# A Petri net-based model for investigating disposable addresses in Bitcoin system

Andrea Pinna[1]

1 DIEE Dept., University of Cagliari
Piazza D'Armi, 09100 Cagliari, Italy
a.pinna@diee.unica.it

**Abstract.** The most Bitcoin users are very careful about keeping their anonymity. In this work we investigate the use of disposable addresses, a very common method adopted by users to protect their identity and propose a method to recognize these addresses, We applied it 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.

**Keywords:** Blockchain, Chains, Disposable Addresses, De-Anonymization, Petri nets

## 1 Introduction

In 2008, Satoshi Nakamoto invented a payment system, called "Bitcoin cash system", which allows electronic payments without the need for centralized system (eg. credit institutions) [4]. Bitcoin is a decentralized system running on a peer-to-peer network. It is free and is based on a public ledger called Blockchain.

The Blockchain is a shared database containing a continuously growing list of blocks of Bitcoin transactions. It does not store information about the users' identity, and hence identifying the users through the only analysis of the Blockchain is a hard work. Bitcoin system guarantees the pseudo-anonymity (pseudonymity). Users can exchange bitcoins through their pseudonyms called addresses. A Bitcoin address is an alphanumeric identifier that represents the sender (input section) or the destination (output section) of the Bitcoin transaction. The balance of each address is computed as the value of the unspent transactions output (UTXO).

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 [6]), and others are based on the TOR network. One of the most known strategy 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 she receives a new payment or executes a new payment. Despite the adoption of these strategies, to

de-anonymize the users' identity can be possible. Recent researches try to de-anonymize the users' identity by using external data [1, 2] and others propose clustering heuristic to form user networks [3, 5] . In this study we propose an original strategy, based on a Petri net formalism with the aim to recognize the disposable addresses, and hence the chains of transactions in which these addresses are involved. We applied it on the first 180,000 blocks finding interesting results about the number and the statistical distribution of the lengths of these addresses chains.

## 2 Proposed Method

We modeled the Blockchain as a Petri net, a bipartite oriented graph $N$, defined as $N = (P_\alpha, T, Pre, Post)$, where $P_\alpha$ is the set of the *places (addresses)*, $T$ is the set of the *transitions (transactions)*, $Pre$ is the *Pre-incidence* matrix and $Post$ is the *Post-incidence* matrix. The element $ij$ in the Pre matrix defines how many times the address $i$ is in the input section of the transaction $j$, instead, in the Post matrix, it defines how many times the address $i$ is in the output section of the transaction $j$. After having built of the Petri net $N$, we focused our attention on the chains of disposable addresses, 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. We created the set of potentially disposable addresses $A_d$, 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 $\theta_{ds} \in \Theta_{ds}$:
   (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.

## 3 Results

We processed the first 180,000 blocks, analysing the Json files downloaded from the website *blockchain.info*. We built a Petri net model composed of 3,730,480 places (or addresses), 3,142,019 transitions (or transactions), and of the connections between nodes, expressed by Pre and Post matrices.

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

Table 1: The dimension of the sets of potentially disposable addresses, the number of involved transactions and the number of the chains.

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 1 e 2. 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. 1). 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. 3. 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 Log-Log scale, is showed in Fig. 1. This distribution follows a Power-law in the tail (starting from a length about 20).

## 4 Conclusion

This work focuses on the problem of the anonymity in the Bitcoin system, and in particular, on the massive use of disposable addresses to protect the user'

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

Table 2: 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 |

Table 3: Top five chains ordered by length. *Blocks* is the number of blocks that contain each chain from the beginning to the end. *Rate* is the average number of transaction per block for the chain

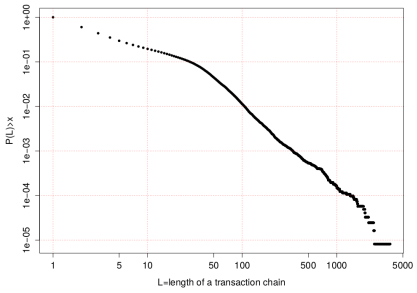


Fig. 1: CCDF of chains lenghts.

identity. We modeled Bitcoin Blockchain as A Petri net-based model, and analyzed the connections between place (bitcoin addresses) and transitions (bitcoin transactions) thanks to the Petri net formalismm, specifically thanks to Pre and Post matrices Results show that over one-third of the addresses registered in the Blockchain are actually disposable addresses. The lengths of the chains are very inhomogeneous and their statistical distributions follow partially a power low. Further, we found that the chains are executed quickly, without waiting for confirmation. In future works we will analyse a higher number of blocks and more complex techniques of anonymization.

## References


1. Biryukov, A., Khovratovich, D., Pustogarov, I.: 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)
2. Koshy, P., Koshy, D., Henze, M., McDaniel, P.: An analysis of anonymity in bitcoin using p2p network traffic. Lecture Notes in Computer Science (2014)
3. Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., McCoy, D., Voelkrt, G.M., Savage, S.: A fistful of bitcoins: Characterizing payments among men with no names. Communications of the ACM, Vol. 59 NO. 4 (April 2016)
4. Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system (May 2009), `http://www.bitcoin.org/bitcoin.pdf`
5. Ron, D., Shamir, A.: Quantitative analysis of the full bitcoin transaction graph. In: Sadeghi, A.R. (ed.) Financial Cryptography and Data Security, Lecture Notes in Computer Science, vol. 7859, pp. 6–24. Springer Berlin Heidelberg (2013), `http://dx.doi.org/10.1007/978-3-642-39884-1_2`
6. Ziegeldorf, J., H., Grossmann, F., Henze, M., Inden, N., Wehrle, K.: 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)