Deanonymization and linkability of cryptocurrency transactions based on network analysis

Alex Biryukov, Sergei Tikhomirov

University of Luxembourg

17 June 2019
Euro S&P
Stockholm, Sweden
Outline

Introduction

Transaction clustering
  Parallel connections
  Weighting timestamp vectors
  Correlation matrix
  Measuring anonymity

Experimental results
  Estimating the source IP

Discussion

Conclusion
Deanonymization and linkability of cryptocurrency transactions based on network analysis

Biryukov, Tikhomirov

Outline

Introduction

Transaction clustering
  Parallel connections
  Weighting timestamp vectors
  Correlation matrix
  Measuring anonymity

Experimental results
  Estimating the source IP

Discussion

Conclusion
Bitcoin

- The first to solve double-spending with proof-of-work
- Senders broadcast transactions into a P2P network
- Miners construct blocks (thus confirming transactions)
Privacy in Bitcoin

- Transactions not linked to "real-world" identity
- Users can generate as many key pairs as they wish
- False sense of privacy?
Taint analysis heuristics

- All transaction inputs *probably* belong to the sender
- One output *probably* also belongs to the sender

*Figure: Bitcoin transaction structure*
Privacy coins hinder blockchain analysis...

- Dash: mixing by masternodes
- Monero: ring signatures
- Zcash: zk-SNARKs

Introduction

Tx clustering
- Parallel connections
- Weighting timestamp vectors
- Correlation matrix
- Measuring anonymity

Experiments
- Estimating the source IP

Discussion

Conclusion
...but what about network analysis?

- How do messages propagate through the network?
- What does a well-connected adversary learn?
- Is it possible to link txs by the same user?
Our contributions

▶ We introduce a **new transaction clustering method** based on weighted vectors of IP addresses

▶ We validate our method with **experiments on Bitcoin and three major privacy-focused cryptocurrencies**
Outline

Introduction

Transaction clustering
  Parallel connections
  Weighting timestamp vectors
  Correlation matrix
  Measuring anonymity

Experimental results
  Estimating the source IP

Discussion

Conclusion
Message propagation in Bitcoin

**Figure:** Bitcoin’s 3-step message exchange
Broadcast randomization in Bitcoin and forks

- trickling: send to a random subset once every 100 ms
- diffusion: send to each neighbor after a random delay
Intuition

Transactions issued from the same node have correlated broadcast patterns.
Outline of our clustering method

- Establish parallel connections to many nodes
- Log timestamps of received tx announcements
- For each tx, consider IPs which announced it to us
- Cluster transactions with "similar" IP vectors
- Measure the decrease in anonymity
Parallel connections

- Default connections: 8 outgoing + up to 117 incoming
- We are unlikely to get a new tx quickly with only one connection per node
- `bcclient` establishes parallel connections to nodes
- Bitcoin and Zcash show similar distribution of free slots
Deanonymization and linkability of cryptocurrency transactions based on network analysis

Biryukov, Tikhomirov

Introduction

Tx clustering

Parallel connections

Weighting timestamp vectors

Correlation matrix

Measuring anonymity

Experiments

Estimating the source IP

Discussion

Conclusion

Bitcoin free slots

Open slots

Share of peers with that number of open slots (%)
Zcash free slots
Weighting timing vectors

IP addresses $p_i$ announce a new tx to us at times $t_i$. We assign exponentially decreasing weights to $p_i$:

$$w(p_i) = e^{-(t_i/k)^2}$$

where the median IP gets weight 0.5:

$$k = \frac{t_{median}}{\sqrt{-\ln(0.5)}}$$
Weighting timing vectors: example

High values indicate higher probability of an IP to be the sender or one of its entry nodes.

Figure: Weight functions for 3 timestamp vectors
Clustering the correlation matrix

- For each pairwise correlations of weight vectors of txs
- Hypothesis: correlation matrix has a *block-diagonal* structure
- With a right permutation of rows and columns, related transactions will form clusters along the main diagonal
Heatmap visualization

- Display correlations between weight vectors as matrix
- Darker color means higher correlation
- Matrix is symmetric by definition: $\text{corr}(i,j) = \text{corr}(j,i)$
- The main diagonal is black: correlation with oneself
Measuring anonymity

We use anonymity degree proposed by Díaz et al.¹:

\[
d = - \sum_{i=1}^{N} p_i \log_2(p_i) / \log_2(N)
\]

where \( p_i \) is the probability of the \( i \)-th tx to originate from the given source.

- \( d = 1 \): users are equally likely to be the senders of a given message
- \( d = 0 \): the attacker knows the senders of all messages

¹Díaz, Seys, Claessens, Preneel. Towards measuring anonymity. 2002
Putting the pieces together

- Connect to many nodes from servers on 3 continents
- Log transaction announcements
- Assign weights to vectors of timestamps
- Calculate pairwise correlations between weight vectors
- Apply the spectral co-clustering algorithm \(^2\)
- Calculate anonymity degree for our txs as ground truth
- Ethical considerations: mostly testnet, our own txs

\(^2\)I.S. Dhillon. Co-clustering documents and words using bipartite spectral graph partitioning. 2001
Deanonymization and linkability of cryptocurrency transactions based on network analysis

Biryukov, Tikhomirov

Outline

Introduction

Transaction clustering
  Parallel connections
  Weighting timestamp vectors
  Correlation matrix
  Measuring anonymity

Experimental results
  Estimating the source IP

Discussion

Conclusion
Bitcoin testnet: anonymity degree = 0.63
Deanonymization and linkability of cryptocurrency transactions based on network analysis

Biryukov, Tikhomirov

Bitcoin mainnet: anonymity degree = 0.88

Only connected to 1/10 of nodes, didn’t occupy all slots.
Deanonymization and linkability of cryptocurrency transactions based on network analysis

Biryukov, Tikhomirov

Introduction

Tx clustering
- Parallel connections
- Weighting timestamp vectors
- Correlation matrix
- Measuring anonymity

Experiments
- Estimating the source IP

Discussion

Conclusion

Zcash: anonymity degree = 0.86
Monero

Experiment without our own transactions.
Dash

dash-mainnet. N = 4, 9 clusters

Experiment without our own transactions.
Estimating the source IP from ADDR messages

- A new node advertises its IP in ADDR messages
- We intersect the announced IPs from ADDRs with the highest-weighted IPs in tx clusters (Bitcoin testnet)
- In most experiments, the source IP appeared among top-5 highest weighted IPs in our transaction cluster
Deanonymization and linkability of cryptocurrency transactions based on network analysis

Biryukov, Tikhomirov

Introduction

Transaction clustering
  Parallel connections
  Weighting timestamp vectors
  Correlation matrix
  Measuring anonymity

Experimental results
  Estimating the source IP

Discussion

Conclusion
Cost of attack

- Feasible for a moderately resourceful attacker
- Main cost components are bandwidth and storage
- We estimate the cost of a full-scale attack on Bitcoin mainnet at hundreds of US dollars
- Our experiments cost $35 on AWS
Countermeasures

- Don’t issue many txs in the same session
- Run nodes with increased number of connections
- Periodically drop and re-establish random connections
- Implement stronger broadcast randomization
Countermeasures (contd): new relay protocols

- **Dandelion++**: two-stage propagation for better anonymity. Only outgoing connections for first phase. Hard to force a remote node to connect to us

- **Erlay** (proposed 2019-05-28): ”[A]nnouncements are only sent directly over a small number of connections (only 8 outgoing ones). [...] We [...] better withstand timing attacks”
Deanonymization and linkability of cryptocurrency transactions based on network analysis

Biryukov, Tikhomirov

Outline

Introduction

Transaction clustering
  Parallel connections
  Weighting timestamp vectors
  Correlation matrix
  Measuring anonymity

Experimental results
  Estimating the source IP

Discussion

Conclusion
Conclusion

- Announcement timings reveal related transactions
- Randomization techniques are not very efficient
- Clustering works better on small networks
Future work: mobile wallets

- In our experiments, txs were issues from a full node
- How are mobile wallets different in terms of networking?
- Can we cluster transactions issued from mobile wallets?
Questions?

- cryptolux.org (we are hiring postdocs)
- s-tikhomirov.github.io
Image credits