Analysis of the de-anonymization of blockchains

Matteo Bonini

De Cifris Athesis Seminar —— 27 November 2019

Anonimity

What is anonimity?

$$anonymity = pseudonymity + unlinkability$$

Pseudonimity

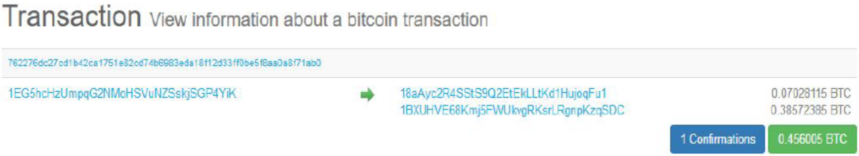
The pseudonym identifies a holder, that is, one human being who possess but do not disclose their true names.

Unlinkability

It means that as a user interacts with the system repeatedly, these different interactions should not be able to be tied to each other from the point of view of some adversary.

Bitcoin-like blockchains

Transactions



Transaction View information about a bitcoin transaction

762276dc27cd1b42ca1751e82cd74b6983eda18f12d33ff0be5f8aa0a8f71ab0

1EG5hcHzUmpqG2NMoHSVuNZSskjSGP4YiK → 18aAyc2R4SStS9Q2EtEkLLtKd1HujoqFu1 0.07028115 BTC
1BXUHVE68Kmj5FWUkvgRKsrLRgnpKzqSDC 0.38572385 BTC

Image taken from [1].

Remark

One key feature about a transaction is that it can have multiple inputs and multiple outputs.

Clustering of Public Keys

Using the publicly available transaction history, a directed acyclic graph representing the transactions can be created.

- Each node in the graph represents a transaction;
- Each directed edge contains the and value of bitcoins where an incoming edge represents the input to a transaction and the outgoing edge represents the output of a transaction.

Network analysis

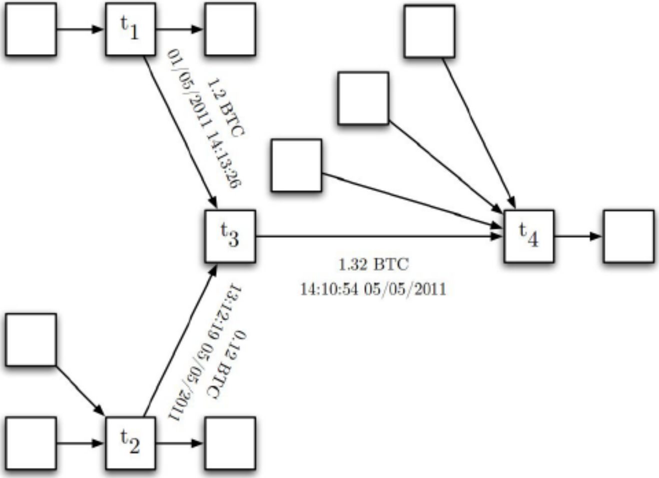


Image taken from [8].

Netowork analysis

Using the transaction graph, it is possible to create an address graph, which represents the transactions between the users.

- Each node in the graph represents the public address of a user;
- Each directed edge represents an input-output pair of a transaction where the input's public-key belongs to sender and the output's public-key belongs to the receiver.

Network analysis

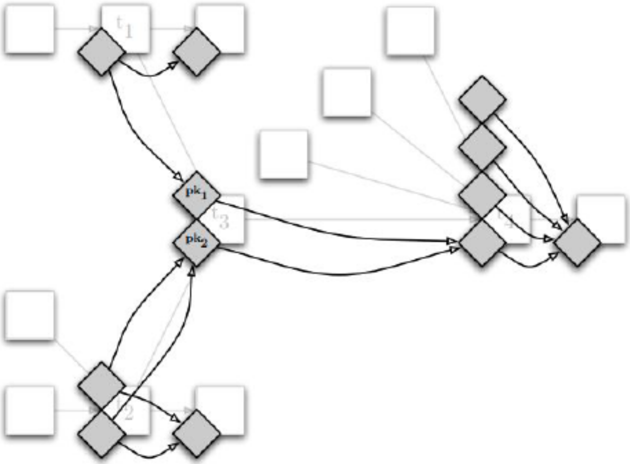


Image taken from [8].

Heuristics

With the address graph, it is then possible to use two types of heuristics to cluster subsets of public keys belonging to the same user.

- “idioms of use”, i.e. all the inputs in a transaction are generated by the same user because different users rarely contribute to a single shared transaction in the real world;
- “change address”, i.e. the excess from the input address of a transaction is sent back to an address belonging to the sender.

Using these heuristics, public keys can be successfully clustered to their respective users.

Network analysis

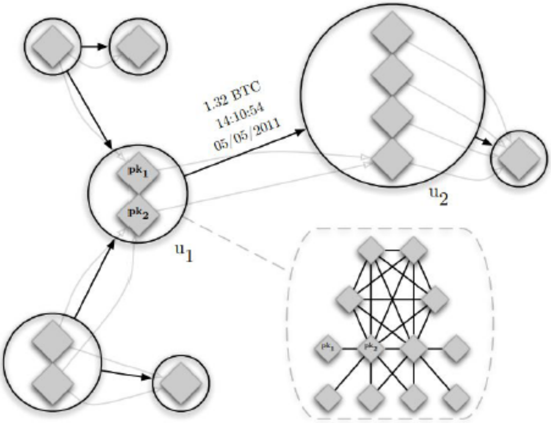


Image taken from [8].

A practical example

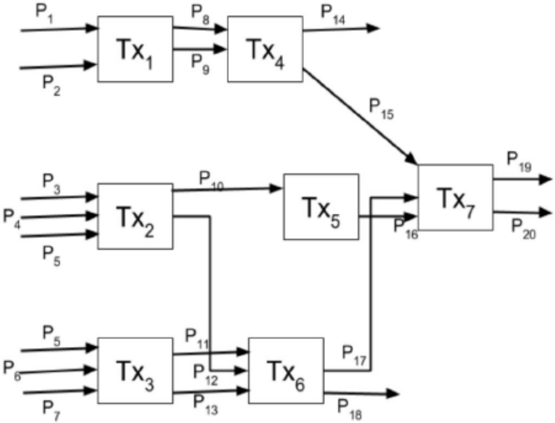


Image taken from [2].

Re-identification from the knowledge of the public key

Common ways that allow Bitcoin users to be associated to their known public-keys:

- Voluntary disclosure of information.
- Trading the cryptocurrency for fiat currency on an exchange.
- Purchasing items with the cryptocurrency.

Once the public keys are linked to real user identities, the deanonymization process is complete.

A different approach

CoinJoin

Multiple transactions are merged by a centralized, trusted mixer, such that the inputs and outputs of the set of users are part of the same transaction, therefore ensuring that each specific output cannot be linked back to a specific input.

Features

- No "idioms of use" heuristics;
- Implementable on almost every cryptocurrency.

CoinJoin

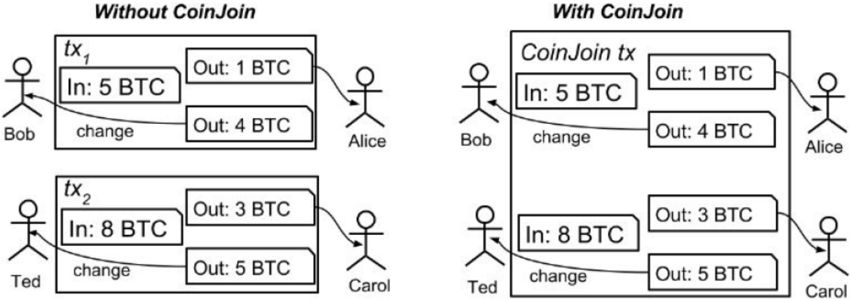


Image taken from [2].

CoinJoin

Issues

- It is very complex to permute the output addresses without revealing the permutation to users within the mixing group.
- If trusted facilitator is used, this leads to accountability issues.

In any case CoinJoin is still an improvement over the Bitcoin protocol because it is resistant to network analysis.

Accountability

CoinJoin guarantees external unlinkability but not internal unlinkability, since the central mixing server learns the relation between input and output addresses and needs to be trusted to ensure anonymity.

Can the mix be trusted?

- Malicious mixes;
- A mix which is malicious, hacked, or subpoenaed might leak its records and undermine user anonymity;
- Mixes badly designed.

Proposals

In order to solve these problems some solutions were proposed

- Verifiable mixes, they provide accountability by enforcing that all mixes issue a proof that their output is a permutation of their input;
- Reputable mixes (Mixcoin), each mix has to prove that each output corresponds to some input, as opposed to the mix itself originating the message;
- Blind signatures (BlindCoin), which extends the Mixcoin protocol by using blind signatures to conceal cryptographically the mapping between the user input and outputs, at the cost of requiring two extra transactions;
- Decentralized mix (Coinshuffle), coordinates CoinJoin transactions using a cryptographic decentralized protocol that allows users to mix their coins with those of other interested users.

Side-channel attacks

There are other privacy attacks that are possible even if the protocol isn't vulnerable:

- Timing information (Intersection Attack);
- Precise payment sizes (Packet Counting Attacks);
- IP Address information (Network Layer attack).

Mixnets

- Introduced by Chaum in 1981 for anonymous communication;
- Not relying on any particular mixer;
- Work by chaining multiple mixes together, that take in and shuffles messages from a group of senders, and sends these messages out in a random sequence to the next mix node, until the messages eventually reach their final destinations.
- The messages have a layer of public key cryptography;
- The main weakness is the size of the anonimity set.

Mixnets

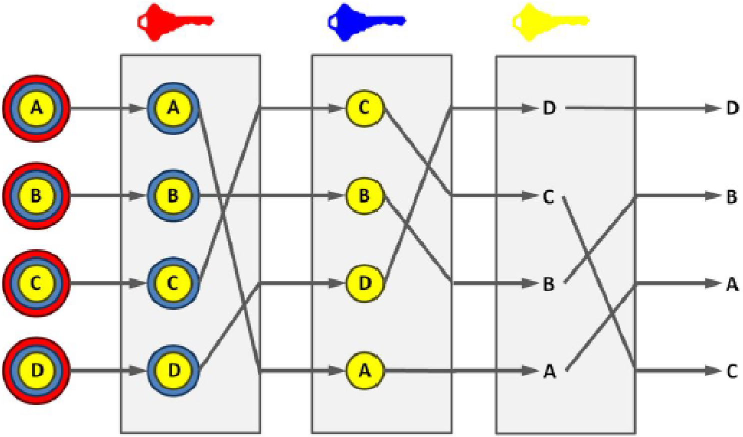


Image taken from [4].

Monero-like blockchains

Features

- **Ring Confidential Transactions**: anonymize the transaction amount. They can protect against blockchain analysis using the "change address" heuristic and Packet Counting Attack.
- **Ring Signatures**: include both the real sender's public key as well as several other users public keys as a possible source of the funds being sent.
- **Stealth Addresses**: compose of two public keys owned by the recipient, which the sender will use to produce new one-time addresses to send the coins to.

Weaknesses

Chain-reaction analysis

Firstly, they showed that 62% of transaction inputs with one or more mixins are vulnerable to “chain-reaction” analysis, because the real input could be deduced by elimination of inputs that are already spent by 0-mixin transactions.

Recent research showed that 64.04% of all Monero transaction outputs prior to February 2017 were indeed such 0-mixin transactions.

Heuristics

Spending behaviour

The spend-time distribution of Monero is highly right skewed, and users tend to spend coins soon after receiving them. Moreover, the Monero client's sampling mechanism samples from a distribution that does not represent the real spending behavior.

Heuristic

The real input is the "newest" input 92.33% of the time using a simulation.

Zcash blockchain

Zero-knowledge

Main idea

Zero-knowledge proofs permit users to convert bitcoins to other types of cryptocurrencies and spend these new coins using anonymous proof of ownership instead of explicit public-key based digital signatures, thus effectively shielding the transaction history of a coin.

Transactions

- z-transactions, private.
- t-transactions, public.

Zcash

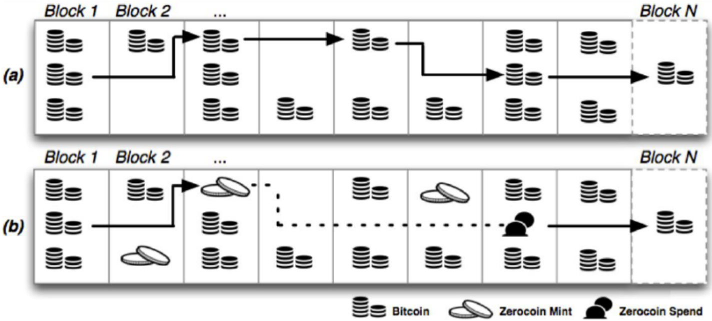


Image taken from [10].

Zcash

Heuristic

The sender moves the currency with a non-shielded transaction from the transparent address to the shielded address, then pays through a shielded transaction. Just after receiving the payment, the receiver moves the currency from the shielded address to the non-shielded address.

Results

A study found that 31.9% of coins being shielded conformed to this pattern and out of these traceable coins.

Comparisons

Protocol	**Anonimity**	**Weakness**
Bitcoin	Pseudonymous	Network analysis
CoinJoin	Pseudonymous, unlinkable	Side Channels
Mix Nets	Pseudonymous, unlinkable	Size of the anonymity set
Monero	Pseudonymous, unlinkable	Size of the anonymity set.
Zcash	Pseudonymous, unlinkable	Side-channel.

Bibliography

[1] https://blockchain.info/.

[2] M. Conti, S. K. E, C. Lal, S. Ruj, "A survey on security and privacy issues of bitcoin." IEEE Communications Surveys & Tutorials 20.4 (2018): 3416-3452.

[3] A. Kumar, C. Fischer, S. Tople, and P. Saxena, “A traceability analysis of monero’s blockchain,” in European Symposium on Research in Computer Security, 2017, pp. 153–173.

[4] https://en.wikipedia.org/wiki/Mix_network

[5] M. Möser et al., “An Empirical Analysis of Traceability in the Monero Blockchain,” Proceedings on Privacy Enhancing Technologies, vol. 1, p. 21.

[6] S. Noether and A. Mackenzie, “A note on chain reactions in traceability in cryptonote 2.0,” Research Bulletin MRL-0001. Monero Research Lab, 2014.

[7] J. Quesnelle, “On the linkability of Zcash transactions” arXiv:1712.01210.

[8] H. Reid,M. Harrigan. "An analysis of anonymity in the bitcoin system." Security and privacy in social networks. Springer, New York, NY, 2013. 197-223.

[9] W. Wu, B. Falk. "Limitations of Privacy Guarantees in Cryptocurrency." (2018).

[10] http://zerocoin.org/.

THANK YOU
FOR THE ATTENTION