Research and Challenges on Bitcoin Anonymity

Conference Paper · September 2014

DOI: 10.1007/978-3-319-17016-9_1

1 author:

Jordi Herrera-Joancomartí
Autonomous University of Barcelona
111 PUBLICATIONS **1,238** CITATIONS

Research and Challenges on Bitcoin Anonymity*

Jordi Herrera-Joancomartí

Dept. d'Enginyeria de la Informació i les Comunicacions
Universitat Autònoma de Barcelona
08193 Bellaterra, Catalonia, Spain
jordi.herrera@uab.cat

Abstract. Bitcoin has emerged as the most successful crypto currency since its appearance back in 2009. Besides its security robustness, two main properties have probably been its key to success: anonymity and decentralization. In this paper, we provide a comprehensive description on the details that make such cryptocurrency an interesting research topic in the privacy community. We perform an exhaustive review of the bitcoin anonymity research papers that have been published so far and we outline some research challenges on that topic.

1 Introduction

Bitcoin is an online virtual currency based on public key cryptography, proposed in 2008 in a paper [1] authored by someone behind the Satoshi Nakamoto pseudonym. It became fully functional on January 2009 and its broad adoption, facilitated by the availability of exchange markets allowing easy conversion with tradicional currencies (EUR or USD), has brought it to be the most successful virtual currency.

However, in contrast to other virtual payments systems appeared so far, the seminal paper [1] describing the Bitcoin system was not published in the scientific arena but as a forum post on the Internet.[1] Furthermore, the practical development of the ideas proposed in such paper took place on January 2009, when the same author created the first block of the Blockchain and implemented a fully functional bitcoin wallet which allows to operate with such new cryptocurrency. For this reason, the deployment of bitcoin took off without so much attention from the research community and the first research papers on the topic did not appear until late 2011 in the arXiv repository and later published conferences and journals ([2, 3]).

During the 2014, there has been an explosion in the publication of bitcoin research papers, and well established conferences included the topic of *cryptocurrecies* as a"topic of interest". Furthermore, specific workshops were created, like

* This work is published in the proceedings of the 9th International Workshop on Data Privacy Management. Springer. LNCS 8872, pp. 1-14. (2014)

[1] http://web.archive.org/web/20090131115053/http://bitcoin.org/
http://p2pfoundation.ning.com/forum/topics/bitcoin-open-source

the 1st Workshop on Bitcoin Research, held jointly with the 18th International Conference Financial Cryptography and Data Security. The research performed so far related to bitcoin has been very broad, not only in the technical research arena but also in other disciplines, like business and economy, law or sociology.

In this paper, we provide a comprehensive description of the key issues of the bitcoin system in order to allow to new comers to understand the scientific review performed later on. Then, we provide an exhaustive review of the papers that dealt with anonymity issues. Throughout the paper we identify and discuss interesting research challenges.

2 The Bitcoin system

In this section, we point out the main ideas that allow to understand the basic functionality of the bitcoin virtual currency. Such background is needed to understand the meaning of the research performed so far. However, the complexity of bitcoins makes impossible to provide a fully description of the system in this review, so interested readers can refer to [4] for a detailed and more extended explanation on the bitcoin system.

Bitcoin is a cryptocurrency based on accounting entries. For that reason, it is not correct to look at bitcoins as digital tokens since bitcoins are represented as a balance in a bitcoin account. A **bitcoin account** is defined by an Elliptic Curve Cryptography key pair[2]. The bitcoin account is publicly identified by its **bitcoin address**, obtained from its public key using an unidirectional function. Using this public information users can send bitcoins to that address[3]. Then, the corresponding private key is needed to spend the bitcoins of the account. Regarding this definition, it is easy to understand that any user can create any number of bitcoin addresses (generating the key pair) either using any standard crypto-software or self purpose created programs, like bitcoin wallets. Notice that if the user creates such bitcoin accounts in a private manner then, a priori, nobody can link the identity of the user with the value of a bitcoin address.

2.1 Bitcoin payments

Payments in the bitcoin system are performed through transactions between bitcoin accounts. A **bitcoin transaction** indicates a bitcoin movement from source addresses to destination addresses. Source addresses are referred as **input addresses** in a transaction and destination addresses are named **output addresses**. As it can be seen in Figure 1, a single transaction can have one or multiple input addresses and one or multiple output addresses.

A transaction details the exact amount of bitcoins to be transfered from each input address. The same applies to the output addresses, indicating the

[2] Bitcoin uses ECDSA with the curve secp256k1 implying private keys of 256 bit length.

[3] Notice that public key, address or bitcoin account are referring to the same concept.

Inputs

Previous output (index)	Amount	From address	Type	ScriptSig
e631567f352f...:1	3.02887912	1CGVyAgAx9gg1va5pGNVJtF6gdKpPUVTSf	Address	304402201700305a3d79a[....]2b985b15daa0ab9c50cd61449ca037dc9f0
c284ec14325f....:0	3.04042789	1GY84QPLfM9d4KqTjTbbHsb9BX9FF1kYQx	Address	3045022100e724004f2d3[....]91d95b56ad29f817f3e3259daffbd72f2a98
0fbec1d29b8e....:0	2.99934316	1CGVyAgAx9gg1va5pGNVJtF6gdKpPUVTSf	Address	304402200f6e9b4281cb0[....]2b985b15daa0ab9c50cd61449ca037dc9f0
232715b3c51a....:1	3.00515088	17ALqzZFPbSqXz9aQhzgK6ts9htZfV8Mwu	Address	3044022073114957478c1d[....]8d4656bf7613d47dd4e6a5b062d9fb6a34

Outputs

Index	Amount	To address	Type	ScriptPubKey
0	0.51682435	1LUHXNTsHPUGVJJeefPdb2rpdxtWoHrcKv	Address	OP_DUP OP_HASH160 d5936a017660c48be2adaa9a77153eccfdb8b0b8 OP_EQUALVERIFY OP_CHECKSIG
1	11.5569767	1HzAb4E1kZH4pDKoxML4KXBLPPyUootw4s	Address	OP_DUP OP_HASH160 ba51b9aee7595c72a2cbc1d4e3e90e356f77804 OP_EQUALVERIFY OP_CHECKSIG

Fig. 1. Bitcoin transaction example: four input addresses and two output addresses (data from blockexplorer.com).

total amount of bitcoins that would be transfered at each account. For consistency, the total amount of the input addresses (source of the money) must be greater or equal than the total amount of the output addresses (destination of the money)[4]. Furthermore, the bitcoin protocol forces that input addresses must spend the exact amount of a previous received transaction[5] and for that reason, in a transaction, each input address can unambiguously indicate the index[6] of the transaction in which the bitcoins were received (the field *Previous output (index)* in Figure 1).

Finally, the owner of the input addresses should perform a digital signature using his private keys, proving that he is the real owner of such accounts[7].

Before accepting a payment from a standard transaction, the receiver should:

- Validate that the bitcoins of the input addresses are not previously spent.
- Validate that the digital signature is correct.

The first validation prevents doublespending in the bitcoin system and to allow such validation the system needs a ledger where all previous transactions are annotated. Before accepting the payment, the receiver needs to be sure that there is no any other transaction already in the ledger that has an input address with the same *Previous output (Index)* of the input addresses of the transaction that has to be validated. For that reason, the integrity of the system is based on the

[4] Although apparently both amounts should be the same, we will discuss later on in which situation the input value could be greater than the output value.

[5] Notice that in Figure 1, there is two input addresses that are exactly the same which indicates that bitcoins have arrived in this bitcoin account in two separate transactions.

[6] A transaction is identified in the bitcoin system by its hash value.

[7] Although this is the standard form of bitcoin verification for regular bitcoin transfer transactions, the verification of a transaction can be much more complex and is based on a bitcoin transaction script language, a stack-based execution language (more details can be found in Chapter 5 of [4]).

fact that this ledger is not modifiable, although it should be possible to add new transactions. In the bitcoin system, this append-only ledger is called blockchain[8]. The second validation can be performed with the information included in the transaction itself together with the information of the transaction identified in the *Previous output (Index)*. Finally, it is worth to mention that the enforcement of spending the total amount of a previous transaction makes very difficult to perform exact payments in the bitcoin system (transactions with exactly a single input address and a single output address), and then users should collect the "change" of the payment in one of his addresses, as it is shown in Figure 2. The address that collects the change in a transaction is referred as a *shadow address* and it belongs to the same user that performs the payment.

Inputs

Previous output (index)	Amount	From address	Type	ScriptSig
073a12d29e11...:0	0.706	1NYB35emL1yQunpExWhRM6CHBAzbJVx9Sd	Address	304402205d2b1[....]0a9b96e22abb02da6e3a03c1aa8c

Outputs

Index	Amount	To address	Type	ScriptPubKey
0	0.4	13osnkmwyYaER5tBPp5f9zWjWhpHwNgD66	Address	OP_DUP OP_HASH160 1ecdc8400fe436056bc1b18f9927ee1a7ce46443 OP_EQUALVERIFY OP_CHECKSIG
1	0.3059	1ATkLdK5icinT2c5F2NWoJYs8QWs4y5NUg	Address	OP_DUP OP_HASH160 67c81fc63d214d19696f25d1fd1fe360dabdf371 OP_EQUALVERIFY OP_CHECKSIG<

Fig. 2. A Bitcoin transaction where the owner of the address 1NYB35emL1yQunpExWhRM6CHBAzbJVx9S performs a payment of 0.4 bitcoins to the address 13osnkmwyYaER5tBPp5f9zWjWhpHwNgD66 and collects the change in the address 1ATkLdK5icinT2c5F2NWoJYs8QWs4y5NUg, the shadow address of this transaction (data from blockexplorer.com).

2.2 The blockchain and the mining process

The **blockchain** is a general append-only ledger containing all bitcoin transactions performed since the system started to operate, back in 2009. Such approach implies that the size of the blockchain is constantly increasing (21 GB by September 2014) and, for that reason, scalability is probably the biggest challenge that the system faces. The blockchain is freely replicated and stored in different nodes of the bitcoin network, making the bitcoin a completely distributed system.

Transactions are included in the blockchain at time intervals, rather than in a flow fashion, and such addition is performed by collecting all new transactions of the system, compiling them together in a data structure, called blocks, and

[8] Note that the non-modifiable property of the blockchain imply that bitcoin payments are non reversible.

including the block at the top of the blockchain. Every time that a block containing a specific transaction is included in the blockchain such transaction is said to be a **confirmed transaction** since it has been already included in the blockchain and can be checked for doublespending prevention.

Blocks are data structures that mainly contain a set of transactions that have been performed in the system (see Figure 3). To achieve the append-only property, addition of a block in the blockchain is a hard problem, so adding blocks to the blockchain is time and work consuming. Furthermore, every block is indexed using its hash value and every new block contains the hash value of the previous one (see the field *Previous block* in Figure 3). Such mechanism ensures that the modification of a block from the middle of the chain would imply to modify all remaining blocks of the chain from that point to the top in order to match all hash values.

Block 125552

Hash: 00000000000000001e8d6829a8a21adc5d38d0a473b144b6765798e61f98bd1d
Previous block: 00000000000008a3a41b85b8b29ad444def299fee21793cd8b9e567eab02cd81
Time: 2011-05-21 17:26:31
Difficulty: 244 112.487774
Transactions: 4
Total BTC: 84.52
Size: 1.496 kilobytes
Merkle root: 2b12fcf1b09288fcaff797d71e950e71ae42b91e8bdb2304758dfcffc2b620e3
Nonce: 2504433986

Transactions

Transaction	Fee	Size (kB)	From (amount)	To (amount)
51d37bdd87...	0	0.135	Generation: 50 + 0.01 total fees	15nNvBTUdMaiZ6d3GWCeXFu2MagXL3XM1q: 50.01
60c25dda8d...	0	0.259	1HuppjXz7dPrt2a67LqacDW5T4VanFrpqC: 29.5	1B8vkT58i8KUPVJvvyQfrbc8Wjwu3vEarQ: 0.5 1BQbxzgRSLEsmv1JNc8MG76wdUgMwbsaww: 29
01f314cdd8...	0.01	0.617	1NdzSE6sHubscXJrv7jJn2gd4fL9L3ai6E: 0.03 1Jjv9m5VrRUE7VoktCsj18KUSqkqchhbum: 0.02 1HsYJJPqTn34DEjMnTb3VfKckX7ZcWPibm: 4.82	175FNxcLc1YrTwwG6TcsywcsHYdVqyhbwC: 0.01 1MueNMRJmcqVQeqE7v4dqogpNbhyxqq8R6: 4.85
b519286a10...	0	0.404	12DCoCVvDCkQShZ5RTh9bysgCkmkRMNQbT: 0.14 13CJwnnXJPwkzY4Xnaoqf8dnyNBwrHG9fe: 0.01	1Mos7p8fqJKBcYNRG1TdT5hBRxdMP6YHPy: 0.15

Fig. 3. Example of a bitcoin block (data from blockexplorer.com).

Adding a block to the blockchain is known as the **mining process**, a process that is also distributed and that can be performed by any user of the bitcoin network using specific-purpose software (and hardware). The mining process uses a hashcash proof-of-work system, first proposed by Adam Back as an anti-spam mechanism. The proof-of-work consists in finding a hash of the new block with a value lower than a predefined target[9]. This process is performed by brute force

[9] Notice that the value of the target determines the difficulty of the mining process. Bitcoin system adjusts the target value depending on the hash power of the miners in order to set the throughput of new blocks to 1 every 10 minutes (in mean).

varying the nonce value of the block and hashing the block until the desired value is obtained. Once the value has been found, the new block becomes the top block of the blockchain and all miners discard their work on that block and move to the next one, by collecting new transactions and taking the hash of the top block as the previous block hash.

Mining new blocks is a structural task in the bitcoin system since it helps to confirm the transactions of the system. For that reason, and also assuming that mining implies a hard work, miners have to be properly rewarded. In the bitcoin system, miners are rewarded with two mechanisms. The first one provides them with newly created bitcoins. Every new block includes a special transaction, called generation transaction, (see the first transaction in Figure 3) in which it does not appear any input address and the output address is determined by the miner who creates the block, who obviously indicates one of its own addresses[10]. The second rewarding mechanism is the fees that each transaction pays to the miner. The fee for each transaction is calculated by computing the difference between the total input amount and the total output amount of the transaction (notice that in example block of Figure 3 the first transaction does not provide any fee while the second one generates a 0.01 fee). All fees collected from transactions in a block are included in the generation transaction.

2.3 The bitcoin network

The bitcoin system needs to disseminate different kinds of information, essentially, transactions and blocks. Since both data are generated in a distributed way, the system transmits such information over the Internet through a distributed peer to peer (P2P) network. Such distributed network is created by bitcoin users in a dynamic way, and nodes of the bitcoin P2P network [5] are computers running the software of the bitcoin network node. This software is included by default into bitcoin's full-client wallets, but it is not usually incorporated in light wallet versions, such as those running in mobile devices. It is important to stress such distinction in case to perform network analysis, because when discovering nodes in the P2P bitcoin network, depending on the scanning techniques, not all bitcoin users are identified, but only those running a full-client and those running a special purpose bitcoin P2P node. Furthermore, online bitcoin accounts, provided by major bitcoin Internet sites, can also be considered as a light weight bitcoin clients, so they do not represent a full bitcoin P2P node neither.

3 Bitcoin Anonymity

Anonymity is probably one of the properties that has been key for the success of the currency deployment. Anonymity in the bitcoin network is based on the fact

[10] The amount of a generation transaction is not constant and it is determined by the bitcoin system. Such value, started in 50 bitcoins, is halved every four years, fixing asymptotically to 21 millions the total number of bitcoins that will be ever created.

that users can create any number of anonymous bitcoin addresses that will be used in their bitcoin transactions. This basic approach is a good starting point, but the underlaying non-anonymous Internet infrastructure, together with the availability of all bitcoin transactions in the blockchain, has proven to be an anonymity threat. In order to review the papers published on bitcoin anonymity, we group them in three different categories: those papers that exploit mainly data obtained from the blockchain to derive some information from users or more general properties like usage patterns; papers that use bitcoin network information to identify users; and papers that propose mixing techniques to protect users anonymity.

3.1 Blockchain Analysis

A direct approach to analyze the anonymity offered by the bitcoin system is to dig information out of the blockchain. Since the blockchain includes all transactions performed by the system, a simple analysis provides information from which bitcoin addresses the money comes and to which bitcoin addresses it goes. However, since users in the bitcoin system can create any number of addresses, the main goal is to cluster all addresses in the blockchain that belong to the same user. As we will see, authors apply different techniques to perform such clustering.

The first research article on Bitcoins was published by Reid and Harrigan [2], a first version of which appeared in arXiv in July 2011. From the blockchain information, authors construct the transaction network and the user network. The former represents the flow of bitcoins between transactions, where each vertex represents a transaction and each directed edge indicates whether or not there is an input/output address that links the transactions. The latter represents the flow of bitcoin users over the time. To construct the user network, authors cluster addresses of the same user assuming that all input addresses of a transaction belong to the same user. Then, external information on bitcoin addresses is obtained from different Internet resources (like twitter posts, forums, specialized bitcoin applications -like bitcoin faucet-) to help the clustering process and to identify the users behind such clusters. All such information allow them to perform egocentric analysis and visualization, context discovery, flow and temporal analyses and they conclude that it is possible to associate many bitcoin addresses with each other, and with external identifying information. Furthermore, with appropriate tools, the activity of known users can be observed in detail.

In [6], Androulaky *et alter* take another step into clustering addresses. Taking into account the same idea of [2], where all input addresses of the same transaction are clustered, they added another heuristic using the output addresses of a transaction. Assuming that most transactions have only two output addresses, in the case that one of the two has already appeared in the blockchain, the other one will be a shadow address and can be clustered with the input addresses. Furthermore, they also apply behavior-based clustering techniques, K-Means and Hierarchical Agglomerative Clustering, to enhance the cluster creation. In

order to perform such analysis, the authors generate synthetic data from a specific purpose bitcoin simulator that they developed. Data from the simulation has also the advantage to provide a ground truth for evaluating their clustering measures. With this simulation environment and the proposed techniques, authors indicate that the profiles of 40% of bitcoin users can be unveiled.

Ron and Shamir [7] perform an analysis of bitcoin user behavior from the blockchain data, rather than trying to deanonymize user information. They also use the assumption that multiple input addresses belong to the same user in order to characterize user behavior. They conclude that until May 13th 2012 most of the new created coins remain unexpended in the minted addresses and that there was a huge number of tiny transactions that move fractions of bitcoins. Furthermore, they carefully analyze the largest transactions of the network until that moment and provide a detailed graph structure of their movements.

Papers reviewed so far perform a passive analysis in the sense that information of the blockchain is processed without any previous intervention. In [8] in order to better understand the traceability of Bitcoin flows, Meiklejohn *et alter* perform an active analysis. By performing payments from owned bitcoin addresses to known services (like mining pools, on-line wallets, gambling services, exchange sites, ...) they can identify such services later on in the corresponding blockhain transactions. Furthermore, they also browse the Internet in order to obtain user identification of other addresses. Then, they used two heuristics for clustering: the first one is the all input addresses belong to the same user (already used in [2, 6, 7]) and the second one identifies the shadow address of a transaction by looking the one between all output addresses that appeared for the first time in the blockchain (a similar approach than in [6], but not limited to two output address transactions). With their analysis, authors conclude that for large bitcoin transactions, it is possible to trace their movements and the bitcoin network does not offer enough anonymity, for instance for money laundry. Such traceability is even more sharp in case the analyzer is (or has access) to a central service, like a mining pool, an eWallet provider or a bitcoin exchange site.

In [9] Ober *et alter* empirically study global properties of the bitcoin transaction graph and their time evolution since bitcoin creation until January 6th, 2013. They distinguish from all bitcoin addresses what they call *used addresses*, those that have been used to perform a payment (that is an address present as an input address in some transaction). They also define an *active entity* as the owner of such addresses and, similar to other authors, cluster in a single *active entity* different used addresses that appear together as input in a transaction. The size of an entity is then the number of addresses included in the cluster. Authors deal the anonymity in the bitcoin network through the measure of k-anonymity. They conclude that to estimate the level of k-anonymity provided by bitcoin system is necessary to estimate the number of active entities since, for instance dormant coins (those included in an address not active for a long time) reduce the anonymity set. Furthermore, they also indicate that to better estimate the k-anonymity at a certain point of time, active entities should be defined based on a window time around this period (hours, days, weeks, ...).

Then, an active entity is the one that have performed a payment within this window time. With their analysis, they conclude that the best strategy, which maximizes the anonymity set, is to be as small as possible (the best case, only one address for each cluster) and be active for the shortest possible time (the best case, single use address). Their analysis provides interesting facts, like for instance, that speculation is good for anonymity since it raises the bitcoin price and, so, the total number of active entities which in turns increase the anonymity set.

Spagnuolo *et alter* [10] present BitIodine, a tool to analyze the blockchain information. BitIodine parsers the blockchain information and provides a frontend to obtain different information. From basic address account balance, received or sent import amount or total number of transactions to more sophisticate information like address clustering (using multi-input addresses and shadow addresses), addresses labeling based on public information on the web or path computation between addresses. As a use case of the proposed tool, authors provide an interesting analysis on payments to CryptoLocker ransomware. They found a high correlation between the dates that infections were reported and the dates of payments performed to bitcoin addresses provided by the ransomware for unlocking the files. This is the first analysis performed with public available information (not backed up with information graved from underground forums) in which it is possible to estimate the amount of money generated by a ransomware software.

In [11], Ron and Shamir present an in deep analysis of bitcoin transactions performed by Dread Pirate Roberts, the person who ran first Silk Road marketplace. Based on the blockchain information and a published account, authors made a detailed analysis which provides enough information to show the power of data mining techniques to analyze specific transactions in a public ledger system like bitcoin.

As we have seen, clustering addresses of the bitcoin system belonging to the same user is the key research topic in blockchain analysis. Although some proposals have been performed so far, the dynamism of the bitcoin system still offers room for further analysis. For instance, some hypothesis on the heuristics to cluster such addresses depend on the behavior of the wallets and how are they programed to perform and receive payments. Such behavior has been modified since the publication of some papers and new functionalities of the system have emerged. For those reasons, as we will see later on, some of those hypotheses may not hold at present time and new heuristics should be analyzed. Furthermore, most research works perform address clustering with the help of external data (like forums post, tweets, etc.), then to cluster addresses with only the information provided by the blochchain is still an open challenge.

3.2 Traffic analysis

As we already mentioned, the anonymity degree of users in the bitcoin system is also bounded by the underlying technologies used. Transactions in the bitcoin system are transmitted through a P2P network, so, as it was first pointed out in

[2], the TCP/IP information obtained from that network can be used to reduce the anonymity of the system. Although it is true that most wallets are able to work over anonymous networks (TOR[11] or I2P[12]) a high number of bitcoin users do not use such services, and then, there is still room for network analysis.

Koshy *et alter* [12] perform an anonymity study based on real-time transaction traffic collected during 5 month. For that purpose, authors develop CoinSeer, a bitcoin client designed exclusively for data collection. For more than 5 million transactions, they collected information on the IP address from where the CoinSeer received such transaction and, in the general case, they assigned as the IP corresponding to the transaction the one that broadcast the transaction for the first time. In order to perform a pure network analysis, authors do not apply any address clustering process, so only single input transactions (almost four million) are taken into account in the analyzed data set. Then, to match an IP with a bitcoin address, they consider a vote on the link between IP_i and $address_j$ if a transaction first broadcasted form an IP_i contains the bitcoin $address_j$ as input address. Authors also perform a similar analysis for output addresses and model the problem as an evaluation of association rules, identifying the corresponding confidence scores and the support counts for the rule. After their analysis, authors conclude that it is difficult to map IP addresses with bitcoin addresses by performing traffic analysis if bitcoin peers act properly, since the bindings authors could obtain between IP addresses and bitcoin addresses mainly come from anomalous transactions patterns. Furthermore, authors also indicate that some network configuration, like mixing services or eWallets, might conduct to erroneous assumptions when linking IP and bitcoin addresses.

In contrast to blockchain analysis, traffic analysis has received less attention from the researches probably due to the fact that the blockchain is ready available for analysis and network data has to be gathered. In fact, bitcoin network analysis is a hard topic due to the dynamism and size of such P2P network. The anonymity analysis performed by Koshy *et alter* seems to show that no information can be derived with this technique, but it is difficult to completely discard such approach since in their work authors do not provide any estimation regarding which part of the bitcoin P2P network represent the 2,678 peers they were able to monitor, and for the period of the analysis, no data of the size of the network is available from other sources. So, with only one work performed, whether or not network analysis can reveal private information from bitcoin users still remains an open problem. Furthermore, network analysis can be performed to identify not only the owner of an address but also the identity of other actors in the bitcoin community.

3.3 Mixing

In order to enhance the anonymity properties of the bitcoin system, some authors propose the use of mix services, a procedure that shuffles the information in order

[11] `https://www.torproject.org/`
[12] `https://geti2p.net/`

to hinder the relation between then input and the output values.[13] The goal is to allow bitcoin users to send bitcoins from one address to a mix service and receive from the mix service the bitcoins to another address that could not be linked with the original one. This service can be run by a central authority which receives payments and pays back to different addresses. However, such authority should be a trusted party since, on one hand, it is able to link addresses and, on the other hand, regarding the non-reversibility of the bitcoin payments, the mixer can receive the payment without sending back the bitcoins.

A basic mix service can be implemented using a multiple-input and multiple-output transaction, as it is described in CoinJoin [14]. The idea is that multiple users can jointly create a transaction with multiple input addresses[14] and multiple output addresses. To be a valid transaction, the transaction should be signed by all users participating in the mixing. Notice that partially signed transaction should circulate between users that mix their coins, although a meeting point server can be used. In that case, users should use an anonymous channel (TOR/I2P) to protect them in front of network attacks performed by the meeting point server. Furthermore, blind signatures may enforce that the meeting server does not learn linkability information between input and output address transactions. One of the problems of this proposal is that one of the anonymous users of the mix service can perform a DoS attack. Since the final valid transactions should be signed by all users that include bitcoins in the transactions, each mixing transaction never becomes valid in case the attacker simply does not sign any transaction in which he takes part. Despite these drawbacks, CoinJoin has been implemented in SharedCoin[15] or DarkWallet[16].

Möser *et alter* [15] perform an active analysis using reverse-engineering to understand the mode of operation of three mixing services: Bitcoin Fog[17], BitLaundry[18] and SharedCoin[19]. They perform mix procedures for each mix service for small bitcoin values using as a destination addresses one or multiple new generated ones. Then, they visualize the transaction graph of the addresses involved in the mixing. They conclude that while in Bitcoin Fog and SharedCoin it is hard to relate input and output transactions, for the BitLaundry, deirect connections of the transaction graphs can be found and it cannot be considered as a reliable anonymizer.

Barber *et alter* propose in [16] a Fair Exchange Protocol that can be used as a two-party mixing protocol. The protocol uses the scripting functionality

[13] The main application of the mix concept, proposed by D. Chaum in [13] is the TOR network.

[14] At that point, it is important to note that some bitcoin uses, like the one described by CoinJoin, break the assumption that multiple input addresses in a transaction implies the same owner for all those input addresses, assumption that is taken as an heuristic for clustering addresses by almost all the anonymity papers.

[15] `https://sharedcoin.com/`

[16] `https://www.darkwallet.is/`

[17] `http://bitcoinfog.com/`

[18] `http://app.bitlaundry.com/`

[19] `https://sharedcoin.com/`

that bitcoin transactions provide and a cut-and-choose protocol. The paper only provides the description of each protocol phase as an isolated two party protocol assuming that both users have already been meet.

In [17], Bonneau *et alter* present Mixcoin, a centralized mixing system that relies on accountability. Users of the system obtain, prior the mixing phase, a signed warranty that can be used to prove, in case of the event, that the mixer entity has misbehaved. Authors point out that such public verifiable proof of misbehavior would discourage malicious mixing. However, there is still the possibility that the mixer could deanonymize users using his stored information. This thread is solved by concatenating several mixer services, thus reducing the strategy of a malicious mixer to a collusion with the other mixers. However, the mixer concatenation is not straight forward, since the proposed mixing protocol does not allow the users to choose the number of bitcoins to mix, because the scheme fixes a predefined amount. For that reason, mixing fees (that can be seen as the difference between the incoming and the outcomming bitcoin values) are difficult to apply without affecting the anonymity of users. To solve that point, authors propose randomized mixing fees so the fee is not a fraction of the mixed value, but the entire value that the user wants to mix, and the fee can be charged, or not, by the mixer with some predefined probability. Using such approach, the input addresses of the mix has the same value than output addresses or, in case the fee has been applied, there is no output address. This approach allows sequential mixing, but imposes a restriction on the fixed amount to be mixed and the minimum number of coins that users can mix, in order to keep a reasonable fee for the service.

Finally, Bissias *et alter* propose in [18] a system called Xim, a two-party mixing protocol designed as a multi-round protocol to enhance its anonymity properties. In fact, the core proposal is an anonymous partnering system that allows to find anonymously partners. Then, the mixing is performed using the Fair Exchange protocol proposed in [16]. They perform a comparative analysis of Xim with other proposals (MixCoin, SharedCoin, DarkWallet, CoinShuffle) and analyze different attacks on Xim. Sybil attacks and DoS attacks are discouraged by means of a carefully designed fee system.

Mix services provide a mechanism to mix bitcoin from different users in order to increase bitcoin user's anonymity. Proposals have moved from centralized systems to distributed protocols in order to increase user's privacy protection. Research in this topic ranges from atomic protocols that do not take into account the entire practical scenario (like how users can be paired or grouped anonymously) to specific proposals on how mixing fees can be calculated. In this field, open challenges include side channels attacks (within the mixing service and at communication level) and the integration of multiple mix services that, in its extreme case, yields the interesting concept of continual mixing, already proposed in [17]. Finally, It is also worth to mention that some proposals that initially were focused on improving bitcoin anonymity, implied a deep modifica-

tion of the bitcoin protocol and, due such impossibility, some of those proposals have evolved in the creation of different currency proposals, like Zerocoin [19].

4 Conclusions

Bitcoin is a payment system based on a decentralized architecture that provides a mechanism to obtain multiple anonymous credentials, bitcoin addresses, that can be used to perform and receive payments. However, research performed so far has proven that the way the system uses such addresses may unveil some information from their owners. Since all transactions performed by the system are freely available in the blockchain for analysis, it allows to cluster different addresses of the same user and characterize some uses. Furthermore, if one of the addresses of the cluster can be mapped to a real identity, then the payment history of the entire cluster may disclose relevant information of that user. Although interesting research has been performed in this topic, the dynamism of the bitcoin ecosystem that constantly modifies and enhances the bitcoin usage implies that some of the hypotheses assumed for those blockchain analysis may not completely hold and, for that reason, blockchain analysis still presents interesting open questions.

Apart form the blockchain analysis, anonymity of the bitcoin system can be analyzed by gathering information from the P2P network used for payment communication. Since the P2P network uses the TCP/IP protocol, traffic analysis may reveal private information from users. However, such analysis is much more difficult to perform than the blockchain analysis since the bitcoin P2P network is highly dynamic. Although very few papers have been presented regarding this topic and results are not apparently optimistic, we think that there is still interesting network analysis that can be performed over the bitcoin P2P network.

In order to mitigate the anonymity reduction of the bitcoin system that can be performed using the techniques described above, the use of mix services have been proposed. Bitcoin mixes are services that allow a user to anonymize his bitcoins by mixing them with bitcoins of other users. Different proposals have been presented in this field showing that it is possible to design a mix service with a considerable level of security for the user. However, it is important to indicate that research in bitcoin mix services has to be performed carefully since developing this kind of services can be considered, from an economical or legal point of view, money laundering.

Finally, it is worth mention that research in the bitcoin ecosystem can be performed in other topics than anonymity, like for instance cryptography, network security or P2P network to name a few. On the other hand, besides the research lines that can be performed directly on the study of the bitcoin system itself, other approaches perform research using the bitcoin system as a tool. Examples of such approach are the design of secure multiparty computation or coin toss protocols. Furthermore, some structural parts of the bitcoin system, like the blochchain approach as an append-only ledger, may open interesting challenges for future developments on secure decentralized systems.

Acknowledgments

This work was partially supported by the Spanish Ministerio de Ciencia y Tecnologia (MCYT) funds under grants TIN2010-15764 "N-KHRONOUS" and TIN2011-27076-C03 "CO-PRIVACY".

References

1. Nakamoto, S.: Bitcoin: A Peer-to-Peer Electronic Cash System. (2008)
2. Reid, F., Harrigan, M.: An analysis of anonymity in the bitcoin system. In Altshuler, Y., Elovici, Y., Cremers, A.B., Aharony, N., Pentland, A., eds.: Security and Privacy in Social Networks. Springer New York (2013) 197–223
3. Babaioff, M., Dobzinski, S., Oren, S., Zohar, A.: On bitcoin and red balloons. In: Proceedings of the 13th Association for Computing Machinery (ACM) Conference on Electronic Commerce. EC '12, New York, NY, USA, ACM (2012) 56–73
4. Antonopoulos, A.M.: Mastering Bitcoins. O'Reilly Media (December 2014)
5. Donet, J.A., Pérez-Sola, C., Herrera-Joancomartí, J.: The bitcoin P2P network. In Böhme, R., Brenner, M., Moore, T., Smith, M., eds.: Financial Cryptography and Data Security. Volume 8438 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2014) 87–102
6. Androulaki, E., Karame, G., Roeschlin, M., Scherer, T., Capkun, S.: Evaluating user privacy in bitcoin. In Sadeghi, A.R., ed.: Financial Cryptography and Data Security. Volume 7859 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2013) 34–51
7. Ron, D., Shamir, A.: Quantitative analysis of the full bitcoin transaction graph. In Sadeghi, A.R., ed.: Financial Cryptography and Data Security. Volume 7859 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2013) 6–24
8. Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., McCoy, D., Voelker, G.M., Savage, S.: A fistful of bitcoins: Characterizing payments among men with no names. In: Proceedings of the 2013 Conference on Internet Measurement Conference. IMC '13, New York, NY, USA, ACM (2013) 127–140
9. Ober, M., Katzenbeisser, S., Hamacher, K.: Structure and anonymity of the bitcoin transaction graph. Future Internet **5**(2) (2013) 237–250
10. Spagnuolo, M., Maggi, F., Zanero, S.: Bitiodine: Extracting intelligence from the bitcoin network. In Christin, N., Safavi-Naini, R., eds.: Financial Cryptography and Data Security. Volume 8437 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2014) 457–468
11. Ron, D., Shamir, A.: How did dread pirate roberts acquire and protect his bitcoin wealth? In Böhme, R., Brenner, M., Moore, T., Smith, M., eds.: Financial Cryptography and Data Security. Volume 8438 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2014) 3–15
12. Koshy, P., Koshy, D., McDaniel, P.: An analysis of anonymity in bitcoin using p2p network traffic. In Christin, N., Safavi-Naini, R., eds.: Financial Cryptography and Data Security. Volume 8437 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2014) 469–485
13. Chaum, D.L.: Untraceable electronic mail, return addresses, and digital pseudonyms. Commun. ACM **24**(2) (February 1981) 84–90
14. Maxwell, G.: Coinjoin: Bitcoin privacy for the real world. post on bitcoin forum https://bitcointalk.org/index.php?topic=279249.

15. Moser, M., Bohme, R., Breuker, D.: An inquiry into money laundering tools in the bitcoin ecosystem. In: eCrime Researchers Summit (eCRS), 2013. (Sept 2013) 1–14
16. Barber, S., Boyen, X., Shi, E., Uzun, E.: Bitter to better - how to make bitcoin a better currency. In Keromytis, A., ed.: Financial Cryptography and Data Security. Volume 7397 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2012) 399–414
17. Bonneau, J., Narayanan, A., Miller, A., Clark, J., Kroll, J.A., Felten, E.W.: Mixcoin: Anonymity for bitcoin with accountable mixes. In Christin, N., Safavi-Naini, R., eds.: Financial Cryptography and Data Security. Volume 8437 of Lecture Notes in Computer Science. Springer International Publishing (2014) 486–504
18. Bissias, G., Ozisik, A.P., Levine, B.N., Liberatore, M.: Sybil-resistant mixing for bitcoin. In: Proceedings of the 13th ACM Workshop on Workshop on Privacy in the Electronic Society. WPES '14, New York, NY, USA, ACM (2014)
19. Miers, I., Garman, C., Green, M., Rubin, A.: Zerocoin: Anonymous distributed e-cash from bitcoin. In: Security and Privacy (SP), 2013 IEEE Symposium on. (May 2013) 397–411