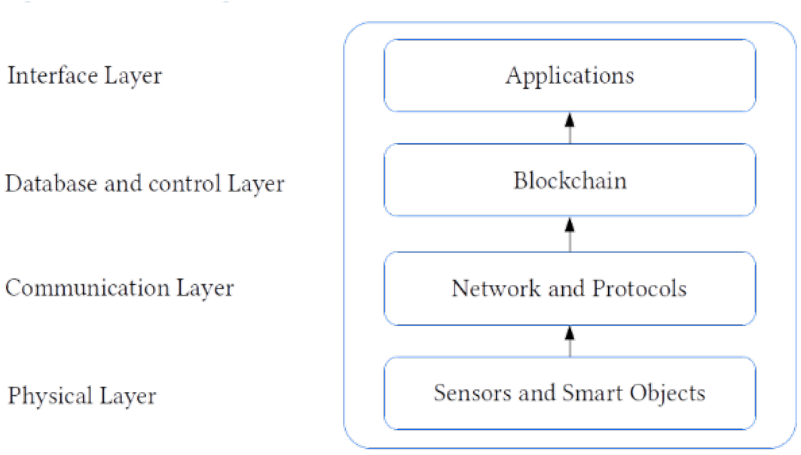


Figure 5.1: Layers of the Citysense system

be connected to a mobile device (for instance a smartphone) to run, or can be autonomous systems able to connect to the peer-to-peer network. Different communication mechanisms, such as WiFi, 3G or Bluetooth, can be used. Specific communication protocols should be defined considering the cost of a transaction in the Ethereum system (which is proportional with the payload size of the message), and designed according to a convenient data format.

The variety of applications enabled by the CitySense system compose the application layer. Thanks to the collaborative nature of our system, the same people who participate to the creation of the CitySense are also able to take advantage of the information that is made available. For instance, a CitySense mobile application could be useful to people who want to know the environment quality of a specific area in the city. Furthermore, web applications could be used by a public administration to check if and where the city has pollution problems.

### 5.1.3 The blockchain solution

The blockchain is the disruptive technology which will drive the development of future smart-cities related applications. Key features, such as to be a shared, transparent, distributed, secure, available and smart technology, make the blockchain an opportunity to improve potentialities of IoT and smart-cities development. As it is known, blockchain data are publicly available. In a smart-city, transparency makes citizen aware and able to know the contribution of each of them and how public governments use data. In case of need (i.s. sensitive or

personal data) it is always possible to encrypt data before they are stored inside the blockchain.

In a blockchain based service, nodes participate to the objective in a democratic way, under the constraint of the consensus mechanism. Therefore, blockchain enables trust-free transactions without the need of a central control authority. IoT devices participate in the peer-to-peer network sending messages to it, as nodes of the system and under the blockchain rules. In a blockchain system, smart contracts, computer programs located and working in the blockchain, can automatically acquire data from IoT devices and produce computed outputs. Because of the distributed nature, smart contracts cannot be modified or interrupted. For this reason, their usage could improve the reliability of a smart-city system. Blockchain is a decentralised and public available shared database, reachable by accessing to a peer-to-peer network. The blockchain is an enabling technology which allows to obtain the satisfaction of security and availability of data, and provides the computational power that makes it able to control the communication between nodes.

As discussed before, a transactions can be seen as a way to send messages between two nodes, in a communication system. In a blockchain, each node of the communication system is called account and is identified by a fixed length hexadecimal number named address. Considering the Ethereum blockchain, the state is composed by all variables recorded in the blockchain and their values. These variables include all addresses balance and contracts internal variables.

A contract is a special typology of account. It is recorded in a block of the blockchain, and can receive and transmit messages from and to other account, by means of transactions. Messages can request the execution of a specific contract functions. For instance, a message can contain the address of the receiver, the name of the function, and a list of parameters.

In CitySense the blockchain consists in the database layer in which measurements are organized and stored. In our system we use contracts to control and save data. A first contract is designed to be the receiver of messages coming from sensors. We call it acquisition and sorting contract (ASC). Depending on the geographical position, the ASC sends the measurement to one of the set of specialized contracts that we call geographic contracts (GC). Inside each GC only measurements coming from a specific geographic area are stored. With specific implemented functions, GCs can be queried to provide measurements organized in several ways (for instance measurement can be organized by their typology, by timestamp, by value).

### 5.1.4 Metodology

IoT, SCRUM methodology and blockchain form the basis of digital transformation. A trust distributed technology ensure privacy, scalability, transparency and reliability. In a Smart City the number of linked objects is set to increase and will produce higher and higher operating costs. The blockchain will be a key element for the cost-cutting in the tracking and coordination of physical devices and will solve failure problems of traditional networks. Currently all IoT systems depend on client and server communication protocols, such as SSL and TLS and on cryptographic mechanisms such as the Public Key system, used in order to make the communication system verifiable and to authenticate the nodes of the network. Such architecture will soon have to face several problems (delay transmission for example) caused by data traffic congestion. Therefore, the decentralization, which is an intrinsic property of blockchain, could be the right solution in order to increase the network efficiency and reduce management costs. CitySense will allow a direct communication between smart devices and will verify the transactions without a central server.

We propose to use SCRUM development methodology in order to implement this innovative blockchain-based system. We chose to apply a SCRUM process because of its capabilities of being flexible, adaptive and iterative. Scrum methodology will be also used in order to ensure software quality, reduce the time-to-market, enhance the support of the citizens and create an infrastructure that allows the transfer of data in real time using a sensor networks with a low energy consumption [28]. The support of the citizens is essential to make this project succeed, but at the same time implies more complexity. For this reason we estimated a duration of 30 days for every sprint: this choice improves the collaboration of stakeholders. Unit test are planned for each iteration [32]. According to SCRUM methodology, in order to verify the correctness of system developed, the client's role is extremely important in several aspects at all the stages of the process. The contribution of customers impacts directly on product quality. For this reason, the clients must be involved in any decision adopted and shall collaborate closely with the development team. In this project, clients are identified with all citizens who participates in the collaborative data acquisition and provides feedbacks. The Product Owner is who analyses information provided by users, in order to interpret consumer expectations and requirements, to filter communications, to identify priorities and to distribute tasks within the development team. Finally, the local government is involved by the Product Owner in the analysis of information communicated by users in order to provide feedback to the developers and take decisions. This method favours the effective cooperative approach.

## 5.2 Energy and business: Crypto-Trading

This section presents the research for the implementation of blockchain based system which extends the features of cryptocurrency exchanges to the renewable Energy Market: the Crypto-Trading system. We presented this case study during the 2017 AEIT international conference[74].

Aiming at promoting smart growth, sustainable development and social inclusion, identifies the regional excellences in terms of research and innovation and their growth potential.

Starting from careful reading of technological and development opportunities, our research idea takes inspiration from the research on smart grids area whose general objective is to improve the technologies adopted to generate energy that is collected from renewable resources and to allow a more economic and efficient management of the local resources.

The purpose of the Crypto-Trading system is to perform two main objective:

1. Efficient management of energy demand and supply in order to improve the distribution networks and regulate the consumption in an energy saving perspective;

2. Monitoring and analysis of electricity consumption by final consumers (private and business) independently of the electricity supplier.

In the scenario of an intelligent energy distribution network, objective will be achieved through the creation of a platform based on a token system for the purchase and sale of energy enabled to record the amount of energy purchased. Each user can become an holder of a certain amount of energy and resell it at any time at the bid price. The transition from virtual trading to energy delivery is decentralized, thanks to the use of energy-oriented systems that are geared to new and advanced paradigms. According to the Barenergy Report, six typologies of barriers are slowing the progress in efficient energy and diffusion of renewable energy systems: physical, political, socio-cultural, economic, knowledge based and individual barriers. In particular, socio-cultural barriers are obstacles to the changing of consumers' behaviour. For this reason, the civil society should be involved in the changing by providing new positive points of view. The Crypto-Trading system responds to the need to promote the transformation of an energy model characterized by a centralized production to a decentralized and intelligent production and distribution, tailored to the needs of proximity and aimed at satisfying local consumption.

### 5.2.1 Background

The use of blockchain technology for the implementation of a decentralized electricity market is currently being studied. In recent years, there has been a growing interest to study the scenarios of the creation of a decentralized and local Energy Market (LEM) by using the blockchain technology to enable development of intelligent energy networks and the advancement of measurement and control systems [102] . In [59], the authors give particular emphasis to the innovation that blockchain technology could give to power distribution networks. There are many possible scenarios for self-produced energy from prosumers resident in European territories, as reported by Bitcoin Magazine [108]. In [82], [81] Mihaylov et al. introduce NRGcoin, the value of which is determined on an open currency exchange market. In their paper the prosumers in the smart grid trade locally produced renewable energy by using NRGcoins. On the same line of Crypto-Trading project, there are similar initiatives such as powerledger [3] operating in Australia and currently being studied. Energy market projects, based on blockchain and smart grids, should take in consideration the issue of sustainability. In a recent work, Mengelkamp et al. [80] proposed an agent model in order to investigate the feasibility and the sustainability of a LEM based on a private blockchain. During the development of the Crypto-Trading project, we will take into account all the previous proposed solutions. Furthermore, the Crypto-Trading project intends to consider the opportunity to develop a solution based on a public blockchain implementation, and aims at a continental-scale Energy Market. So the sustainability issues will be faced during the project. In particular, using the blockchain instead of conventional ICT centralized system leads to some energy issues, as discussed in Chapter 3. Note that the study of blockchain applications must include the study of specific solutions in order to preserve the users' privacy. This important aspect is crucial in the development of Crypto-Trading system [49, 4].

### 5.2.2 The Crypto-Trading system

The realization of the Crypto-Trading system require the application and extension of the knowledge of financial trading and blockchain technology in the field of the energy trading exploiting the cryptocurrency technology. The main goals will be two:

1. The introduction of an European Energy Market trading system by using "token" and smart contracts that will simplify the trade of electricity distributed by intelligent networks;

[3] see: https://powerledger.io

2. the development of a platform for the optimal allocation of a cryptocurrencies portfolio.

The system takes advantage of an existent financial trading platform and web application called Selfiewalth[4]. Such platform provides advanced financial services, empowered by a robot-advisor. This Robot-advisor is the result of the implementation of a prediction algorithm studied and developed by the company itself. This algorithm is developed in order to perform the financial analysis on over 80,000 different stocks or funds and over the principal fiat currency. Very recently, Selfiewealth added the financial analysis of many popular cryptocurrency. The platform provides a guided creation of a personal portfolio. A simple preference survey identifies users' financial preferences, in particular in terms of the risk-reward ratio. The Crypto-Trading system aims to bring that solution in the world of the Energy Market, taking advantage of the cryptocurrency and blockchain technologies.

System developing includes the creation of a specific trading platform focused on cryptocurrencies, by studying and adopting new typologies of market indicators, which are evaluated including the analysis of blockchain data. Such analysis will include transaction volumes and users behaviors [122, 17, 57] .

Crypto-Trading can enable local prosumers to buy and sell energy, and at the same time, produce it from renewable sources. Furthermore, Crypto-Trading introduces a new typology of Energy Market. It is characterized by simplicity and reliability and it promises a new business opportunity for small producers. Using Crypto-Trading, the electricity could be traded following the free market rules, at the price established when supply and demand meet. The system will provide advanced tools which will make the prosumers aware about the current Energy Market trend and historical trend, making easier the price formation and overcoming regional boundaries. Furthermore the users could know the energy origin and the typology of the source. The Energy Market will became an individual based, decentralized and free market. It is reasonable to imagine that in the near future we will see the coverage growth of smart grid and connected technologies devote to the control and distribution of the energy, similar as the Internet network did in the past [84]. For example, the Sardinian region was electrical isolated until the 2010, year when was tested the High Voltage Sardinia Island – Italian Peninsula link (SAPEI) was introduced [112]. For this reason, in the future the majority of the energy sources and the energy consumers, will be physically and digitally connected to smart grid systems. In this scenario, local energy storages are not mandatory. The Crypto-Trading system has no geographic coverage limitation and could base its IT infrastructure on a public blockchain.

[4] see: http://www.selfiewealth.com

### 5.2.3 The blockchain solution

The system will take into account the complexity of a smart and efficient energy distribution and will combine the smart grid technology with the blockchain technology. In particular, in our system, the blockchain technology is the ICT upon with the system works. The blockchain works both as a ledger (i.e. a database of all transactions in which energy sales are recorded and can no longer be modified) and as a control system which by means of existent technology drives smart meters. Overall, the Crypto-Trading system can be described as the composition of three functional subsystems: the prosumer system, the blockchain system and the trading system. The prosumers system is composed of the energy sources, the smart meters, and the final users who sell and buy energy. Each user is the owner of an energy account that is associated to a smart meter. Smart meters have the task of measuring the energy production and energy consumption, and of interrupting the energy availability if the related energy account is empty. Smart meters must be ready for the internet connection and able to host a blockchain light client. In the blockchain, specific smart contracts receives messages from the smart meters and from the trading platform. All the energy trading operations are coinceved to be publicly available in the blockchain. For each prosumer, the energy availability (described by means of the definition of an energy token) is recorded in a specific smart contract that represent the energy account of each user of the system. A token describes the tradable energy unit. Each prosumer can buy and sell tokens at any time and thanks to the blockchain technology it is not possible sell twice the same energy token. The exchange currency, with which users pay the energy, could be the cryptocurrency associated to the blockchain system (i.e. Ether inside the Ethereum blockchain) or a different specific token, having an independent value. The currency availability is recorded inside the users' blockchain accounts. The trading system is the web application which allows the prosumers to access the robot-advised trading services. This system reads the energy production data from the blockchain and sends control messages to the smart meters through related smart contracts. The trading services include personalized solutions based on user's preferences. In particular, the robot-advisor continuously produces reports and suggestions, basing them on the energy price trends, and on the constraints imposed by each user. In Fig.5.2 we show a conceptual representation of the Crypto-Trading system. In this figure, it is possible to identify the three typologies of subsystems. Prosumers are represented as nodes of the power network (red line). In particular, two of those nodes represent the typical domestic prosumer: a node represents the commercial/industrial prosumer and a node represents a power generation station. Each node is provided with a smart meter. Smart meters are at the same time connected to the power line and to the

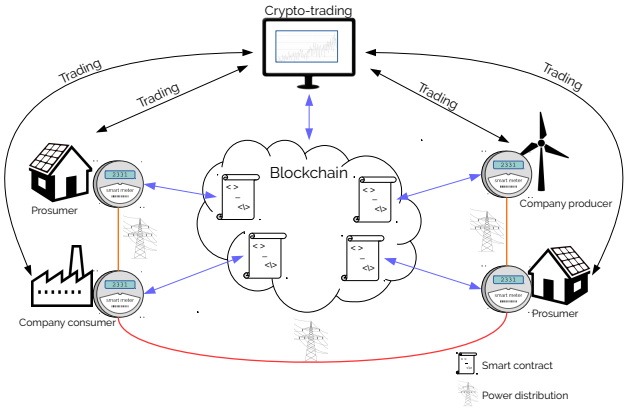


Figure 5.2: Representation of the Crypto-Trading system.

blockchain using an internet connection (blue arrow). Whenever a prosumer produces a unit of energy, its smart meter sends a message to the smart contract connected to it in order to increase the number of energy tokens. Each smart contract receives messages from the smart meter and from the trading system, according with the prosumer's necessity. Whenever an amount of energy is sold by a prosumer, the trading system sends a message to the related smart contract, which, in turn, sends the equivalent number of energy tokens to the buyer's smart contract. In this way, the transaction of energy token prevents the double selling of energy. In parallel, by means of a blockchain transaction, the value of the sold energy is transferred from the buyer to the seller. Prosumers can trade the energy using the web based trading system (black arrows). Basing on the robot-advisor suggestions, prosumers can set up a strategy for energy trading to maximize revenues (or minimize costs) and have a satisfactory coverage of their energetic needs. In particular, the robot-advisor can advise users for the best offering who can plan a personalized strategy, or operate autonomously, managing the users' energy account in order to satisfy user's requirements.

### 5.2.4 Discussion

The Crypto-Trading research activity will be carried out in three stages:

1. analysis and design of the general architecture for the local management of a energy distribution network;

2. integration and development of Smart Contracts for the Energy Market;

3. development of the prototype of Crypto-Trading system.

More specifically, we will discuss the system requirements for the local management of an intelligent energy distribution network, including a market where small producers and users interact to optimally allocate and economically optimize the various resources. Smart Contracts will also be analyzed and designed for the management of the interaction among intelligent tools (IoT) [23], manufacturers, users and the market, including exchanges with other intelligent electrical grids. From the analysis, the design of the Energy Market will be carried out, integrating it into the market with the aim of allowing payments and quotations of electronic exchanges in cryptocurrencies, and reusing prediction and risk control algorithms. Technological architecture and security infrastructure approaches, as well as the security for each element working on the infrastructure, will be defined.

As discussed in Chapter 4, if the blockchain is the core component of the system, the modeling and development phases could face the lack of proper guidelines. In order to properly manage eventual development problems, the system can be modeled and developed using existent Agile methodologies. In facts, one of the characteristics of these methodologies is the adaptability. Furthermore, the heterogeneity of the components of the system leads to the need of a modular and iterative approach that enables the parallel and test driven development. The Agile methods provide the process that allow the incremental development of the Crypto-Trading system [100]. In order to have short iterations and a Use Case driven development, the prototype development (that includes the back-end and the web interface of the trading web application) will be done using an Agile methodology. In particular the practices of continuous integration, automated testing, and the refactoring will be used. The prototype will be validated through a testing phase, including a laboratory simulation in operational conditions.

# Chapter 6

# Conclusion

The blockchain is a disruptive technology, a tool which will be more and more present in our life. The thesis, result of three years of research during my Ph.D., explores the blockchain technology issues looking at them from two main point of view, and for this reason it is organized in two parts

The first part answers the need to understand, analyse and evaluate the blockchain technology. The aim was to infer information on users' behaviours and on global impact, by means of empirical studies and developing analysis algorithms. In particular, chapter 2 introduced a blockchain analysis approach based on a Petri Net. The purpose was to define a single useful model in which all main information about transactions and addresses are represented. By using this model, it is possible to pick out significant and original results. This formalism has proven to be a powerful methodology for performing many kinds of measurements and analysis. The model was tested on the first 180 thousand blocks of the Bitcoin's blockchain, and it well described the transaction tree. Summarising, the model associates a place for each address and a transition for each bitcoin transaction. Our Petri net includes *pre* and *post-incidence* matrices where all links between addresses and transactions are modeled. All the empirical analysis was obtained by elaborating matrices elements. Analysing the number of pre and post arcs, we had proof of the presence of power-law like distributions. In a nutshell, few addresses receive or send a huge number of transactions. In addition, analysing both incidence matrices, all transactions chains, identifying a typical *disposable addresses* usage by Bitcoin users was identified. By measuring the chains' lengths, we found again power-law like distributions. The model allows us to recognize addresses belonging to the same owner, denoted *Entity*. In order to obtain information about addresses' owners, an algorithm Entities and constructs The Entities Petri Net. The portion of blockchain which we chosen was processed without specific hardware resources. Anyway, the current size of the blockchain (over 494,000 blocks and the total

number of transaction is over 240 million) requires more resources in order to handle all the blocks information. Despite the current blockchain size is about two orders of magnitude greater than the size of the portion that we have studied, our approach can be adopted to study a specific portion of the blockchain, for example starting from a specific set of addresses which we want to investigate and analyze. Finnaly, the Petri Net model can be used for studying a large set of other issues related to other systems based on blockchain technology, such as Ethereum.

Then, Chapter 3 focused on a specific issue of blockchain technology, that concerns the usage of this technology in banking and capital markets. The chapter analyse the Bitcoin system during the years, taking into account its evolution. In particular, it considered that mining hardware has evolved over time, passing trough CPU, GPU, FPGA, and ASIC, taking into account two of its features: the hash rate and the power consumption. The mining evolution was represented using two fitting curves: the "best hash rate per $", $R(t)$ and the "best power consumption function", $P(t)$. We defined three quantities: "economic efficiency" (EE), "operational efficiency" (OE), and "efficient service" (SE). Results show that the EE, defined as as the ratio between the value of bitcoins mined by the power consumption of 1 kWh, is characterised by a strong variability because it is influenced by the Bitcoin popularity and the power consumption of the network. It is currently growing, thanks to the growing of the Bitcoin price. Second, the OE, defined as the ratio between the value of voluntary fees and the energy cost of a transaction, is currently growing, indicating that fees are becoming more and more important to assure the sustainability of the Bitcoin system. In fact, mining operations will be remunerated only until the sum of circulating bitcoins reaches 21 million. Finally, the SE, defined as the ratio between the number of transactions validated by the power consumption of 1 kWh, which describes how much electricity the network spends to perform its main service, i.e., to wire bitcoin. Because transaction blocks are limited in size (1 MB), the number of transactions per block is limited, and the SE can not increase.

The second part of the thesis moves the focus on engineering and design issues in software projects that include the blockchain technology. To start with, Chapter 4 discusses the general issues not yet covered by the software engineering. The chapter introduces the Blockchain-oriented software engineering, in order to point the attention to the challenges and the new directions which will allow effective software development. In the chapter, the most evident issues of state-of-art blockchain-oriented software development are discussed, by advocating the need for new professional roles, enhanced security and reliability, novel modeling languages, and specialized metrics. Statistical information on popularity, collaboration, repository age, and programming languages are

obtained by analysing a dataset of blockchain-oriented software repositories, created using the 2016 Moody's Blockchain Report and the market capitalization of cryptocurrencies. In addition, the chapter provides new directions for blockchain-oriented software engineering, which focus on improving collaboration among large teams, testing activities, and specialized tools for the creation of smart contracts.

To give an idea of the disruptive nature of the blockchain, the thesis concludes proposing two blockchain-oriented software projects elaborated during the PhD research activity. Chapter 5 describe two case studies. The first, Citysense, is a blockchain and IoT project that aims to monitor the environmental data in a city in a collaborative way. The project puts the focus on the decentralization and the collective involvement of all citizens. The blockchain technology allows the realization of a reliable communication layer, which receives data from sensors and elaborates them with smart contracts. In this project, in order to really involve citizens and local council in the process, SCRUM results the most appropriate developing methodology. Thanks to the characteristics of blockchain infrastructures, local government can obtain a low cost real-time map of environmental data of the city, making it able to take real time countermeasures in case of pollution alarms. The second case of study is the Crypto-Trading project. The section highlights the key role of the blockchain technology and smart contracts in the management and control of an innovative typology of Energy Market. Taking inspiration from the Sardinian Region S3 goals, Crypto-Trading aims at facilitate the creation of a decentralised Energy Market, making the final user (the prosumer) able to self manage the supply of energy and the sale of the excess energy. The new business opportunity will help to overcome barriers that slow the growth of the adoption of state-of-the-art technologies in the field of smart grids. In particular, Crypto-Trading will provide a robot-advisor which will help the users to optimise the energy trading. The proposed system could facilitate the transformation of the energetic model in the direction of a decentralized and smart production of electricity.

Summarizing, the thesis work has offered a vast number of results, embracing the thesis topic from many of its points of view. In particular, at the end of this work and in order to summarize the results obtained, we can say the following.

- The systems called blockchain can be modeled and analyzed through Petri nets. Our analysis system, tested on Bitcoin blockchain, allows to effectively compute and deduce a large number of statistics and information on the blockchain usage.

- The blockchain technology can become the core technology of new banking systems.According to our studies, the Bitcoin system can work as a

banking system. The main obstacle is related to its sustainability. In fact, efficiency problems limit its scalability and they will have to be solved before the system can be used widely as a bank.

- The development of blockchain-based software systems are exploding and new software engineering challenges need to be addressed. The Blockchain-oriented Software Engineering (BOSE) proposed by us sets objectives for engineers and practitioners that will allow to arrive at the definition of a specific design and development practices for block chain-based systems.

- Finally, we have demonstrated, through the analysis of two case studies proposed by us, that blockchain technology allows to realize systems to improve people's lives, thanks new democratic and participatory services. The blockchain provides natively the trusty system that needed by these services.

Our work will continue on these issues, starting from the points still open. For example, we will evaluate the validity of the methods proposed for the analysis of the first part of the Bitcoin blockchain on other types of blockchain (for example that of Ethereum). We will study accurate answers to the challenges we have launched for the foundation of BOSE, providing studies on smart-contract metrics and analyzing the development tools that are released. Regarding the case studies, we propose to carry out a SWOT (Strengths, Weaknesses, Opportunities and Threats) analysis to find where our approaches are lacking. In addition, other case studies will be studied in order to increase attention to the use of blokchain in social contexts.

# Bibliography


[1] Mark Aberdour. Achieving quality in open-source software. *IEEE software*, 24(1):58–64, 2007. [cited at p. 76]

[2] Alain Abran, James W Moore, P Bourque, R Dupuis, and LL Tripp. Swebok: Guide to the software engineering body of knowledge 2004 version. *IEEE Computer Society, Los Alamitos, California*, 2004. [cited at p. 71, 72]

[3] Accenture. Banking on blockchain, a value analysis for investment banks. Technical report, Accenture Consulting, 2017. [cited at p. 47, 49]

[4] Nurzhan Zhumabekuly Aitzhan and Davor Svetinovic. Security and privacy in decentralized energy trading through multi-signatures, blockchain and anonymous messaging streams. *IEEE Transactions on Dependable and Secure Computing*, 2016. [cited at p. 85]

[5] F. M. Ametrano. Hayek Money: the Cryptocurrency Price Stability Solution. 2014. [cited at p. 51]

[6] E. Androulaki, G. Karame, M. Roeschlin, T. Scherer, and S. Capkun. Evaluating User Privacy in Bitcoin. In *Proceedings of the Financial Cryptography and Data Security Conference (FC)*. [cited at p. 19]

[7] Nicola Atzei, Massimo Bartoletti, and Tiziana Cimoli. A survey of attacks on ethereum smart contracts (sok). In *International Conference on Principles of Security and Trust*, pages 164–186. Springer, 2017. [cited at p. 5, 70]

[8] F.M.a Aymerich, G.a Fenu, and S.b Surcis. A real time financial system based on grid and cloud computing. In *Proceedings of the ACM Symposium on Applied Computing*, pages 1219–1220, 2009. [cited at p. 12]

[9] Adam Back, Matt Corallo, Luke Dashjr, Mark Friedenbach, Gregory Maxwell, Andrew Miller, Andrew Poelstra, Jorge Timón, and Pieter Wuille. Enabling blockchain innovations with pegged sidechains. *URL: http://www. opensciencereview. com/papers/123/enablingblockchain-innovations-with-pegged-sidechains*, 2014. [cited at p. 72]

[10] Iddo Bentov, Ariel Gabizon, and Alex Mizrahi. *Cryptocurrencies Without Proof of Work*, pages 142–157. Springer Berlin Heidelberg, Berlin, Heidelberg, 2016. [cited at p. 53]

[11] Iddo Bentov, Charles Lee, Alex Mizrahi, and Meni Rosenfeld. Proof of Activity: Extending Bitcoin's Proof of Work via Proof of Stake. Cryptology ePrint Archive, Report 2014/452, 2014. [cited at p. 53]

[12] A. Biryukov, D. Khovratovich, and I. Pustogarov. Deanonymisation of clients in bitcoin p2p network. *CCS '14 Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security Pages 15-29*, April 2016. [cited at p. 18, 37]

[13] K. Biswas and V. Muthukkumarasamy. Securing smart cities using blockchain technology. In *2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, pages 1392–1393, Dec 2016. [cited at p. 79, 80]

[14] Bitcoinwiki. Scalability. 2016. [cited at p. 51]

[15] G. BitFury. Proof of Stake versus Proof of Work. Technical report, BitFury Group, 2015. [cited at p. 50, 53]

[16] J. Bradley. The Energy Efficiency of Bitcoin. `https://www.cryptocoinsnews.com/energy-efficiency-bitcoin/`, 2016. [cited at p. 44]

[17] Oosterlinck K. Brière, M. and A. Szafarz. Virtual currency, tangible return: Portfolio diversification with bitcoin. *Journal of Asset Managemen*, 16(6):365–373, 2015. [cited at p. 10, 86]

[18] V. Buterin. Slasher: A punitive proof-of-stake algorithm. 2014. [cited at p. 52, 53]

[19] V. Buterin. On stake. 2016. [cited at p. 51, 52, 53]

[20] Maria Paola Cabasino, Alessandro Giua, and Carla Seatzu. Identification of petri nets from knowledge of their language. *Discrete Event Dynamic Systems*, 17(4):447–474, 2007. [cited at p. 19]

[21] Nick Caes, Robard Williams, Elena H Duggar, and Michael R Porta. Robust, cost-effective applications key to unlocking blockchain's potential credit benefits. 2016. [cited at p. 73]

[22] J. Carterand. Does Bitcoin Have an Energy Problem? `http://blog.acton.org/archives/82688-does-bitcoin-have-an-energy-problem.html`, 2015. [cited at p. 58]

[23] K. Christidis and M. Devetsikiotis. Blockchains and smart contracts for the internet of things. *IEEE Access*, 4:2292–2303, 2016. [cited at p. 5, 79, 89]

[24] Pavel Ciaian, Miroslava Rajcaniova, and d'Artis Kancs. The economics of bitcoin price formation. *Applied Economics*, 48(19):1799–1815, 2016. [cited at p. 5]

[25] Concas G. Cocco, L. and M. Marchesi. Using an artificial financial market for studying a cryptocurrency market. *Journal of Economic Interaction and Coordinations*, 12(2):345–365, 2015. [cited at p. 9]

[26] Pinna A. Cocco, L. and M. Marchesi. Banking on blockchain: Costs savings thanks to the blockchain technology. *Future Internet*, 9(3):25, 2017. [cited at p. 8]

[27] Ethereum community. History of ethereum. 2016. [cited at p. 6, 50]

[28] Giulio Concas, Giuseppe Destefanis, Michele Marchesi, Marco Ortu, and Roberto Tonelli. Micro patterns in agile software. In *International Conference on Agile Software Development*, pages 210–222. Springer, Berlin, Heidelberg, 2013. [cited at p. 8, 83]

[29] Giulio Concas, Michele Marchesi, Giuseppe Destefanis, and Roberto Tonelli. An empirical study of software metrics for assessing the phases of an agile project. *International Journal of Software Engineering and Knowledge Engineering*, 22(04):525–548, 2012. [cited at p. 7]

[30] Monni C. Orrú M. Concas, G. and R. Tonelli. A study of the community structure of a complex software network. In *4th International Workshop on Emerging Trends in Software Metrics (WETSoM)*, pages 14–20, San Francisco, CA, 2013. IEEE Press. [cited at p. 18]

[31] Monni C. Orrù M. Concas, G. and R. Tonelli. Are refactoring practices related to clusters in java software? In *Proceeding of the 15th International Conference on Agile Processes in Software Engineering and Extreme Programming, XP 2014, Rome, Italy, May 26-30*, pages 269–276. Springer, Berlin, Heidelberg, 2014. [cited at p. 19]

[32] Steve Counsell, Giuseppe Destefanis, Xiaohui Liu, Sigrid Eldh, Andreas Ermedahl, and Kenneth Andersson. Comparing test and production code quality in a large commercial multicore system. In *Software Engineering and Advanced Applications (SEAA), 2016 42th Euromicro Conference on*, pages 86–91. IEEE, 2016. [cited at p. 8, 83]

[33] N.T. Courtois, M. Grajek, and R. Naik. The Unreasonable Fundamental Incertitudes Behind Bitcoin Mining. `http://arxiv.org/pdf/1310.7935v3.pdf`, 2014. [cited at p. 55]

[34] M. Czarnek. Nxt Network Energy and Cost Efficiency Analysis. `https://www.scribd.com/document/254930279/Nxt-Network-Energy-and-Cost-Efficiency-Analysis`, 2014. [cited at p. 6, 53, 55]

[35] S. Deetman. Bitcoin could consume as much electricity as Denmark by 2020. http://motherboard.vice.com/read/bitcoin-could-consume-as-much-electricity-as-denmark-by-2020, 2016. [cited at p. 45, 47]

[36] Kevin Delmolino, Mitchell Arnett, Ahmed Kosba, Andrew Miller, and Elaine Shi. *Step by Step Towards Creating a Safe Smart Contract: Lessons and Insights from a Cryptocurrency Lab*, pages 79–94. Springer Berlin Heidelberg, Berlin, Heidelberg, 2016. [cited at p. 5]

[37] Giuseppe Destefanis, Marco Ortu, Steve Counsell, Michele Marchesi, and Roberto Tonelli. Software development: do good manners matter? *https://dx.doi.org/10.7287/peerj.preprints.1515v1*, (e1892), 2015. [cited at p. 8]

[38] Giuseppe Destefanis, Marco Ortu, Steve Counsell, Michele Marchesi, and Roberto Tonelli. Measuring affectiveness and effectiveness in software systems. *arXiv preprint arXiv:1703.01642*, 2017. [cited at p. 6, 72]

[39] Giuseppe Destefanis, Marco Ortu, Steve Counsell, Stephen Swift, Roberto Tonelli, and Michele Marchesi. On the randomness and seasonality of affective metrics for software development. In *Proceedings of the Symposium on Applied Computing*, pages 1266–1271. ACM, 2017. [cited at p. 72]

[40] Giuseppe Destefanis, Marco Ortu, Simone Porru, Stephen Swift, and Michele Marchesi. A statistical comparison of java and python software metric properties. In *Emerging Trends in Software Metrics (WETSoM), 2016 IEEE/ACM 7th International Workshop on*, pages 22–28. IEEE, 2016. [cited at p. 7]

[41] Jules DuPont and Anna Cinzia Squicciarini. Toward de-anonymizing bitcoin by mapping users location. In *Proceedings of the 5th ACM Conference on Data and Application Security and Privacy*, CODASPY '15, pages 139–141, New York, NY, USA, 2015. ACM. [cited at p. 5]

[42] Eduardo Castello Ferrer. The blockchain: a new framework for robotic swarm systems. *CoRR*, abs/1608.00695, 2016. [cited at p. 79]

[43] International Institute for Sustainable Development. Sustainable banking. http://www.iisd.org/business/banking/sus_banking.aspx, 2016. [cited at p. 48]

[44] S. Fortunato and C. Castellano. Community structure in graphs. *Computational Complexity: Theory, Techniques, and Applications*, pages 490–512, 2012. Springer New York, New York, NY, USA. [cited at p. 18]

[45] Linux Foundation. Linux foundation's hyperledger project announces 30 founding members and code proposals to advance blockchain technology. https://www.linuxfoundation.org/news-media/announcements/2016/02/linux-foundation-s-hyperledger-project-announces-30-founding, 2016. [cited at p. 49]

[46] Gautham. The dominance of bitcoin network by mining pools. `http://www.newsbtc.com/2016/06/30/dominance-bitcoin-network-mining-pools/`, 2016. [cited at p. 61]

[47] A Giua and A Di Febbraro. Sistemi ad eventi discreti, 2002. [cited at p. 19]

[48] Florian Glaser. Pervasive decentralisation of digital infrastructures: a framework for blockchain enabled system and use case analysis. In *Proceedings of the 50th Hawaii International Conference on System Sciences*, 2017. [cited at p. 6]

[49] Steven Goldfeder, Harry Kalodner, Dillon Reisman, and Arvind Narayanan. When the cookie meets the blockchain: Privacy risks of web payments via cryptocurrencies. *arXiv preprint arXiv:1708.04748*, 2017. [cited at p. 85]

[50] H. Halpin and M. Piekarska. Introduction to security and privacy on the blockchain. In *2017 IEEE European Symposium on Security and Privacy Workshops (EuroS PW)*, pages 1–3, April 2017. [cited at p. 5]

[51] Adam Hayes. A cost of production model for bitcoin. *Available at SSRN*, 2015. [cited at p. 47]

[52] Bashir M. Jeon G. Hernandez, I. and J. Bohr. Are bitcoin users less sociable? an analysis of users' language and social connections on twitter. *Proceedings of the HCI International Conference 2014, Heraklion, Crete, Greece, June 22-27*, pages 26–31, 2014. [cited at p. 10]

[53] Garrick Hileman. State of blockchain q1 2016. 2016. [cited at p. 73]

[54] Simona Ibba, Andrea Pinna, Gavina Baralla, and Michele Marchesi. Icos overview: should investors choose an ico developed with the lean startup methodology? In *Proceedings of the XP2018 conference*, page 15. Springer (in press), 2018. [cited at p. 70]

[55] Simona Ibba, Andrea Pinna, Matteo Seu, and Filippo Eros Pani. Citysense: blockchain-oriented smart cities. In *Proceedings of the XP2017 Scientific Workshops*, page 12. ACM, 2017. [cited at p. 8, 77]

[56] Poon J. and Dryja T. The bitcoin lightning network: Scalable off-chain instant payments. Technical report, 2016. [cited at p. 54]

[57] Zhengyao Jiang and Jinjun Liang. Cryptocurrency portfolio management with deep reinforcement learning. *arXiv preprint arXiv:1612.01277*, 2016. [cited at p. 86]

[58] Jiong Jin, Jayavardhana Gubbi, Slaven Marusic, and Marimuthu Palaniswami. An information framework for creating a smart city through internet of things. *IEEE Internet of Things Journal*, 1(2):112–121, 2014. [cited at p. 79]

[59] Luke Patrick Johnson, Ahmed Isam, Nick Gogerty, and Joseph Zitoli. Connecting the blockchain to the sun to save the planet. 2015. [cited at p. 85]

[60] S. King and S. Nadal. Ppcoin: Peer-to-peer crypto-currency with proof-of-stake. Technical report, 2012. [cited at p. 52]

[61] Pavneet Singh Kochhar, Tegawendé F Bissyandé, David Lo, and Lingxiao Jiang. Adoption of software testing in open source projects–a preliminary study on 50,000 projects. In *Software Maintenance and Reengineering (CSMR), 2013 17th European Conference on*, pages 353–356. IEEE, 2013. [cited at p. 75]

[62] Pósfai M. Csabai I. Vattay G. Kondor, D. and L Dobos. Do the rich get richer? an empirical analysis of the bitcoin transaction network. *PLoS ONE*, 9(2):e86197, 2014. [cited at p. 5, 19]

[63] A. Kosba, A. Miller, E. Shi, Z. Wen, and C. Papamanthou. Hawk: The blockchain model of cryptography and privacy-preserving smart contracts. In *2016 IEEE Symposium on Security and Privacy (SP)*, pages 839–858, May 2016. [cited at p. 5]

[64] P. Koshy, D. Koshy, M. Henze, and P. McDaniel. An analysis of anonymity in bitcoin using p2p network traffic. *Lecture Notes in Computer Science*, 2014. [cited at p. 37]

[65] What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. `http://arxiv.org/abs/1406.0268`, 2014. [cited at p. 5]

[66] Daniel Larimer. Introducing bitshares object graph. [cited at p. 72]

[67] L.Cocco and M.Marchesi. Modeling and Simulation of the Economics of Mining in the Bitcoin Market. *PLOS ONE*, 11(10):1–31, 10 2016. [cited at p. 6, 10, 52, 55, 56, 59]

[68] A. Lees and M. King. World Payments. Technical report, Capgemini and The Royal Bank of Scotland, 2015. [cited at p. 47, 51]

[69] Yu-Pin Lin, Joy R Petway, Johnathen Anthony, Hussnain Mukhtar, Shih-Wei Liao, Cheng-Fu Chou, and Yi-Fong Ho. Blockchain: The evolutionary next step for ict e-agriculture. *Environments*, 4(3):50, 2017. [cited at p. 5]

[70] M. Lischke and B. Fabian. Analyzing the bitcoin network: The first four years. *Future Internet*, 8(1):7, 2016. [cited at p. 5, 19]

[71] Michal Lom, Ondrej Pribyl, and Tomas Zelinka. Hybrid-agile approach in smart cities procurement, 2016. [cited at p. 79]

[72] Loi Luu, Duc-Hiep Chu, Hrishi Olickel, Prateek Saxena, and Aquinas Hobor. Making smart contracts smarter. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, pages 254–269, New York, NY, USA, 2016. ACM. [cited at p. 5]

[73] C. Malmo. Bitcoin is unsustainable. `http://motherboard.vice.com/read/bitcoin-is-unsustainable`, 2015. [cited at p. 45, 47]

[74] K. Mannaro, A. Pinna, and M. Marchesi. Crypto-trading: Blockchain-oriented energy market. In *AEIT International Annual Conference*, 2017. [cited at p. 8, 84]

[75] Katiuscia Mannaro, Gavina Baralla, Andrea Pinna, and Simona Ibba. A blockchain approach applied to a teledermatology platform in the sardinian region (italy). *Information*, 9(2), 2018. [cited at p. 77]

[76] G. Mark. Virtual Bitcoin Mining Is a Real-World Environmental Disaster. `www.Bloomberg.com`, 2013. [cited at p. 55]

[77] H. McCook. Under the Microscope: The True Costs of Banking. `http://www.coindesk.com/microscope-true-costs-banking/`, 2014. [cited at p. 50]

[78] J. McLean. Banking on blockchain: charting the progress of distributed ledger technology in financial services. Technical report, Finextra Research Ltd, 101 St Martin's Lane, London, WC2N 4AZ, United Kingdom, 2016. [cited at p. 44, 46, 47, 49, 51]

[79] S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelkrt, and S. Savage. A fistful of bitcoins: Characterizing payments among men with no names. *Communications of the ACM, Vol. 59 NO. 4*, April 2016. [cited at p. 19, 37]

[80] Esther Mengelkamp, Benedikt Notheisen, Carolin Beer, David Dauer, and Christof Weinhardt. A blockchain-based smart grid: towards sustainable local energy markets. *Computer Science-Research and Development*, pages 1–8, 2017. [cited at p. 85]

[81] Mihail Mihaylov, Sergio Jurado, Narcís Avellana, Kristof Van Moffaert, Ildefons Magrans de Abril, and Ann Nowé. Nrgcoin: Virtual currency for trading of renewable energy in smart grids. In *European Energy Market (EEM), 2014 11th International Conference on the*, pages 1–6. IEEE, 2014. [cited at p. 85]

[82] Mihail Mihaylov, Sergio Jurado, Kristof Van Moffaert, Narcís Avellana, and Ann Nowé. Nrg-x-change-a novel mechanism for trading of renewable energy in smart grids. In *SMARTGREENS*, pages 101–106, 2014. [cited at p. 85]

[83] Harlan D Mills, Michael Dyer, and Richard C Linger. Cleanroom software engineering. *IEEE Software*, 4(5):19, 1987. [cited at p. 71]

[84] Khosrow Moslehi and Ranjit Kumar. A reliability perspective of the smart grid. *IEEE Transactions on Smart Grid*, 1(1):57–64, 2010. [cited at p. 86]

[85] T. Murata. Petri nets: Properties, analysis and applications. *Proceedings of the IEEE*, 77(4):541–580, 1989. [cited at p. 19]

[86] Alessandro Murgia, Marco Ortu, Parastou Tourani, Bram Adams, and Serge Demeyer. An exploratory qualitative and quantitative analysis of emotions in issue report comments of open source systems. *Empirical Software Engineering*, pages 1–44, 2017. [cited at p. 6]

[87] Kumiyo Nakakoji, Yasuhiro Yamamoto, Yoshiyuki Nishinaka, Kouichi Kishida, and Yunwen Ye. Evolution patterns of open-source software systems and communities. In *Proceedings of the international workshop on Principles of software evolution*, pages 76–85. ACM, 2002. [cited at p. 76]

[88] S. Nakamoto. Bitcoin: A Peer-to-Peer Electronic Cash System. Technical Report www.bitcoin.org, 2009. [cited at p. 50]

[89] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. *URL: http://www.bitcoin.org/bitcoin.pdf*, 2009. [cited at p. 5]

[90] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. May 2009. [cited at p. 9, 11]

[91] S. Naumann, M. Dick, E. Kern, and T. Johann. The GREENSOFT Model: A reference model for green and sustainable software and its engineering. *Sustainable Computing: Informatics and Systems*, 1:294–304, 2011. [cited at p. 43]

[92] Jude Nelson, Muneeb Ali, Ryan Shea, and Michael J Freedman. Extending existing blockchains with virtualchain. In *Workshop on Distributed Cryptocurrencies and Consensus Ledgers (DCCL'16),(Chicago, IL)*, 2016. [cited at p. 72]

[93] M. E. J. Newman. The structure and function of complex networks. *SIAM Review, Philadelphia, PA, USA*, 2013. [cited at p. 18]

[94] Nxt. Whitepaper: Nxt. Technical report, From Nxt Wiki, 2016. [cited at p. 53]

[95] K. J. O'Dwyer and D. Malone. Bitcoin mining and its energy footprint. In *25th IET Irish Signals Systems Conference 2014 and 2014 China-Ireland International Conference on Information and Communications Technologies (ISSC 2014/CIICT 2014)*, pages 280–285, June 2014. [cited at p. 5, 48]

[96] Svein Ølnes, Jolien Ubacht, and Marijn Janssen. Blockchain in government: Benefits and implications of distributed ledger technology for information sharing, 2017. [cited at p. 5]

[97] Matteo Orrú, Ewan Tempero, Michele Marchesi, Roberto Tonelli, and Giuseppe Destefanis. A curated benchmark collection of python systems for empirical studies on software engineering. In *Proceedings of the 11th International Conference on Predictive Models and Data Analytics in Software Engineering*, page 2. ACM, 2015. [cited at p. 7]

[98] M Ortu, G Destefanis, S Counsell, S Swift, R Tonelli, and M Marchesi. Arsonists or firefighters? affectiveness in agile software development. 2016. [cited at p. 6]

[99] Marco Ortu, Giuseppe Destefanis, Steve Counsell, Michele Marchesi, and Roberto Tonelli. Connecting the dots: measuring effectiveness and affectiveness in software systems. In *Proceedings of the 2nd International Workshop on Emotion Awareness in Software Engineering*, pages 52–53. IEEE Press, 2017. [cited at p. 6]

[100] Marco Ortu, Giuseppe Destefanis, Mohamad Kassab, Steve Counsell, Michele Marchesi, and Roberto Tonelli. Would you mind fixing this issue? an empirical analysis of politeness and attractiveness in software developed using agile boards. In *Agile Processes, in Software Engineering, and Extreme Programming*, pages 129–140. Springer International Publishing, 2015. [cited at p. 89]

[101] Marco Ortu, Giuseppe Destefanis, Stephen Swift, and Michele Marchesi. Measuring high and low priority defects on traditional and mobile open source software. In *Proceedings of the 7th International Workshop on Emerging Trends in Software Metrics*, pages 1–7. ACM, 2016. [cited at p. 78]

[102] Aafaf Ouaddah, Anas Abou Elkalam, and Abdellah Ait Ouahman. Fairaccess: a new blockchain-based access control framework for the internet of things. *Security and Communication Networks*, pages n/a–n/a, 2017. SCN-16-0184. [cited at p. 79, 85]

[103] Sunoo Park, Krzysztof Pietrzak, Albert Kwon, Joël Alwen, Georg Fuchsbauer, and Peter Gazi. Spacemint: A Cryptocurrency Based on Proofs of Space. *IACR Cryptology ePrint Archive*, 2015:528, 2015. [cited at p. 53]

[104] Andrea Pinna. A petri net-based model for investigating disposable addresses in bitcoin system. In CEUR Workshop Proceedings, editor, *2nd international workshop on Knowledge Discovering on the Web (KDWEB)*, volume 1748. ceur-ws.org, 2016. [cited at p. 7, 18]

[105] Andrea Pinna and Simona Ibba. A blockchain-based decentralized system for proper handling of temporary employment contracts. *CoRR*, abs/1711.09758, 2017. [cited at p. 77]

[106] Andrea Pinna, Roberto Tonelli, Matteo Orrú, and Michele Marchesi. A petri nets model for blockchain analysis. *The Computer Journal*, 2018. [cited at p. 7]

[107] Simone Porru, Andrea Pinna, Michele Marchesi, and Roberto Tonelli. Blockchain-oriented software engineering: challenges and new directions. In *Proceedings of the 39th International Conference on Software Engineering Companion*, pages 169–171. IEEE Press, 2017. [cited at p. 8, 9, 70]

[108] G Prisco. An energy blockchain for european prosumers, 2016. [cited at p. 85]

[109] J. Quiggin. Bitcoins are a waste of energy - literally. `http://www.abc.net.au/news/2015-10-06/quiggin-bitcoins-are-a-waste-of-energy/6827940`, 2016. [cited at p. 58]

[110] J. Redman. The Segregated Witness Concept: A 'Turning Point' for Bitcoin? `https://news.bitcoin.com/segregated-witness-concept-turning-point-bitcoin/`, 2016. [cited at p. 54]

[111] Fergal Reid and Martin Harrigan. An analysis of anonymity in the bitcoin system. *CoRR*, abs/1107.4524, 2011. [cited at p. 18]

[112] R Rendina, MR Guarniere, R Niccolai, G Pazienza, A Gualano, S Malgarotti, A Danelli, B Jansson, F Alvarez, W Lovison, et al. The realization and commissioning of the±500 kv 1000 mw hvdc link sardinia island–italian peninsula (sapei). *CIGRE paper B1-101*, pages 1–11, 2012. [cited at p. 86]

[113] P Rizzo. French bank bnp is testing blockchain for mini-bonds. 2016. [cited at p. 49]

[114] P Rizzo. Hong kong's central bank to test blockchain. 2016. [cited at p. 49]

[115] D. Ron and A. Shamir. Quantitative Analysis of the Full Bitcoin Transaction Graph. Technical report. [cited at p. 5]

[116] Dorit Ron and Adi Shamir. Quantitative analysis of the full bitcoin transaction graph. In Ahmad-Reza Sadeghi, editor, *Financial Cryptography and Data Security*, volume 7859 of *Lecture Notes in Computer Science*, pages 6–24. Springer Berlin Heidelberg, 2013. [cited at p. 18, 25, 28, 37]

[117] M. Roth. Sustainability report. Technical report, DZ Bank, Platz der Republik, 60325 Frankfurt am Main, Germany, 2015. [cited at p. 49]

[118] T. Ruffing, P. Moreno-Sanchez, and A. Kate. Coinshuffle: Practical decentralized coin mixing for bitcoin. *MMCI, Saarland University*, 2014. [cited at p. 18]

[119] A. Saxena, J. Misra, and A. Dhar. Increasing anonymity in bitcoin. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8438:122–139, 2014. cited By 0. [cited at p. 10]

[120] Pradip Kumar Sharma, Seo Yeon Moon, and Jong Hyuk Park. Block-vn: A distributed blockchain based vehicular network architecture in smart city. *Journal of Information Processing Systems*, 13(1):84 195, 2017. [cited at p. 79]

[121] David Siegel. Understanding the dao hack for journalists. *Medium*, 2016. [cited at p. 70]

[122] M. Stocchi, I. Lunesu, S. Ibba, G. Baralla, and M. Marchesi. The future of bitcoin: a synchrosqueezing wavelet transform to predict search engine query trends. In CEUR Workshop Proceedings, editor, *2nd international workshop on Knowledge Discovering on the Web (KDWEB)*, volume 1748. ceur-ws.org, 2016. [cited at p. 86]

[123] Marco Stocchi and Michele Marchesi. Fast wavelet transform assisted predictors of streaming time series. *Digital Signal Processing*, 2017. [cited at p. 5]

[124] Melanie Swan. *Blockchain: Blueprint for a new economy*. O'Reilly Media, Inc., 2015. [cited at p. 6, 69]

[125] Nick Szabo. Formalizing and securing relationships on public networks. *First Monday*, 2(9), 1997. [cited at p. 6]

[126] Feng Tian. An agri-food supply chain traceability system for china based on rfid blockchain technology. In *2016 13th International Conference on Service Systems and Service Management (ICSSSM)*, pages 1–6, June 2016. [cited at p. 79]

[127] Concas G. Marchesi M. Pinna S. Turnu, I. and R. Tonelli. A modified yule process to model the evolution of some object-oriented system properties. *In Information Sciences*, 181(4):883–902, 2011. Amsterdam, Netherlands. [cited at p. 17, 18, 19]

[128] Unicredit. Blockchain technology and applications from a financial perspective. 2016. [cited at p. 69]

[129] Andrew Urquhart. The inefficiency of bitcoin. *Economics Letters*, 148:80 – 82, 2016. [cited at p. 5, 47]

[130] Harald Vranken. Sustainability of bitcoin and blockchains. *Current Opinion in Environmental Sustainability*, 28:1 – 9, 2017. [cited at p. 5, 47]

[131] B. Weber. Bitcoin and the legitimacy crisis of money. *Cambridge Journal of Economics*, 40(1):17–41, 2016. [cited at p. 10]

[132] M. G. Wilson and A. Yelowitz. Characteristics of bitcoin users: An analysis of google search data. *SSRN Electronic Journal*. [cited at p. 10]

[133] A. V. Wirdum. Segregated witness officially introduced with release of bitcoin core 0.13.1. `https://bitcoinmagazine.com/articles/segregated-witness-officially-introduced-with-release-of-bitcoin-core-1477611260`, 2016. [cited at p. 54]

[134] A. V. Wirdum. The Segregated Witness Timeline: From Idea to Adoption in Six Steps. `https://bitcoinmagazine.com/articles/the-segregated-witness-timeline-from-idea-to-adoption-in-six-steps-1461255570`, 2016. [cited at p. 54]

[135] Gavin Wood. Ethereum: A secure decentralised generalised transaction ledger. 2015. [cited at p. 13]

[136] Z. Xu, Pautasso C., Zhu L., Giramoli V., Ponomarev A., and Chen S. The Blockchain as a Software Connector. In *13th Working IEEE/IFIP Conference on Software Architecture*, 2016. [cited at p. 6]

[137] H Yu. What wall street's obsession with blockchain means for the future of banking. 2016. [cited at p. 49]

[138] Andrea Zanella, Nicola Bui, Angelo Castellani, Lorenzo Vangelista, and Michele Zorzi. Internet of things for smart cities. *IEEE Internet of Things journal*, 1(1):22–32, 2014. [cited at p. 78]

[139] Yu Zhang and Jiangtao Wen. The IoT electric business model: Using blockchain technology for the internet of things. *Peer-to-Peer Networking and Applications*, page 1–12, 2016. [cited at p. 79]

[140] H. Ziegeldorf, J., F. Grossmann, M. Henze, N. Inden, and K. Wehrle. Coinparty: Secure multi-party mixing of bitcoins. *CODASPY 2015 - Proceedings of the 5th ACM Conference on Data and Application Security and Privacy pp. 75-86*, 2015. [cited at p. 18, 36]