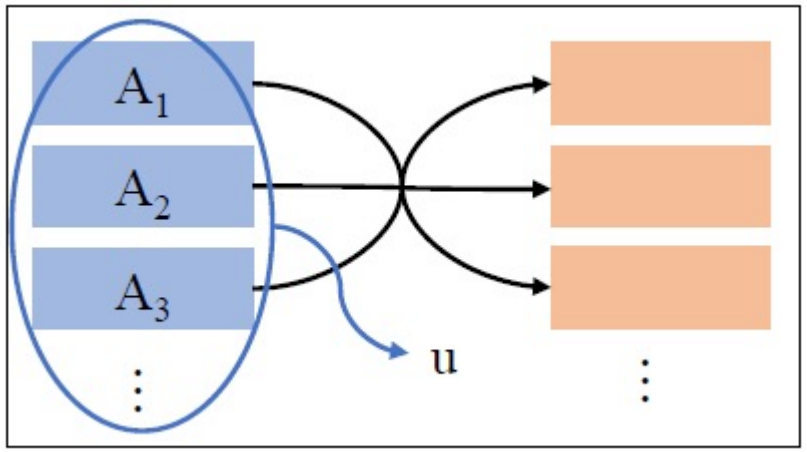
Address Clustering in Bitcoin

Xinyu Liu

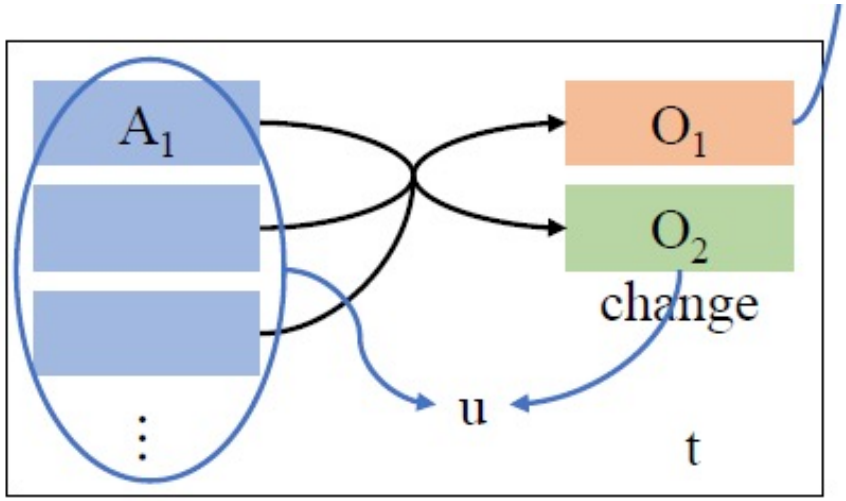
BUPT

Address Clustering

Techniques to group the addresses that belongs to the same user.

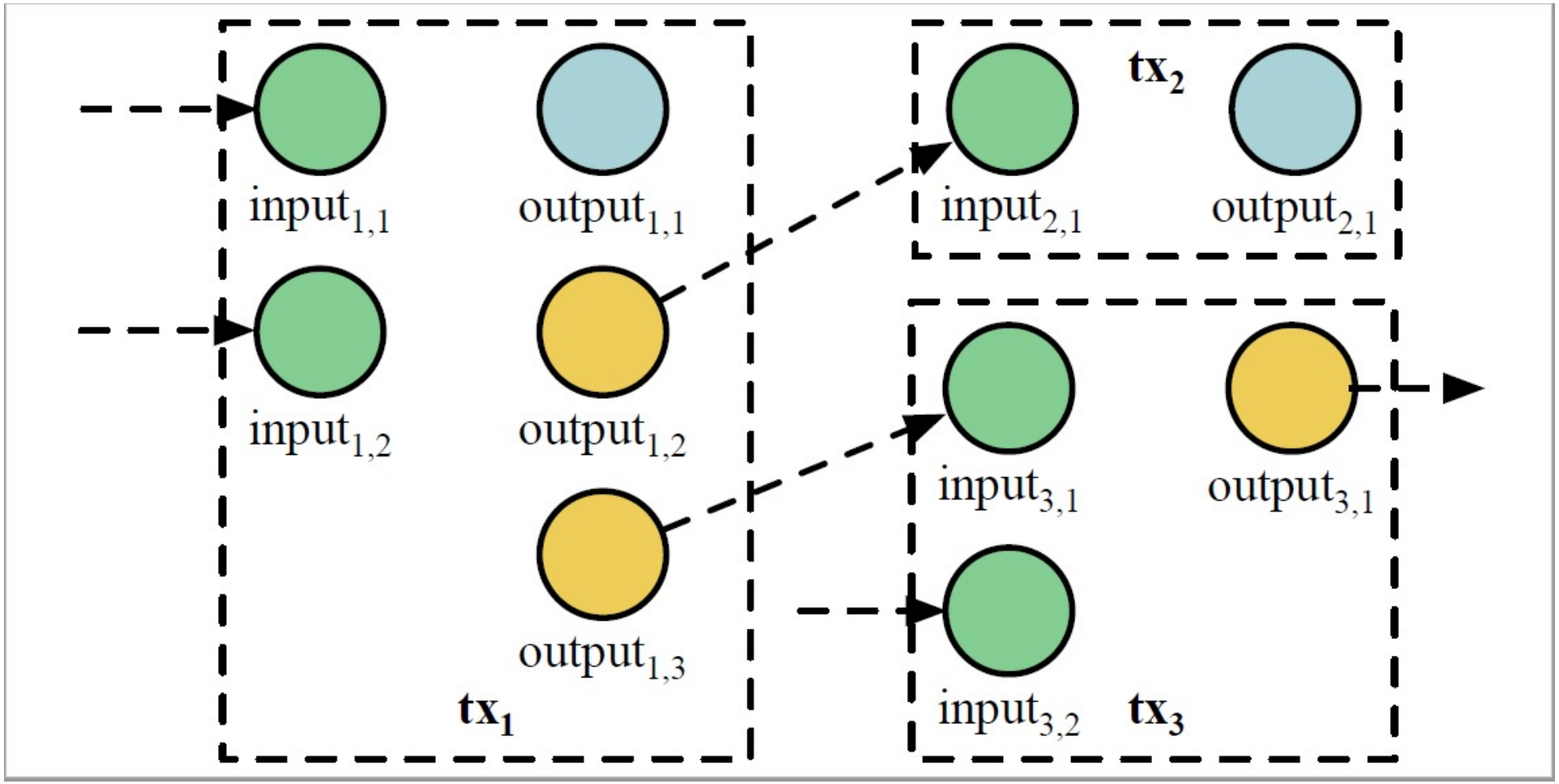


Heuristic 1. Common Spending



Heuristic 1. Change Address

Bitcoin Data Structure



Resurrecting Address Clustering in Bitcoin

Möser, Malte, and Arvind Narayanan. "Resurrecting Address Clustering in Bitcoin." arXiv preprint arXiv:2107.05749 (2021).

Contributions

- A new ground truth method and dataset.
- Evaluating existing heuristics.
- Combine different heuristics to improve prediction.
- Preventing cluster collapse.
- Assessing impact.

Build Ground Truth Data Set

Methodology

Heuristics

- One of the outputs is a payment, the other output receives the change(change address).
- change outputs are sometimes revealed by common spending at a later point in time due to address reuse.

Method

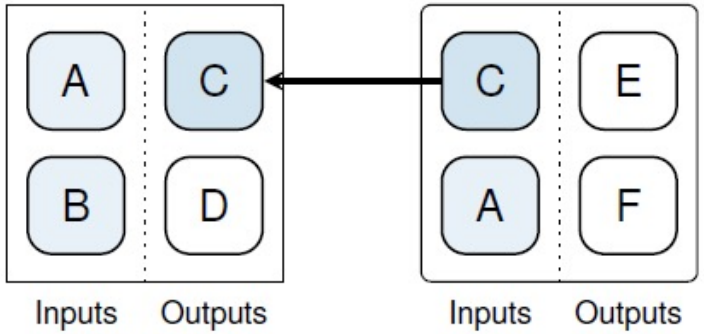


Figure 4: Address *C* is merged into the same cluster as addresses A and B by the multi-input heuristic, thereby revealed as the change address in the first transaction.

Discuss

What are the advantages and disadvantages of this method compared with interactive collection?

Advantages

- Variety
  - Not limit to a small use cases and entities.
  - Not limit to a small period of time.
- Scale

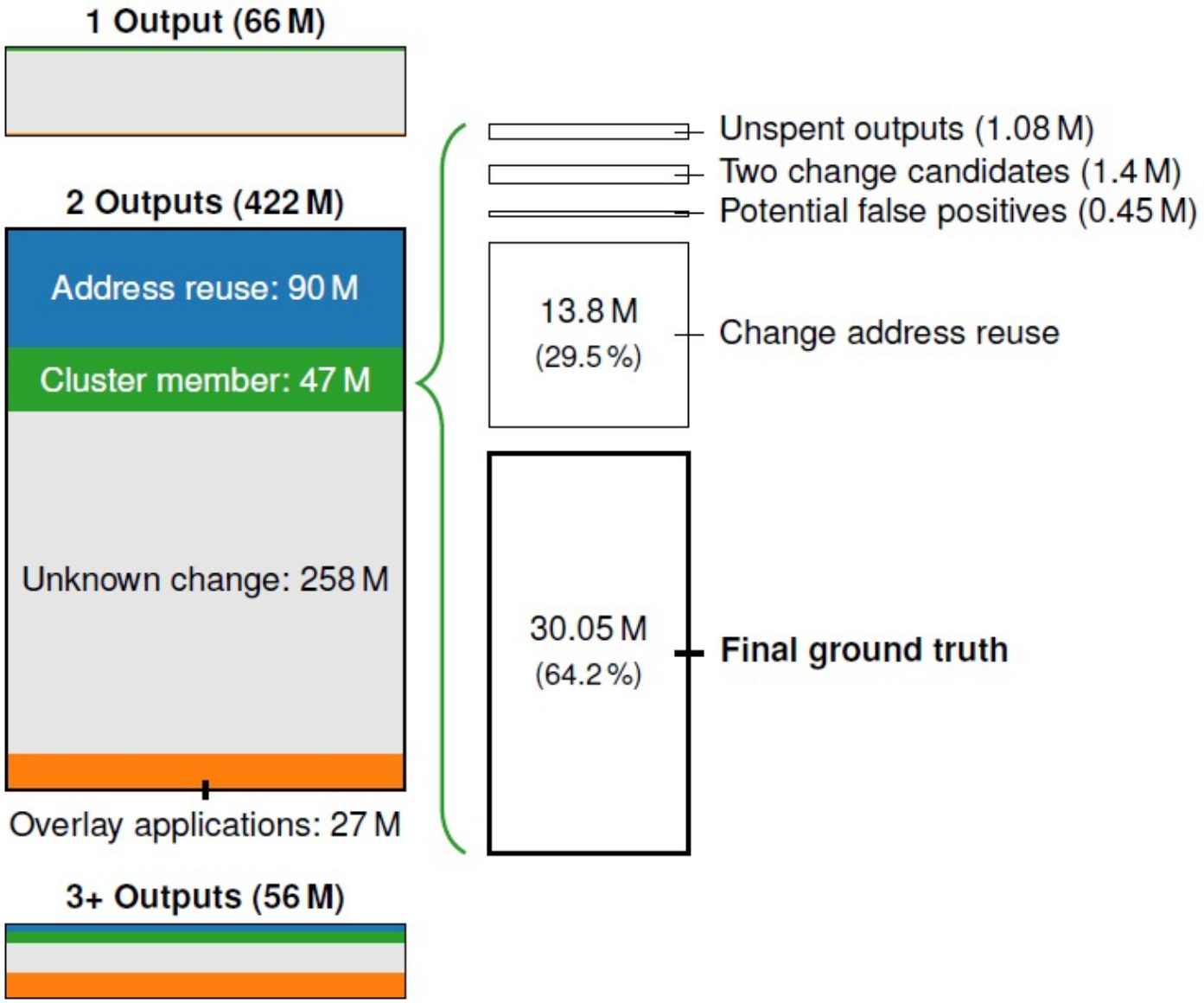
Disadvantages

- Veracity
  - Cannot fully verify the correctness.
- Biased
  - Biased towards to the entities that are more prone to reuse or merge addresses.

Data Collection

- BlockSci v0.7
- Until the end of June 2020

Harry Kalodner, Malte Möser, Kevin Lee, Steven Goldfeder, Martin Plattner, Alishah Chator, and Arvind Narayanan. "Blocksci: Design and applications of a blockchain analysis platform". In: 29th USENIX Security Symposium. 2020, pp. 2721–2738.



Assess

Scale and Time Frame

- about 7.7% of standard transactions
- about 5.6% of all transactions
- The percentages are relatively stable over time.

* Standard Transactions: One output is a payment, and the other output is a change address.

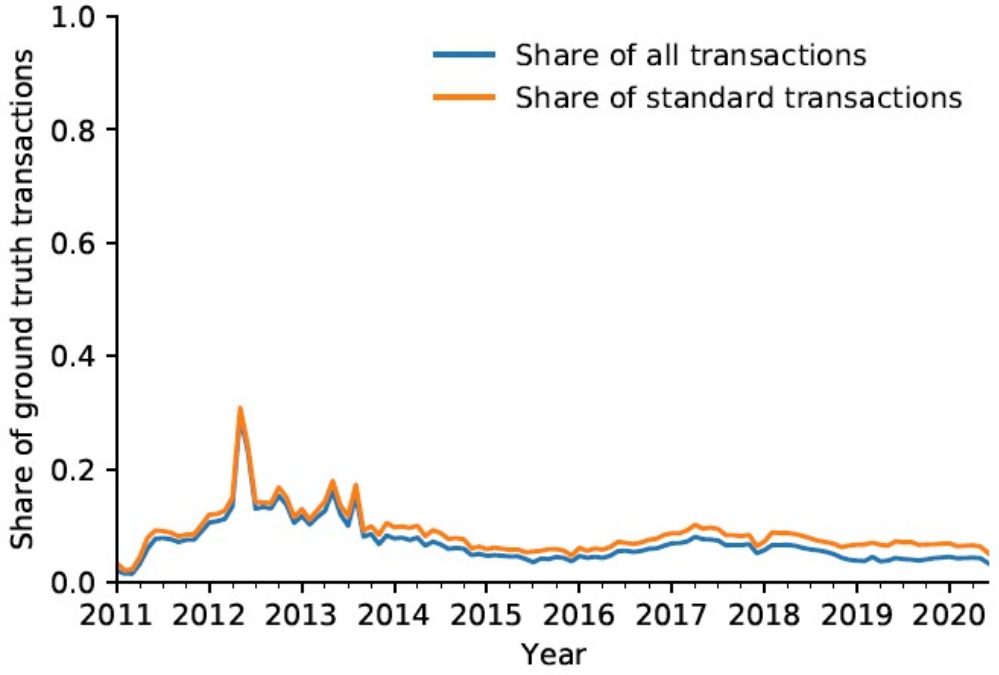
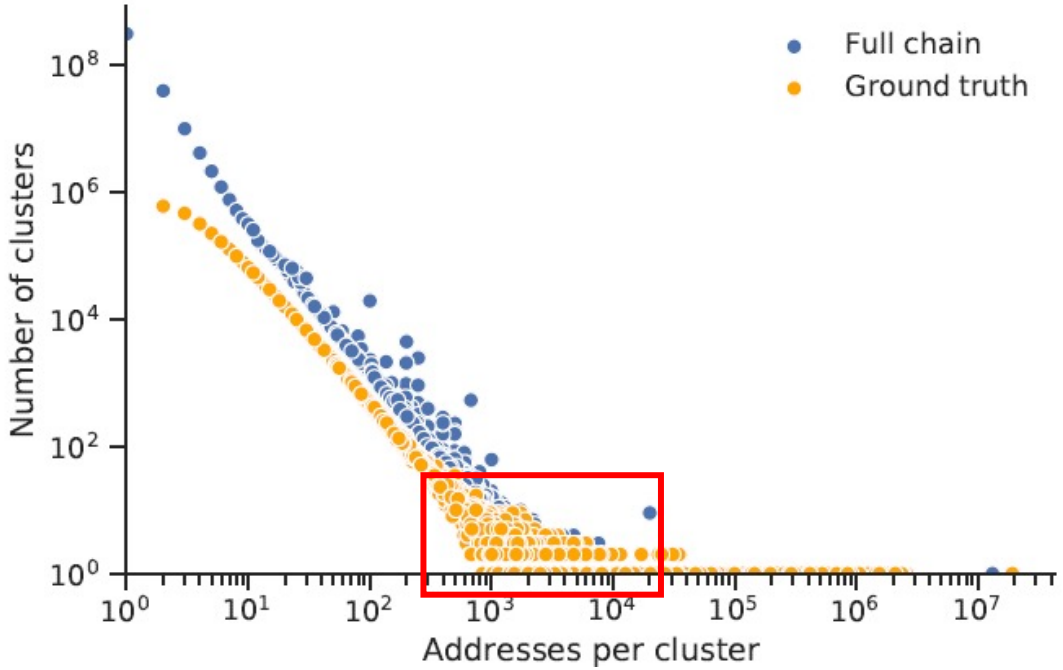


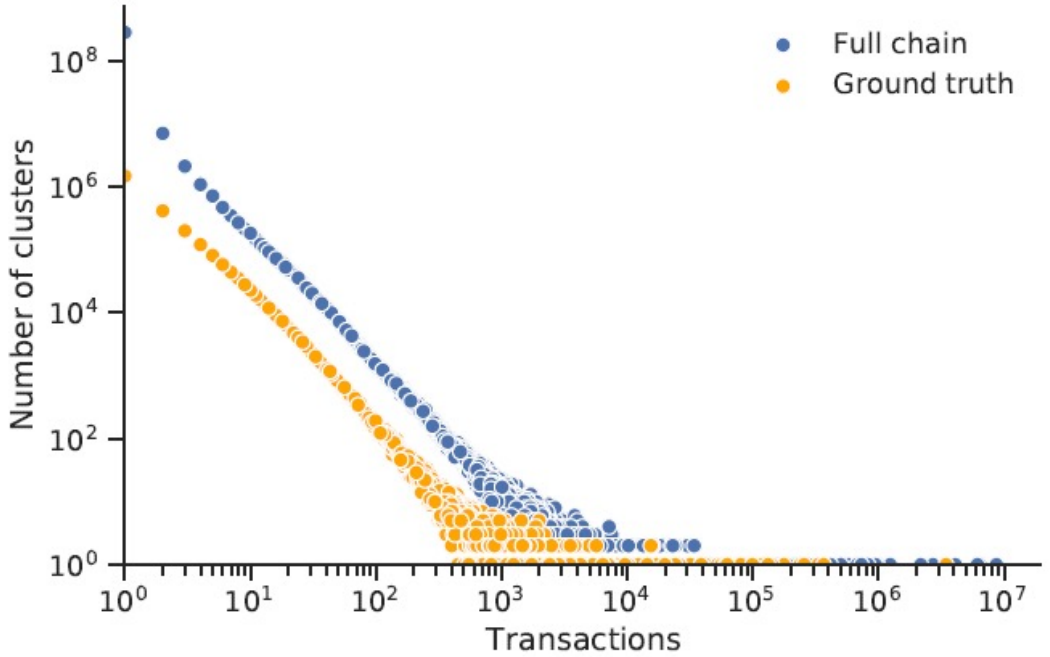
Figure 6: Share of ground truth transactions of all and standard transactions over time.

Assess

Variety of Included Clusters

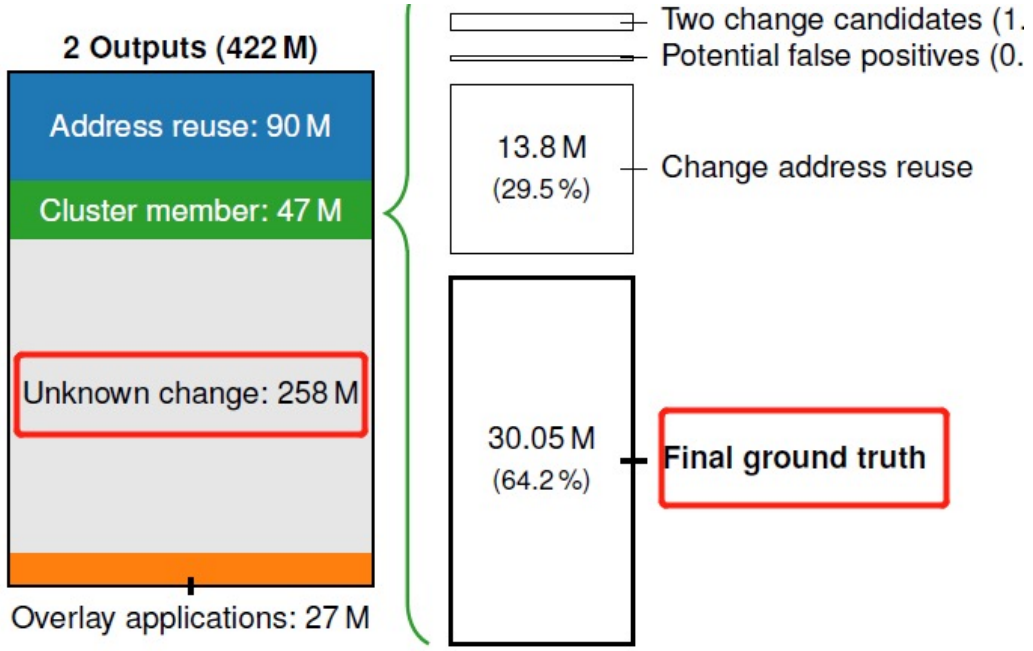
- ground truth includes transactions from 2.7 million base.
- the number of addresses
- the number of transactions





Assess

Transaction composition and protocol features



Characteristic	Ground truth (%)	Remaining (%)
1 Input	38.96	78.28
2 Inputs	21.91	13.92
3+ Inputs	39.13	7.79
Version = 1	81.56	83.36
Locktime > 0	26.10	24.22
RBF	2.77	4.26
SegWit	14.89	23.36
n (in million)	30.05	258.68

Table. Comparison of transaction characteristics between ground truth transactions and transactions with 2 outputs for which change is unknown.

Evaluating Change Output Heuristics

Heuristics

Table . Change heuristics proposed in the literature and used in this paper.

Heuristic	Notes and limitations	Used	Refs.
Optimal change: There should be no unnecessary inputs: if one output is smaller than any of the (2+) inputs, it is likely the change.	Only applies to transactions with 2+ inputs. We use two variants, one ignoring and one accounting for the fee.	✓	[34, 36]
Address type: The change output is likely to have the same address type as the inputs.	Wallets could use different address types to obfuscate the change output.	✓	[24, 36]
Power of ten: As purchase amounts may be rounded, and the change amount also depends on input values and the fee, it is more likely to have fewer trailing zeros.	We use six different variants, which are partially redundant.	✓	[24, 36]
Shadow address: Many clients automatically generate fresh change addresses, whereas spend addresses may be more easily reused.	Modern wallets discourage reuse of receiving addresses. We do not use the heuristic because our ground truth is filtered based on address freshness.	x	[2, 30]
Consistent fingerprint: The transaction spending a change output should share the same characteristics. We use 17 variants based on the following characteristics: • input/output counts and order • version • locktime • serialization format (SegWit) • replace-by-fee (RBF) • transaction fee • input coin age (zero-conf) • address and script types	False positives are possible when a wallet implementation or the protocol change. We only consider characteristics after they are available in the protocol. Appendix A describes the characteristics we use in more detail.	✓	[8, 36]

Evaluate

Table. True and false positive rates of heuristics applied to transactions in the ground truth data set.

Heuristic	Ground Truth TPR	Ground Truth FPR	Remaining Coverage*
Universal heuristics			
Optimal change	0.299	0.027	0.134
• incl. fee	0.232	0.020	0.094
Address type	0.210	0.028	0.339
Power of ten			
• $n = 2$	0.489	0.012	0.405
• $n = 3$	0.444	0.006	0.335
• $n = 4$	0.400	0.005	0.277
• $n = 5$	0.326	0.006	0.191
• $n = 6$	0.229	0.005	0.115
• $n = 7$	0.115	0.001	0.053
Consistent fingerprint			
Output count	0.272	0.125	0.429
Input/output count	0.264	0.108	0.572
Version	0.224	0.003	0.297
Locktime	0.302	0.003	0.356
RBF	0.059	0.002	0.088
SegWit	0.155	0.019	0.224
SegWit-conform	0.023	0.001	0.029
Ordered ins/outs	0.241	0.053	0.430
Zero-conf	0.101	0.055	0.213
Absolute fee	0.134	0.029	0.335
Relative fee	0.045	0.009	0.215
Multisignature	0.137	0.000	0.151
Address type			
• P2PKH	0.214	0.014	0.287
• P2SH	0.236	0.012	0.305
• P2WPKH	0.146	0.017	0.219
• P2WSH	0.050	0.007	0.068
All address types	0.252	0.021	0.348

*Coverage denotes share of standard transactions with yet unidentified change where the heuristic returned exactly one output.

Individual heuristic doesn't perform well both in accuracy and coverage.

Evaluate

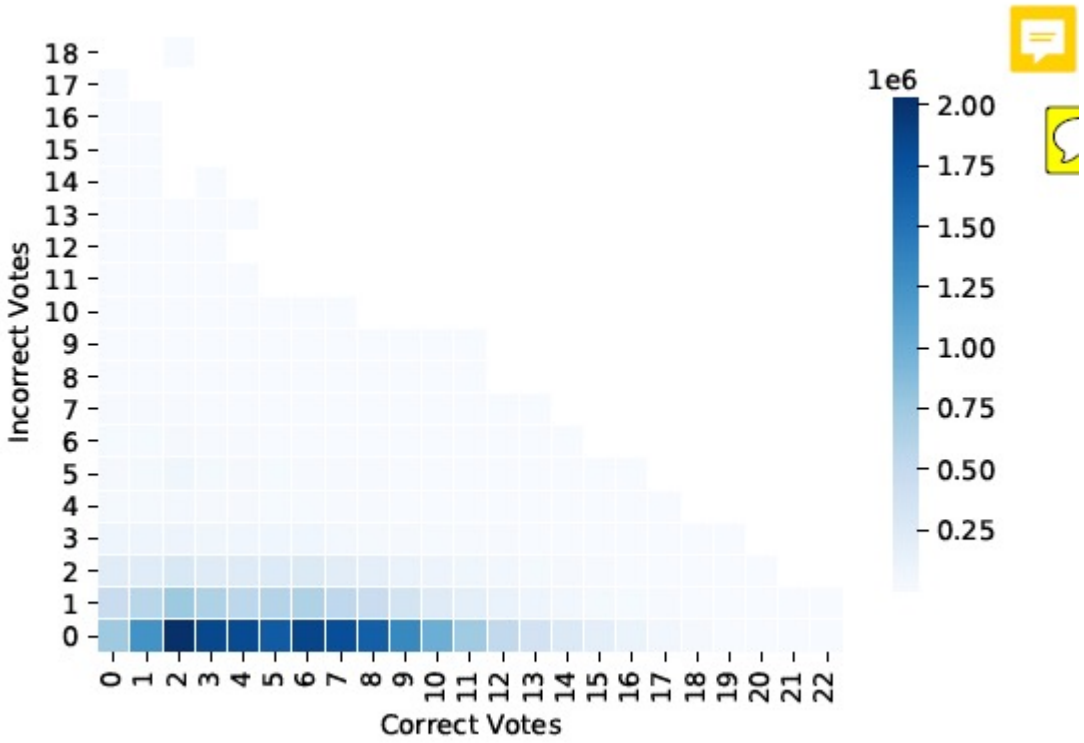


Figure. Number of votes from heuristics

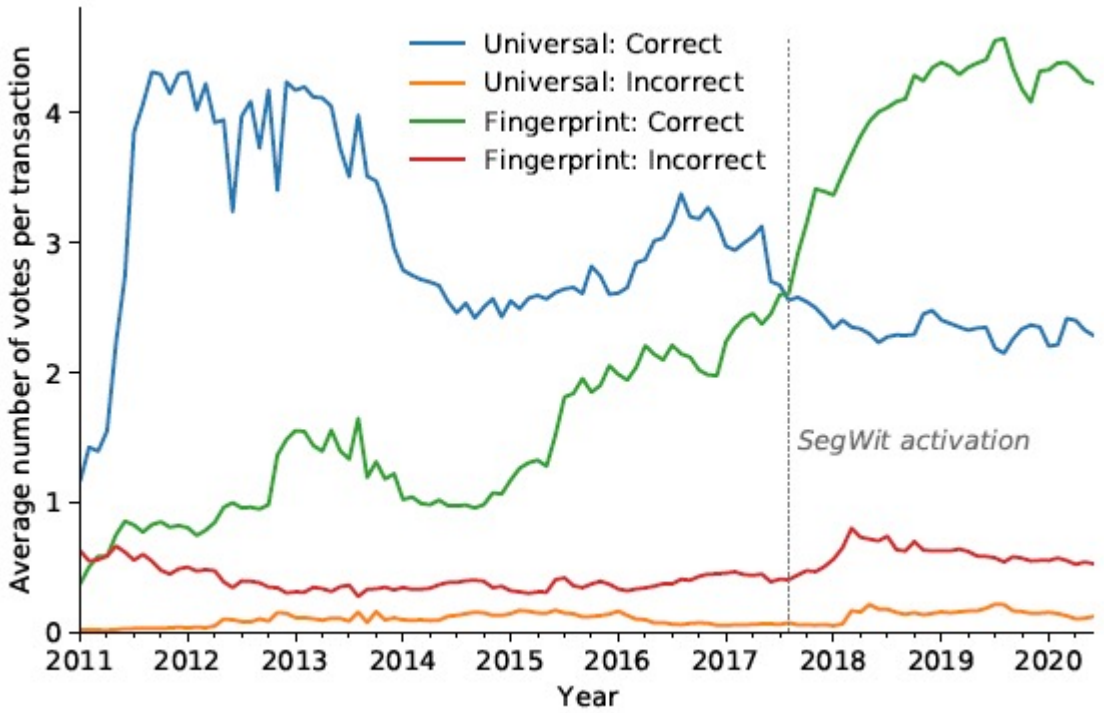


Figure. Average number of correct and incorrect votes per transaction and type of heuristic over time

96.93% has been predicted correctly by at least one heuristic.

Combine Heuristics

- Threshold vote
- Random forest classifier

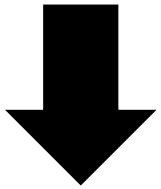
Threshold vote

- Methodology

if there are at least t more votes for output a than for output b, then output a is considered the change.

- Cons

- Heuristics have varying TPR and FPR.
- Dependability may change in different time period



Random Forest Classifier

Random Forest

Features

output features

- an output's value/total output value
- an output's index

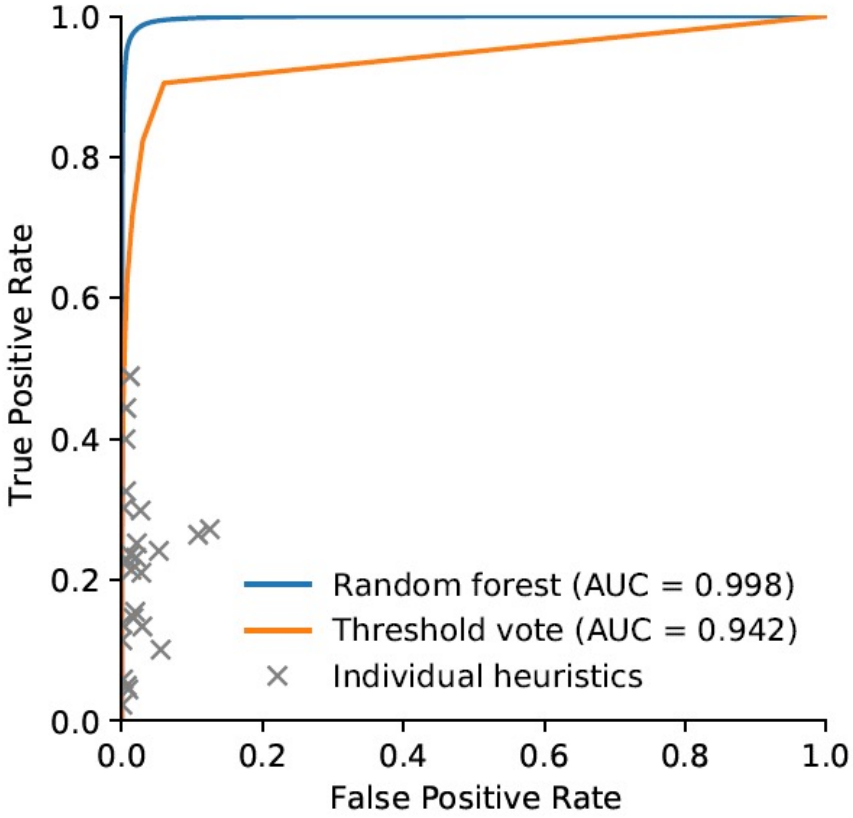
Transaction features

- Total value
- Transaction fee per byte
- Version
- Whether use SegWit
- Whether set a none-zero locktime
- The number of inputs
- The time of inclusion

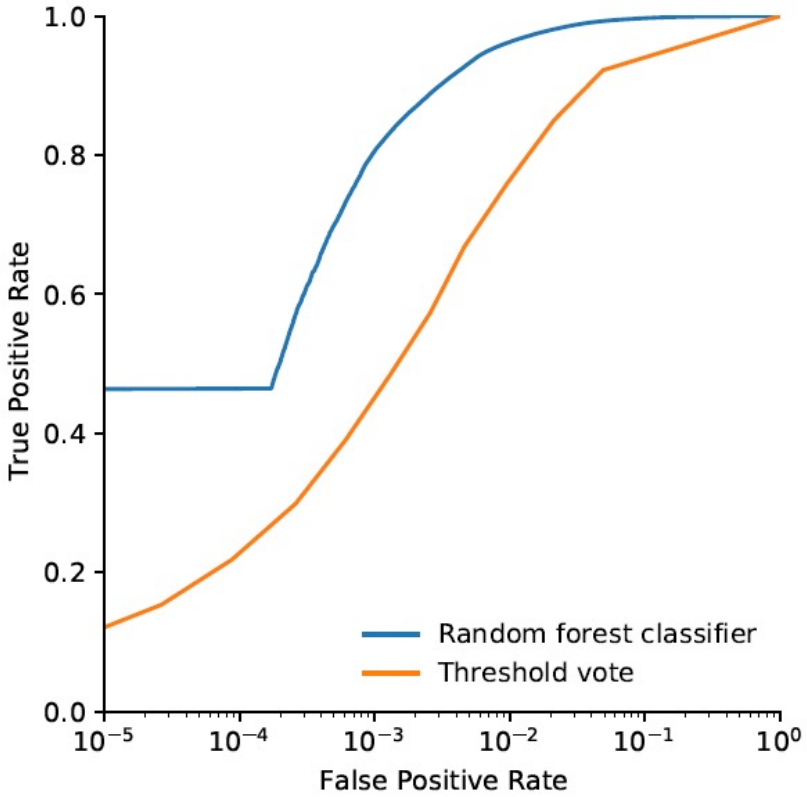
Parameters

- training set : test set = 8 : 2
- max_features: 5
- min_samples_split: 50

Result



The random forest performs better than threshold vote.



The random forest achieves much higher true positive rates at low false positive rates

Model Validation

Dataset 1:

- 11197 transactions
- AUC : 99.6%

Danny Yuxing Huang, Maxwell Matthaios Aliapoulios,Vector Guo Li, Luca Invernizzi, Elie Bursztein, Kylie McRoberts, Jonathan Levin, Kirill Levchenko, Alex C Snoeren, and Damon McCoy. "Tracking ransomware end-to-end". In: IEEE Symposium on Security and Privacy. IEEE. 2018, pp. 618–631.

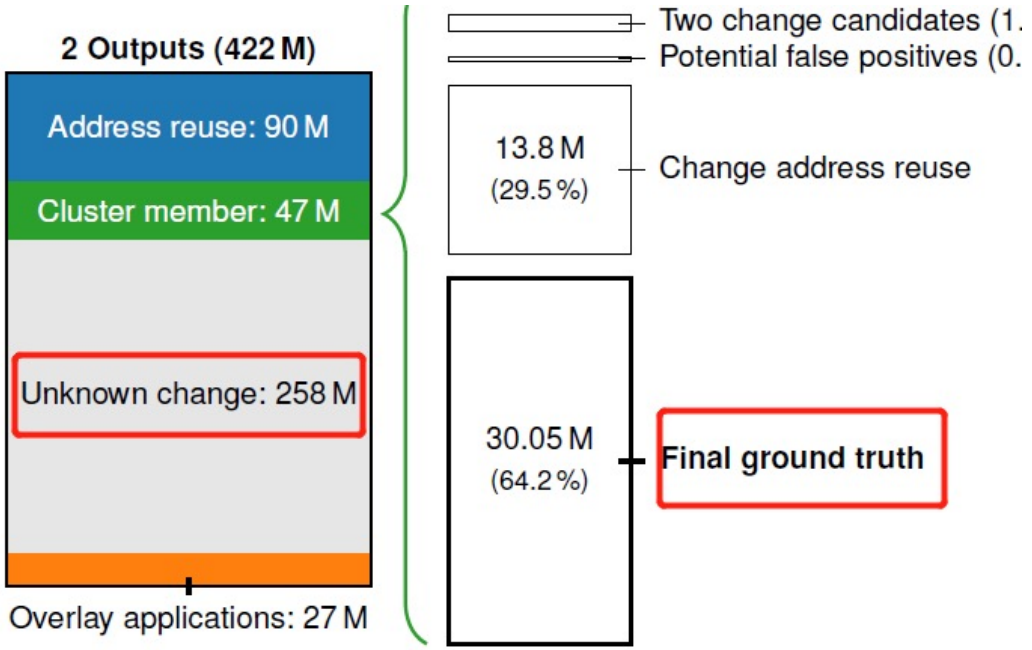
Dataset 2:

- 267993 transactions
- AUC:
- 97.3%

GraphSense Public TagPacks. URL:https://github.com/graphsense/graphsense- tagpacks (visited on 04/01/2021).

Enhance the base clustering using new model

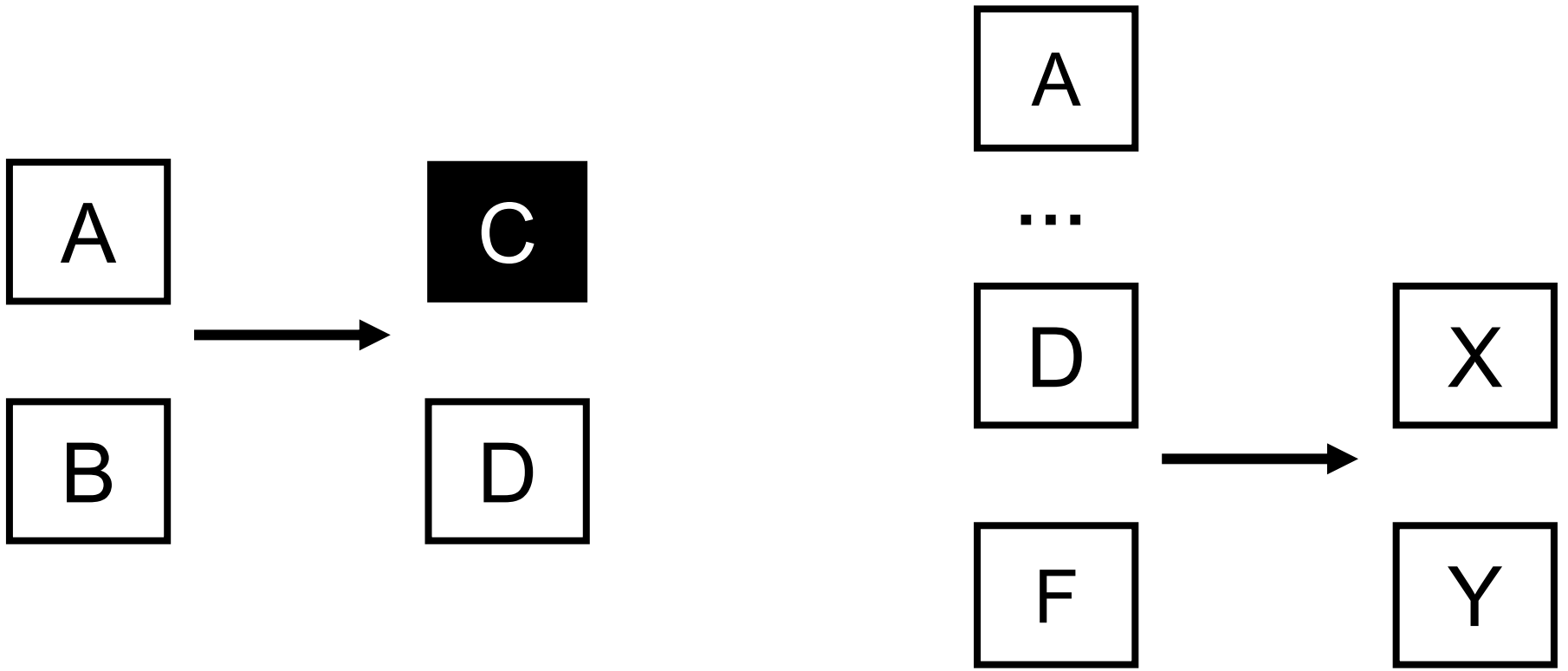
Unknown change dataset



- exclude 8.9 million transactions where no heuristic identified a change output.

Random forest: 119.86 million change outputs (46.34% of transactions)

Constraints prevent cluster collapse



Cannot be merged !!

Result

Naïve cluster leads to sever collapse:
Inspecting the 273 labeled clusters from the Graphsense tag pack, we find that 148 of them have been merged into this supercluster.
Constraint merge prevent the above collapse:
Assessing the 273 labeled clusters, there are only ten instances left where two labeled clusters are merged together

Impact on Blockchain Analyses

Cashout flows from darknet markets to exchanges

- Using address tags from the GraphSense tag pack to identify relevant clusters.
  - 15Darknet markets to 117 exchangesss
- Using the model to predict the unknown change transactions.

The median increase in output value across all 15 exchanges amounts to 13.86%. The total amount of bitcoins flowing from the darknet markets to exchanges increases from BTC 821500 to BTC 961519 (a 17% increase).

Remove the influence of self-payments

Clustering can be used to remove self-payments of users (such as change outputs), which would artificially inflate estimates of economic activity.

reduces the estimate of bitcoins moved per day between January 2017 to June 2020 by about 11.7%.

Harry Kalodner, Malte Möser, Kevin Lee, Steven Goldfeder, Martin Plattner, Alishah Chator, and Arvind Narayanan. "Blocksci: Design and applications of a blockchain analysis platform". In: 29th USENIX Security Symposium. 2020, pp. 2721–2738.