Final Year Technical Project Report

National Institute of Posts and Telecommunications

Implementation of a Blockchain based application to manage lost Bitcoins

Authors:

Mohamed Amine AJINOU

Abdelkarim HABOUCH

Supervisors:

Pr. Meryem AYACHE

Pr.. Amjad GAWANMEH

Cybersecurity and Digital Trust Engineering

Academic Year: 2020 - 2021

Acknowledgments

We are overwhelmed in all humbleness and gratefulness to acknowledge our depth to all those who have helped me to put these ideas, well above the level of simplicity and into something concrete.

We thank God for all the blessings and patience he gave us to complete this project. We would also like to express our special gratitude and thanks to Pr. Meryem AYACHE as well as Pr. Amjad GAWANMEH who gave us this special opportunity to work on this wonderful project on the topic "Implementation of a Blockchain based application to manage lost Bitcoins" , which also helped us to do lot of research and get to know more about the interesting world of blockchain and bitcoin.

We would like to thank our INPT professors who helped us a lot in gathering different information, collecting data and guiding us throughout all our academic years despite their busy schedules.

Any attempt at any level can 't be satisfactorily completed without the support and guidance of our parents and classmates.

Thank you all.

Abstract

While it is not possible to count the exact number of lost Bitcoins, it is estimated that there are around 4 million bitcoins lost out of the 18 million that are mined so far. Another report stated that about 20% of all mined bitcoins are lost, including a study from Wall Street journal!

Bitcoin owners typically hold their tokens in digital wallets, protected by cryptography and accessible only via private key. If an owner loses a private key, its associated bitcoins will be lost permanently.

In this project, we simulated the actual bitcoin blockchain, and established a probabilistic model to analyze the likelihood of a wallet to be lost. This is a starting point to detect the most likely wallets to be lost, recover their bitcoins, get them back into circulation by adding them to the mining pool and use them later for transaction reward fees.

List of Figures

Table of Contents

Introduction

Bitcoin is the leading cryptocurrency nowadays used by the richest investors in the world. It contains many features and is based on the blockchain system, which ensures the integrity and security of all transactions. However, it is very crucial for all bitcoin wallet owners to save their public keys, and most importantly, their private keys to be able to perform any transaction safely. If the private key is lost, access to the wallet is impossible, thus all the balance inside is frozen and this wallet is considered to be permanently lost. Throughout this project, we will try to detect lost wallets based on different parameters and get their lost bitcoins back into network circulation by using them during new transaction rewards.

Chapter I

Introduction about Bitcoin and Blockchain

1. Bitcoin

1.1 History of Bitcoin

Bitcoin was conceptualized by Satoshi Nakamoto in 2008. We do not know whether it is a person or a group of people. This anonymous person or group of persons still remains a mystery. What we know is that Nakamoto has claimed to be a man living in Japan born on April 15th, 1975. However, there are a lot of theories and speculation about the identity of Nakamoto. Some people say that the identity of Nakamoto is based on a number of cryptography and computer science experts living in the US and Europe, not necessarily Japanese people. In November 2017, Nakamoto was believed to own up to roughly 1 million Bitcoins, the value of this 1 million comes to be 7.2 billion US dollars, which is a huge amount of money owned by someone who is not known to anybody in the world. ***(Kulkarni,2018)***

In fact, Bitcoin is the first decentralized digital currency, and as such it is a revolutionary technology invention. It changed the way we compute things and the way we operate software and computers. Bitcoin is considered to be the next big wave of change after the internet.

Figure 1: Evolution of Bitcoin

1.2 Bitcoin properties

- It's an international network of payments.
- It uses cryptography to control its creation and management, rather than relying on central authorities such as governments, banks, union territories, or intermediaries.
- It's not printed but is produced by people using software that solves mathematical problems.
- It is controlled and limited in supply, which arrests the hyperinflation problem. For example, whenever African countries were short of currency notes, they had to print more notes, which resulted in hyperinflation and brought the value of the currency down.
- Since the arrival of Bitcoin, the way Bitcoin programs are written means there will always be a maximum of 21 million Bitcoins available across the globe. The moment 21 million Bitcoins have been mined; the program will not generate anymore new Bitcoins. Hence, Bitcoins will be limited in supply and this will arrest the problem of hyperinflation.

1.3 Alternative cryptocurrencies to Bitcoin

Altcoins is an alternate cryptocurrency to Bitcoin. Once Bitcoin became popular, people realized the value, robustness, and flexibility Bitcoin brought and also started liking the fact that Bitcoin appreciated in value. They simply took the source code of the Bitcoin protocol available from GitHub repositories, forked it, modified it as per their needs, and created alternative cryptocurrencies such as: Ethereum, XRP, Cardano... With the increasing popularity of Bitcoin, the usage and rate of Bitcoin have skyrocketed.

Figure 2: Different cryptocurrencies

2. Blockchain

2.1 What is Blockchain

Blockchain is a system of recording information in a way that makes it difficult or impossible to change, hack, or cheat the system.

A blockchain is essentially a digital ledger of transactions that is duplicated and distributed across the entire network of computer systems on the blockchain. Each block in the chain contains a number of transactions, and every time a new transaction occurs on the blockchain, a record of that transaction is added to every participant's ledger. The decentralized database managed by multiple participants is known as Distributed Ledger Technology (DLT). ***("Blockchain explained",2021)***

Blockchain is a type of DLT in which transactions are recorded with an immutable cryptographic signature called a hash.

2.2 Blockchain features

- Secure: It is really impossible for anyone to tamper with transactions or ledger records present in the blockchain, which makes it more secure, so it is seen as are liable source of information.
- Global reach: Blockchain has been adopted worldwide and has the backing of many investors from banking and non-banking sectors, which makes it a globally accepted technology stack.
- Automated operations: Operations are fully automated through software. Private companies are not needed to handle operations, which is why there is no mediation required to carry out the transactions, and trust is assured, so people can carry out their own transactions.
- Open source: Blockchain is an open-source technology. All the operations are carried out by the open-source community.

- Distributed: Blockchain works in a distributed mode, in which records are stored in all nodes in the network. If one node goes down, it doesn't impact any other nodes or any other records, because they are globally distributed across all the nodes.
- Flexible: Blockchain is programmable, using basic programming concepts and programming semantics, which makes blockchain very flexible.

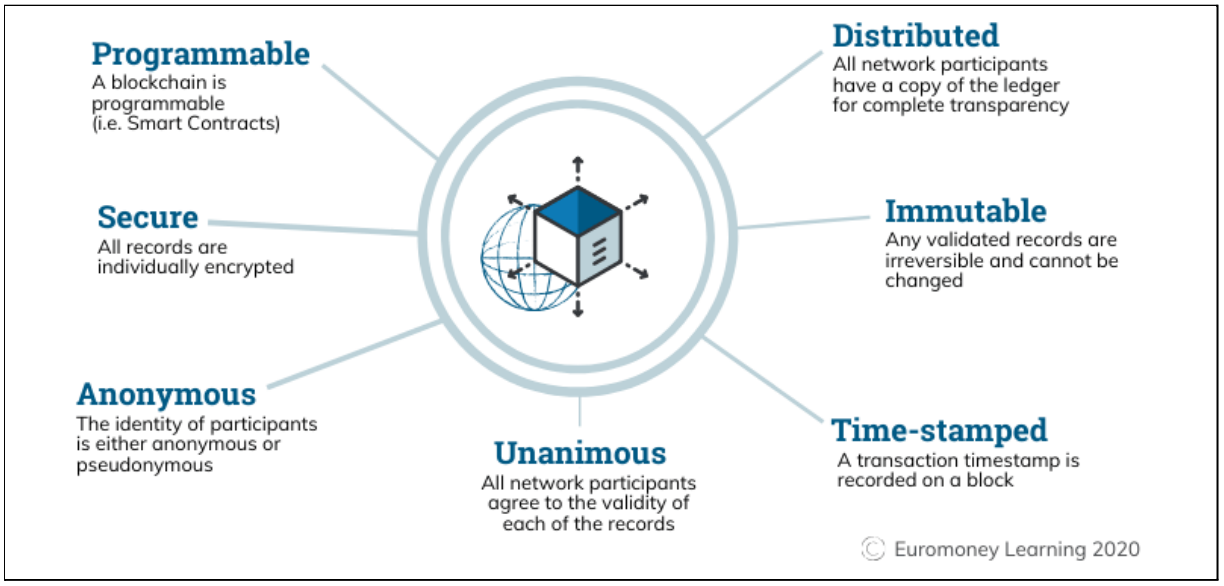

Figure 3: Blockchain features

2.3 Blockchain use cases

- **International Payments**

Blockchain provides a way to securely and efficiently create a tamper-proof log of sensitive activity. This makes it excellent for international payments and money transfers.

- **Trade Finance**

Notable strategies of exchange financing have been a major torment point for businesses since the moderate forms frequently hinder commerce and make liquidity difficult to oversee. Cross-border exchange includes a huge number of factors when communicating data – such as the country of origin and product details – and exchanges produce tall volumes of documentation. Blockchain has the capacity to streamline exchange back bargains and

streamline the method over borders. It empowers ventures to more effortlessly execute with each other past territorial or geographic boundaries.

- **Insurance:**

Seemingly the most noteworthy blockchain application for insurance is through smart contracts. These contracts permit clients and safeguards to oversee claims in a straightforward and secure way. All contracts and claims can be recorded on the blockchain and approved by the network, which would kill invalid claims, since the blockchain would dismiss numerous claims on the same accident. ***(Business Insider,2020)***

- **Health care:**

Health information that's reasonable for blockchain incorporates common data like age, sex, and possibly essential medical history information like immunization history or crucial signs. On its own, none of this data would be able to particularly distinguish any specific patient, which is what permits it to be stored on a shared blockchain that can be accessed by various people without undue protection concerns.

- **Energy:**

Blockchain innovation might be utilized to execute energy supply exchanges, but also to encourage providing the basis for metering, billing, and clearing processes, agreeing to PWC. Other potential applications incorporate reporting ownership, resource administration, origin guarantees, emission allowances, and renewable energy certificates.

3. Relation between Bitcoin and Blockchain

The operation of Bitcoins and the blockchain is particular, as the blockchain requires a virtual currency. When one person sends a certain amount of Bitcoins to another, this transaction (or exchange) is immediately recorded in a block. This block is then validated by nodes also called "miners" which correspond to a huge network of several tens of thousands of computers. This is done using cryptographic techniques called "Proof-of-Work". The block is dated and then added to the larger whole that is the blockchain. All users can have access to the information, which cannot be erased. Finally, the person who ordered the Bitcoins can then receive them.

The great advantage of the blockchain and therefore of using Bitcoin as a currency is that this computer protocol is totally secure, encrypted and invulnerable to cyber-attacks.

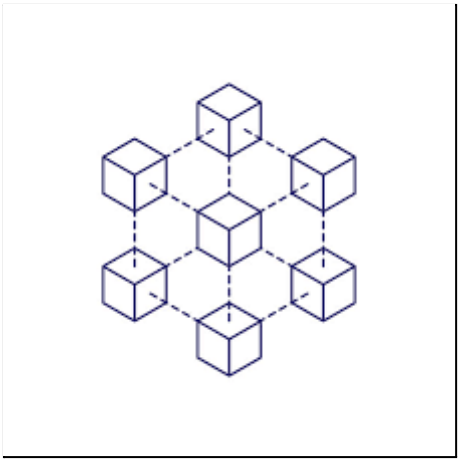

Figure 4: Blocks in a blockchain

Chapter II

Bitcoin Security and Risks

1. Bitcoin Wallets Security

Bitcoin's security (the public/private key system) is based on cryptographic algorithms published by the NSA (National Security Agency, an American government agency): ECDSA, SHA256. These algorithms are unanimously considered as the most secure, they are used in banking, military, intelligence. If they were "cracked", all the electronic systems on the planet would fall. A bitcoin address is of the form: 175tWpb8K1S7mH4Zx6rewF9wQrcPv245W. There are more than 10^48 possible addresses, more than the number of atoms on earth. ***(Philippe Herlin,2013)***

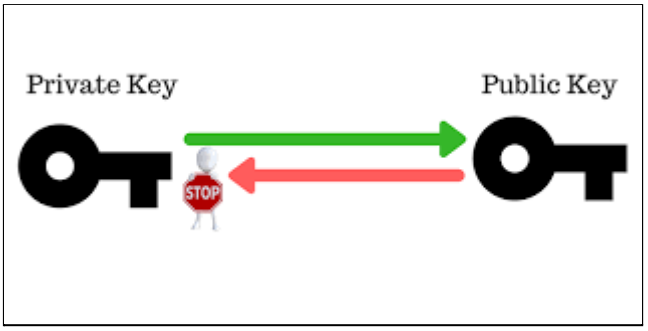

Figure 5: Public and Private Key

Generally speaking, a centralized system that depends on a few passwords is fragile because if the secret is broken, it falls apart (disclosure of PlayStation and LinkedIn accounts, for example). Whereas a decentralized system in which each participant is responsible for his or her private key avoids this risk; there is simply no "one" password that would crack the bitcoin network. There is an official project site (bitcoin.org) but it is separate from the network itself; hacking into it would not give access to personal accounts.

2. Private keys and Bitcoin wallets

Bitcoins do not physically exist. The only evidence of their existence is when they are associated with addresses, which are referred to in transactions. When an address is initially created, a pair of public and private keys are generated with it. The public key is made known to the public and the private key is kept only by the owner of the address. When the owner wants to spend all or a portion of their BTCs, the owner provides a digital signature signed with the private key and sends the BTC request to the Bitcoin network. In other words, one has to know both the address and its private key to spend the BTC. Therefore, it is advised to keep this information in a safe place. ***(Xun Wu and Weimin Sun, 2018)*** It is generally good

practice to keep the address and private keys in separate places. To prevent a digital copy getting lost, an owner should maintain physical copies of printouts. To make conversion easier, an owner can print a QR code and later scan the QR code whenever it is needed. However, if an owner loses a private key, its associated BTCs will be lost permanently.

3. Lost Bitcoin wallets

Around 4 million bitcoins are lost out of the 18 million that are mined so far. Another report expressed that nearly 20% of all mined bitcoins are lost, counting a consider from Wall Street journal! Bitcoin owners typically hold their tokens in digital wallets, secured by cryptography and available as it were by means of private key, it's exceptionally troublesome for others to get to those holdings.

Figure 6: Bitcoin Wallet

Bitcoin can be utilized only through this private key. When a Bitcoin address is created, a public key and a private key is created. The public key is utilized as an address to send and get bitcoins, whereas the private key is utilized to secure the tokens within the public address. The private key is necessary to send transactions, which means one cannot spend Bitcoins without the private key. For occurrence, here is an case of the pair of keys:

- Public: 18uTHBP6PEiyK4hVmazXfdVbN4TcBNY591
- Private: Kz9qXEGX9BxA9VdxpdohPrz6Bq95SxJynMbaZi3BF9NnoXVmJdp1

Unlike passwords, these keys are not conceivable to keep in mind!

In addition, there is a huge number of bitcoin wallets that hold very small value i.e. less than 10000 satoshi = 0.0001 BTC), and quite a large number that holds less than1 $ value. Most of these are considered abandoned and it is unlikely that these wallets will be accessed again. Statistically, this number will increase with time, causing more bitcoins to be lost.

Chapter III

Blockchain simulation and Probabilistic Model for lost wallets

1. Problematic

Bitcoin has limited supply, only 21 Million Bitcoins will be mined, having 20% of this stuck so far, and hence cannot be circulated or used is a great concern! Therefore, our problematic for this project can be stated as the following: How to detect lost wallets get their lost bitcoins back into network circulation by using them during new transaction rewards?

Bitcoin has 659776 blocks created so far, this makes it around 55000 blocks per year. Since the lost bitcoins will never be used, the last time a bitcoin unit (Satushi is moved) can be used as a criteria to determine the likelihood of this unit to be lost. On the other hand, more than 50% of the overall wallets in the world hold a balance less than 0.001 BTC = 100 000 satoshi, this balance will be used as well as a second criteria for our probabilistic model.

2. Environment Simulation

In order to simulate blockchain with its different components, Python can be a very good tool for this aim. Therefore, different modules were put together in this simulation to ensure a randomized dataset with respect to some encryption and hashing constraints.

The structure of the Python project has two main directories: “blockchain” and “crypto”. The first one contains different classes to simulate wallets, transactions and blocks. The second folder contains functions for hashing and encryption since it is very essential to ensure a minimum amount of security and encryption interconnection between all the blocks constituting the whole blockchain for integrity purposes.

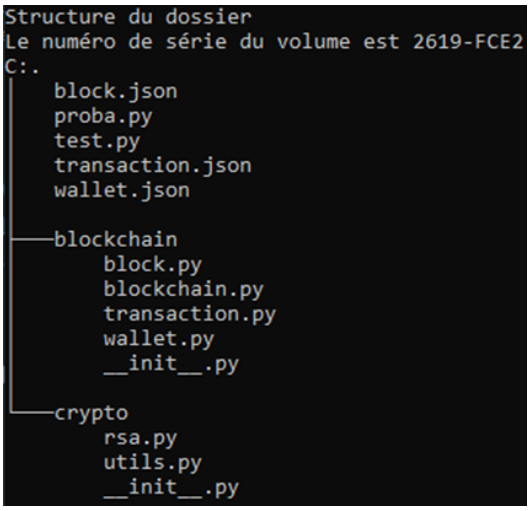

Figure 7: Project structure

The file “test.py” is for running the simulation, and “proba.py” contains the probabilistic functions used for analyzing generated wallets and detecting whether they are lost or not. Concerning the three json files, they store the most important information from generated wallets, transactions and blocks for further analysis and exploitation.

3. Simulation structure

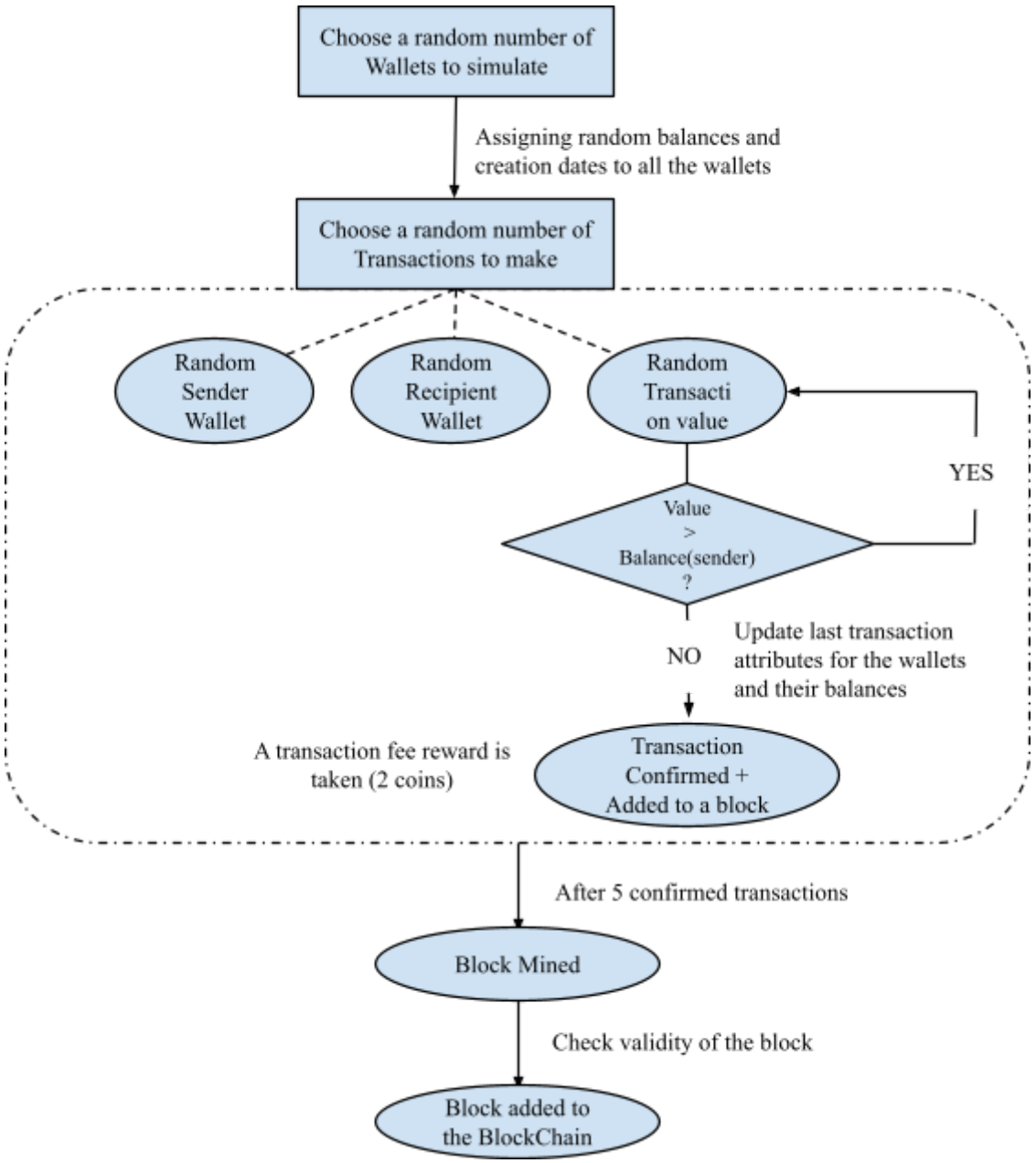

Figure 8: Simulation overview

The program asks for the first input n, which is the number of wallets to simulate. Then, the wallets are created with random balances and random creation dates assigned to each one. To simulate a closer simulation to reality, the program takes into consideration the Bitcoin wallet distribution ***(BitInfoCharts)***

Balance, BTC	Addresses	% Addresses (Total)	Coins	$USD	% Coins (Total)
(0 - 0.001)	19553997	51.05% (100%)	3,978 BTC	134,207,147 USD	0.02% (100%)
[0.001 - 0.01)	9747695	25.45% (48.95%)	37,055 BTC	1,250,223,465 USD	0.2% (99.98%)
[0.01 - 0.1)	5835719	15.23% (23.51%)	189,455 BTC	6,392,108,218 USD	1.01% (99.78%)
[0.1 - 1)	2371517	6.19% (8.27%)	741,813 BTC	25,028,387,721 USD	3.96% (98.77%)
[1 - 10)	651032	1.7% (2.08%)	1,672,357 BTC	56,424,452,410 USD	8.92% (94.81%)
[10 - 100)	130544	0.34% (0.38%)	4,247,435 BTC	143,306,238,911 USD	22.65% (85.9%)
[100 - 1,000)	13947	0.04% (0.04%)	3,948,436 BTC	133,218,163,894 USD	21.06% (63.24%)
[1,000 - 10,000)	2063	0.01% (0.01%)	5,180,826 BTC	174,798,346,375 USD	27.63% (42.19%)
[10,000 - 100,000)	80	0% (0%)	2,047,564 BTC	69,083,721,331 USD	10.92% (14.56%)
[100,000 - 1,000,000)	4	0% (0%)	681,888 BTC	23,006,555,838 USD	3.64% (3.64%)

Figure 9: Bitcoin wallet distribution

The percentages of balance ranges are applied to the number of generated wallets and their balances are therefore updated respecting what is stated on the figure above.

A second input is asked to choose the number of transactions to simulate. The program chooses randomly the sender and recipients ids and also the transaction value. The transaction is not confirmed unless it verifies all conditions, mainly having sufficient funds in the sender wallet to make the transaction. Once it is confirmed, 2 coins are taken away from the mining pool as a transaction fee, then the transaction is automatically added to the current block. Thereafter five consecutive successful transactions, the current block is mined and added to the blockchain after checking the validity of the hashes to ensure the integrity.

NB: For our project, we simulated 100.000 wallets with 200.000 transactions.

At the end of the simulation, all the output is stored on json files for further exploitation and analysis.

4. Probabilistic model

Before moving the model implemented, here is a sample of the dataset generated using pandas dataframes by importing the json files.

```
data_wallet = pd.read_json (r'wallet.json').T
data_wallet.sample(10)
```

	id	name	creation	balance	last_transaction
wallet 5759	5759	wallet 5759	2015-06-09	90	2019-09-19
wallet 2672	2672	wallet 2672	2012-08-19	0	2019-04-21
wallet 6964	6964	wallet 6964	2010-06-03	72	2019-12-29
wallet 1151	1151	wallet 1151	2014-06-28	18	2019-08-04
wallet 8183	8183	wallet 8183	2010-04-25	319	2019-09-16
wallet 3939	3939	wallet 3939	2011-08-09	7	2016-07-28
wallet 3545	3545	wallet 3545	2011-09-16	12	2019-12-29
wallet 7364	7364	wallet 7364	2015-06-07	46	2019-12-25
wallet 4430	4430	wallet 4430	2012-05-12	13	2019-12-17
wallet 8507	8507	wallet 8507	2010-12-25	821	2018-10-30

Figure 10: Wallet pandas dataframe

```
data_transaction = pd.read_json (r'transaction1.json').T
data_transaction.sample(10)
```

	sender	recipient	value	date	transaction_id
transaction 9393	7909	9784	12	2019-12-08	319761a0ddf25576d849afa78218d86f6d63e20469c680...
transaction 5128	2039	1313	11	2018-11-30	efd3d56203a2bf9796be045f390f6fbf1fd51c03208a36...
transaction 2581	8289	9809	11	2019-12-26	9689cd914b93c2f45c6bd1b1ba5913fbecbded2b51e443...
transaction 5182	3752	2926	2	2019-11-07	deabb6141742dd0044e1c7688a7fe07db79280175a6f21...
transaction 6868	3885	4376	5	2019-12-31	0b1e0977d37b200da4060708f8e87a4484a2873c620e32...
transaction 6548	272	6744	10	2019-12-31	58933c3e7860108a46d69d4853d852cc1470b61974f142...
transaction 7678	4886	3945	8	2019-12-31	80394e899a4a1e29027d6b54f6461bc597b195b28afeb3...
transaction 5223	7849	8869	10	2019-12-29	9510ec63ef4177582019690be874ebf3019278f1525f45...
transaction 1365	5959	7010	6	2019-01-13	582df44540463bf6d66a78271027b91c440abb404f2b80...
transaction 7077	6063	9464	19	2019-07-28	7057db0f303e71d744fd4138e8a949f3e1f6539e9a3327...

Figure 11: Transaction pandas dataframe

Now that all the dataset is available for analysis, we chose a probabilistic model that takes the last transaction and balance of wallets as inputs. Two functions were established to calculate the likelihood of a wallet to be lost. They are as the following:

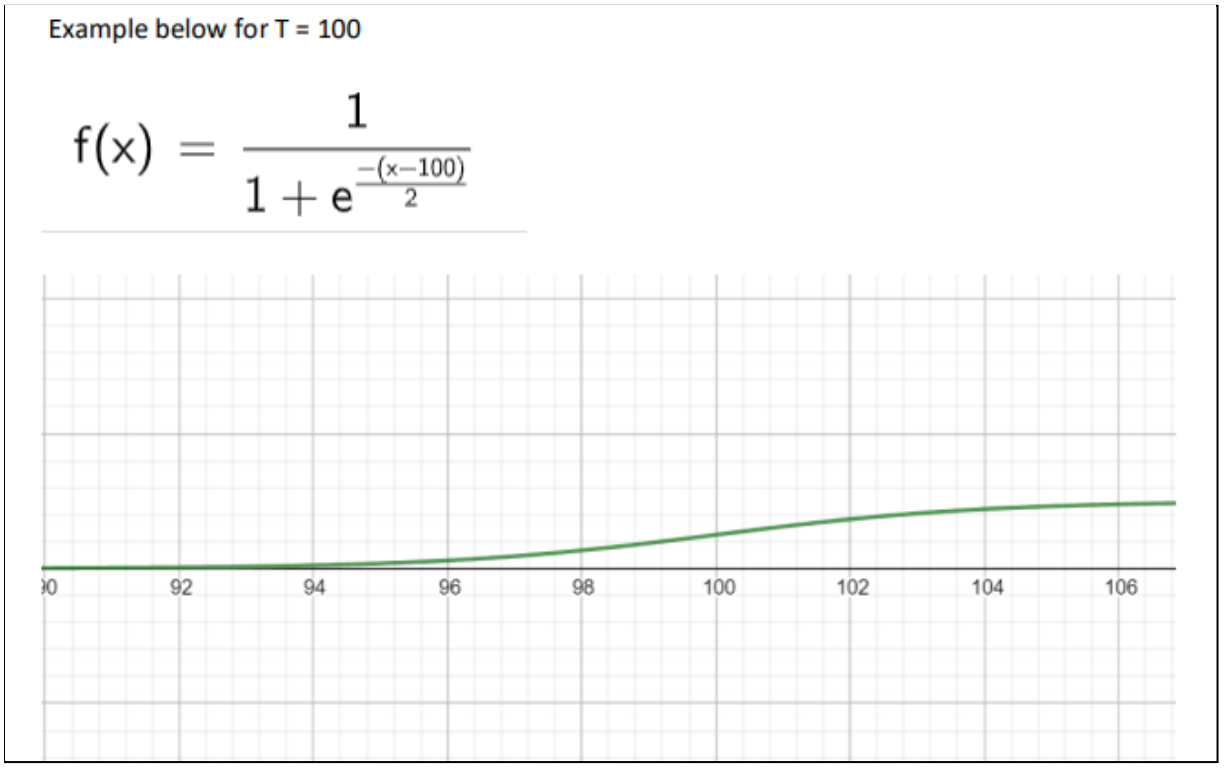

Figure 12: Function 1 (probabilistic model)

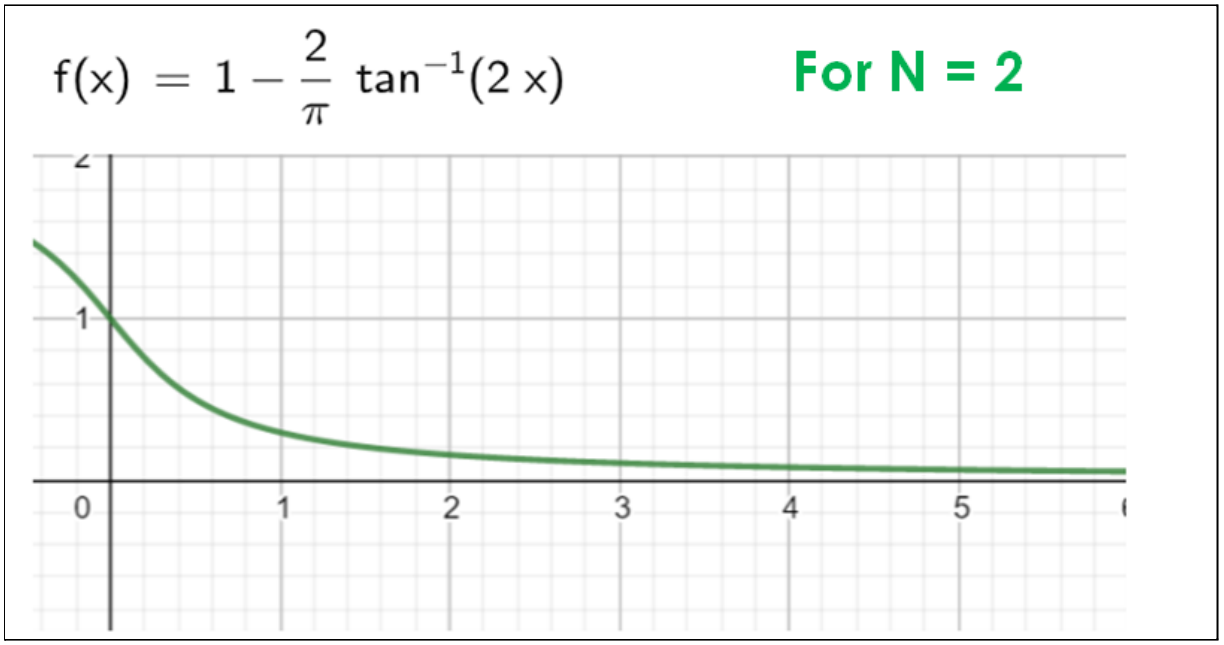

Figure 13: Function 2 (probabilistic model

- The first function takes the last transaction date as an input with the hyperparameter T, it is set to be the median of last transaction date attributes for all wallets to have the most optimal variance for the output values.

- The second function takes the wallet balance as an input with the hyperparameter N, it is set to be the median of all the balances.

After multiplying the output of the functions for each wallet, we get the final probability of the wallet to be lost. The higher the probability, the more likely the wallet is considered lost. The result of mapping the wallets to the calculated probability is plotted using matplotlib python module:

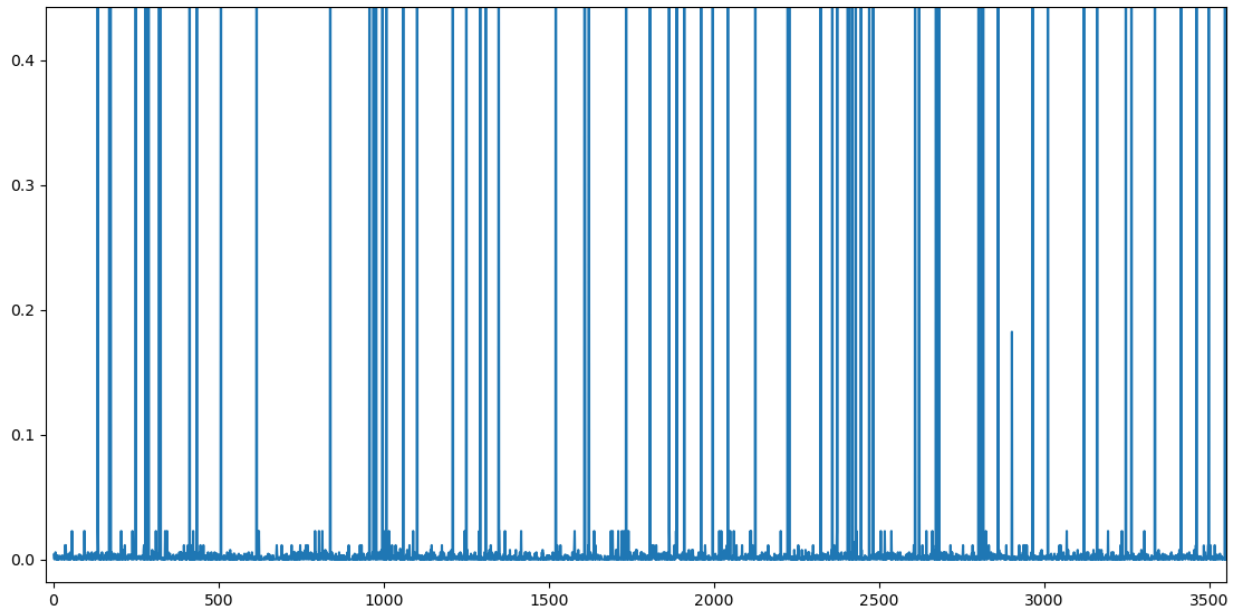

Figure 14: Graph of probabilities and associated wallets

Setting a border value is the best way to decide whether a wallet is most considered lost or not. For our simulation, we set this value to “0.4” Here is some probability values calculated and examples of wallets that are most likely to be lost:

```
The Wallet that is most likely to be lost N° 91:        wallet 1621     Probability ==> 1.0
The Wallet that is most likely to be lost N° 92:        wallet 1609     Probability ==> 1.0
The Wallet that is most likely to be lost N° 93:        wallet 320      Probability ==> 0.9999999999999996
The Wallet that is most likely to be lost N° 94:        wallet 2418     Probability ==> 0.9999999999999993
The Wallet that is most likely to be lost N° 95:        wallet 6690     Probability ==> 0.9999999999999971
The Wallet that is most likely to be lost N° 96:        wallet 2042     Probability ==> 0.9999999999999656
The Wallet that is most likely to be lost N° 97:        wallet 4280     Probability ==> 0.9999999999622486
The Wallet that is most likely to be lost N° 98:        wallet 2801     Probability ==> 0.9999999995400946
The Wallet that is most likely to be lost N° 99:        wallet 4869     Probability ==> 0.9999999992417439
The Wallet that is most likely to be lost N° 100:       wallet 3498     Probability ==> 0.9999999987498471
The Wallet that is most likely to be lost N° 101:       wallet 2860     Probability ==> 0.9999998144608981
The Wallet that is most likely to be lost N° 102:       wallet 134      Probability ==> 0.9999998144608981
The Wallet that is most likely to be lost N° 103:       wallet 3264     Probability ==> 0.999997739675702
The Wallet that is most likely to be lost N° 104:       wallet 434      Probability ==> 0.9999938558253978
The Wallet that is most likely to be lost N° 105:       wallet 324      Probability ==> 0.9999724643088853
```

Figure 15: Wallets more likely to be lost

The step following the application of this probabilistic model consists of taking away the next transaction fees successively from the wallets that are most likely to be lost in order to unlock lost bitcoins and add them back to the mining pool for circulation.

The whole project is hosted on the github following repository:
https://github.com/aminekayy/Lost-Bitcoins

Conclusion

This project is a good start for establishing an algorithm to detect lost wallets and get their balances back into life. However, the probabilistic model can be sharpened using more advanced artificial intelligence techniques and larger datasets. Unfortunately, it cannot be implemented on bitcoin cryptocurrency, but it should be taken into consideration for further similar new altcoins and consolidated by a recovery mechanism if the wallet owner claims that he still has access to his private key.

Bibliography

BitInfoCharts. “Bitcoin wallet distribution.” *Bitcoin Information Charts*, https://bitinfocharts.com/top-100-richest-bitcoin-addresses.html.

“Blockchain explained.” *Euromoney Learning*, 24 May 2021, https://www.euromoney.com/learning/blockchain-explained/what-is-blockchain.

Insider Intelligence. “Use cases of Blockchain Technology.” *Business Insider*, 2 March 2020, https://www.businessinsider.com/blockchain-technology-applications-use-cases.

Kulkarni, Kirankalyan. *Learn Bitcoin and Blockchain*. Birmingham, 2018.

Mohamed Amine AJINOU, and Karim Haboush. “Github repo: Lost Bitcoins.” *Github*, July 2021, https://github.com/aminekayy/Lost-Bitcoins.

Philippe Herlin. *LA RÉVOLUTION DU BITCOIN ET DES MONNAIES COMPLÉMENTAIRES*. Paris, 2013.

Xun Wu, and Weimin Sun. *Blockchain Quick Start Guide*. Birmingham, 2018.