POLITECNICO DI TORINO

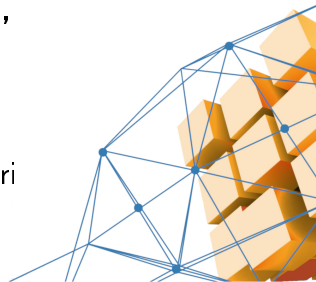
DISMA Dipartimento di Scienze Matematiche G. L. Lagrange

ECCELLENZA 2018 • 2022

CrypTO Conference 2021

An Overview of Blockchains' De-anonymization Attacks

Andrea Gangemi
CrypTO - Gruppo di crittografia e teoria dei numeri
Politecnico di Torino e Università di Torino
28 May 2021

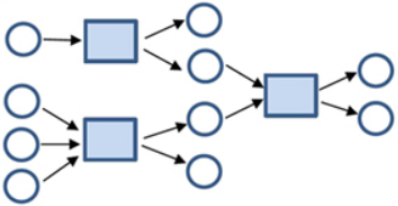


Introduction

- Blockchains were first described in 2008, thanks to the well-known Satoshi Nakamoto's paper that introduced **Bitcoin**.
- Blockchains have a lot of interesting properties: among them, we have a supposed high level of **pseudoanonymity**.
- Theoretically, *a new key pair should be used for each transaction to keep them from being linked to a common owner*.
- We can distinguish between two different record-keeping models: the **UTXO** model and the **Account** model.

The UTXO Model

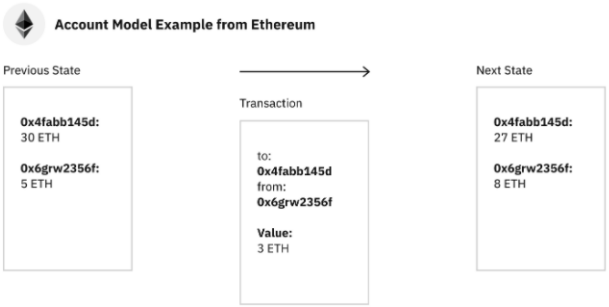
- In the UTXO model, transaction inputs are UTXOs from previous transactions recorded on the blockchain.



- Bitcoin, Monero and ZCash use this model.

The Account Model

- The account model keeps track of the balance of each account as a global state.



- Ethereum uses this model.

The De-anonymization Problem

- There are mainly three types of strategies an attacker can use to de-anonymize blockchain data:
  - Study of the peer-to-peer network to link the user IP with its addresses.
  - Use of Machine Learning (ML) algorithms to deduce information from transaction data.
  - Analysis of the blockchain transaction graph.

De-anonymization of Bitcoin

Let's start our analysis from Bitcoin.

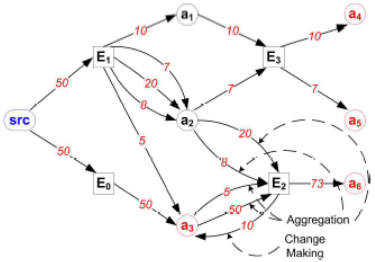
- A first, classic, de-anonymization technique is the **clustering** of addresses. There are mainly two strategies to form clusters:
  - All the inputs of a transaction belong to the same user.
  - If in a transaction there are exactly two outputs and one of them has appeared before on the blockchain, while the other one is fresh, then the latter is the change address of the user who is sending the transaction.
- *ChainAnalysis* is a company which clustered, identified, and categorized a huge number of Bitcoin addresses.
- We call **entity** the owner of a set of addresses.

De-anonymization of Bitcoin

- A good percentage of these clusters can be labelled thanks to off-chain data, but some of them remain unlabelled without further analysis.
- **Supervised Learning algorithms** can be used to label the clusters which are still not categorized.
- Starting from the known clusters, supervised algorithms can be trained, and their performance is compared with different metrics like the *F1-score*.
- The best algorithms were **random forests** and **gradient boosting classifier**.

De-anonymization of Bitcoin

- A different and more modern approach represents the blockchain network like a weighted bipartite graph.



- In this graph, square nodes represent transactions, while circle nodes represent entities.

De-anonymization of Bitcoin

- The key idea takes advantage of the concept of **motif**, a small subgraph with statistical significance.
- To characterize entities, researchers used different **graph neighborhood features**, like temporal features, centrality features and motif features.
- The algorithm used for the classification is the **decision tree**.

- The main limitation of this approach is it does not scale: *the graph is considered as a static model and most of the features are hand-engineered*.
- Bitcoin users should start using privacy-enhancing wallets like *Wasabi*.

De-anonymization of ZCash

A second interesting blockchain which uses the UTXO model is **ZCash**.

- ZCash's first block was mined in 2016. This blockchain is a fork of Bitcoin whose aim is the improvement of the user's privacy.
- On this blockchain, we must distinguish between two kinds of addresses: **t-addresses**, which act exactly like Bitcoin addresses, and **z-addresses**, which are private and do not reveal the coins (ZEC) that have been spent thanks to the use of zk-SNARKs.
- ZEC contained in the *z*-addresses form the *shielded pool*.

De-anonymization of ZCash

- There are four different kinds of transactions in ZCash.
- t-addresses can be clustered with the same techniques we have described for Bitcoin addresses. Similarly, t-to-t transactions can be studied with the same ML algorithms.
- However, if a t-to-z transaction can be linked to a z-to-t transaction, we can effectively reduce the size of the shielded pool.

Cryptography is good to enhance privacy only if the protocol enforces it for every transaction!

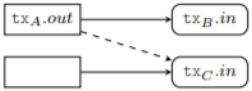
De-anonymization of Monero

A third blockchain that uses the UTXO model is **Monero**.

- Monero launched in 2014 with the idea of using modern cryptographic algorithms to improve the blockchain privacy.
- Monero obscures the transaction graph by including decoy transaction inputs, called **mixins**, and the sender of a transaction is protected by the use of **ring signatures**.
  The recipient must use every time a new **one-time address**.
- Starting September 2017, the amount of a transaction is kept hidden thanks to **confidential transactions** (RingCT).

De-anonymization of Monero

- At first, users were not obliged to use mixins for their transaction inputs.
- Researchers designed some heuristics that were able to reveal in most situations the real input even for transactions that used at least one mixin.



- Monero solved this problem by updating the protocol, so that each user had to use at least one mixin. As of 2021, every transaction input has exactly 10 mixins.

De-anonymization of Monero

- RingCT also helped resolving this information leak, because it hid transaction amounts so mixins could be chosen with more randomness.
- There have been no known attacks to the Monero blockchain after the RingCT protocol became mandatory for every user.
- Monero is, to date, the only blockchain that seems to have convincingly solved the problem of de-anonymization.

De-anonymization of Ethereum

Differently from the blockchains described until now, **Ethereum** is based on the account model.

- Given this different structure, Machine Learning techniques like address clustering are useless on Ethereum.
- However, clustering can be utilized in a different way: we can group together smart contracts deployed on the blockchain.
- In fact, they can be compared and grouped together using specific distance measures, like the **Levenshtein distance**.

De-anonymization of Ethereum

- Researchers found that the majority of smart contracts are used to generate tokens. There have also been a huge number of DeFi-related contracts in the recent years.

A classic de-anonymization approach is linking different Ethereum accounts owned by the same users.

- A recent work has described some models that can characterize a subset of users, applying *quasi-identifiers* for address de-anonymization activities.
- Specifically, these quasi-identifiers are the active time of the day, the selected gas price and the location of the addresses in the Ethereum transaction graph.

De-anonymization of Ethereum

- The Ethereum transaction graph was analyzed with **Graph Representation Learning** techniques, a novel field of Machine Learning applied to graphs.
- The study tested several algorithms, and the best results were obtained from **Diff2Vec** and **Role2Vec**.
- This heuristic showed really promising results and has shown that it can limit the real size of the anonymity set.

References I

- Antonopoulos, A., *Mastering Bitcoin: Unlocking Digital Cryptocurrencies*, O'Reilly (2017).
- Ermilov, D. et al, *Automatic Bitcoin Address Clustering*, (2017).
- https://www.chainalysis.com/
- Sun Yin, H. et al, *Regulating Cryptocurrencies: A Supervised Machine Learning Approach to De-Anonymizing the Bitcoin Blockchain*, (2019).
- Ranshous, S. et al, *Exchange Pattern Mining in the Bitcoin Transaction Directed Hypergraph*, (2017).

References II

- Jourdan, M. et al, *Characterizing Entities in the Bitcoin Blockchain*, (2018).
- Möser, M. et al, *An Empirical Analysis of Traceability in the Monero Blockchain*, (2018).
- Kappos, G. et al, *An Empyrical Analysis of Anonymity in ZCash*, (2018).
- Norvill, L. et al, *Automated Labeling of Unknown Contracts in Ethereum*, (2017).
- Béres, F. et al, *Blockchain is Watching You: Profiling and Deanonymizing Ethereum Users*, (2020).
- https://wasabiwallet.io/