Airdrops and Privacy: A Case Study in Cross-Blockchain Analysis

Preprint · September 2018

CITATIONS
0

READS
57

3 authors, including:

Martin Harrigan
Institute of Technology, Carlow
27 PUBLICATIONS **421** CITATIONS

SEE PROFILE

Airdrops and Privacy: A Case Study in Cross-Blockchain Analysis

Martin Harrigan*[1], Lei Shi†[1], and Jacob Illum‡[2]

[1]Institute of Technology, Carlow
[2]Chainalysis Inc., New York

Abstract

Airdrops are a popular method of distributing cryptocurrencies and tokens. While often considered risk-free from the point of view of recipients, their impact on privacy is easily overlooked. We examine the Clam airdrop of 2014, a forerunner to many of today's airdrops, that distributed a new cryptocurrency to every address with a non-dust balance on the Bitcoin, Litecoin and Dogecoin blockchains. Specifically, we use address clustering to try to construct the one-to-many mappings from entities to addresses on the blockchains, individually and in combination. We show that the sharing of addresses between the blockchains is a privacy risk. We identify instances where an entity has disclosed information about their address ownership on the Bitcoin, Litecoin and Dogecoin blockchains, exclusively via their activity on the Clam blockchain.

1 Introduction

An airdrop is a free distribution of a cryptocurrency or token. It generally involves a large number of recipients and uses external systems, for example, existing blockchains or centralised identity providers, to prevent Sybil attacks. In practice, the parameters of an airdrop are arbitrarily set by the distributor. Airdrops are often considered risk-free since the recipients do not part with anything of value. However, this is not true in the general case. For example, one method of preventing Sybil attacks uses addresses on existing blockchains to select the recipients. In order to claim the free cryptocurrency or tokens, a recipient must reuse his or her existing private-keys to create a transaction on the blockchain belonging to the airdrop. The contents of this transaction, in particular the reuse of addresses across blockchains, its relationship with other

*Email: martin.harrigan@itcarlow.ie
†Email: lei.shi@itcarlow.ie
‡Email: jacob@chainalysis.com

transactions, and the very act of broadcasting the transaction itself may inadvertently disclose information about the recipients to interested third parties.

This paper considers the impact of the Clam airdrop on privacy across three blockchains: Bitcoin, Litecoin and Dogecoin. We use *address clustering* to try to construct the one-to-many mappings from entities to addresses on the blockchains, individually and in combination. An address clustering partitions the set of addresses observed on a blockchain into maximal subsets of addresses that are controlled by the same entity. Each component of the partition is an address cluster. When combined with address tagging, that is, associating real-world identities with addresses, and graph analysis, it is an effective means of analysing and deanonymising blockchain activity at both the micro- and macro-levels, see, e.g., [1, 5, 10, 12, 16, 18, 20].

Experimental analysis has shown that a single heuristic for address clustering, the *multi-input heuristic*, can identify more than 69% of the addresses in the wallets stored by lightweight clients [14]. This heuristic assumes that the addresses referenced in transaction outputs spent in a single multi-input transaction are controlled by the same entity [13]. Although vulnerable to techniques such as CoinJoin [11] and its kin, it is a useful heuristic in practice [7]. The analyses in this paper use the multi-input heuristic exclusively but can be extended to any number of other heuristics, including those that reduce the rate of false positives [6].

We examine the impact of the address clustering for the Clam blockchain on the address clusterings for the Bitcoin, Litecoin and Dogecoin blockchains. In other words, we identify cases where an entity has disclosed information about their address ownership on the Bitcoin, Litecoin and Dogecoin blockchains, exclusively via their activity on the Clam blockchain.

The paper is organised as follows. In Section 2 we briefly list related work in the areas of cross-blockchain analysis, and airdrops. In Section 3 we describe the four blockchains and quantify the sharing of addresses between them. We analyse the address cluster sizes and coverage and we quantify the levels of address reuse and address cluster merging. Section 4 explores the impact of the address clustering for the Clam blockchain on its counterparts individually. Section 5 generalises the approach to handle many blockchains in combination. We conclude with some future work in Section 6.

2 Related Work

We can categorise related work into two areas: cross-blockchain analysis and airdrops.

In the first category, Nieves [15] evaluated heuristics for recognising cross-blockchain transactions, in particular those facilitated by centralised services such as ShapeShift and Changelly[1]. The heuristics involve the identification of pairs of transactions from different blockchains that can be linked to known service addresses and have similar timestamps and output values. Popuri and

[1] https://shapeshift.io, https://changelly.com

Gunes [17] performed a network analysis of the address graphs derived from the Bitcoin and Litecoin blockchains. They identify power laws in their in- and out-degree distributions, in-out and out-in degree correlations that vary between negative (disassortative) and neutral (non-assortative), and global clustering coefficients that decrease with time. Kalodner et al. [9] present BlockSci, an open-source platform for blockchain analysis that supports many blockchains that are schematically similar to Bitcoin. In an example usage, the authors compute a form of velocity for Bitcoin and propose it as a useful metric for making comparisons across cryptocurrencies. There exist many comparative studies of the market performance of cryptocurrencies. For example, ElBahrawy et al. [4] compare 1400 cryptocurrencies by market capitalisation over a four year period. However, these studies are only tangentially related to the underlying blockchains.

Although airdrops are a popular method of distributing cryptocurrency, our second category has few entries. The MIT Bitcoin Project [19] organised a pseudo-airdrop by offering every MIT undergraduate $100 worth of bitcoin in October 2014. The experimenters used the airdrop to examine the behaviour of natural early adopters (NEAs) [2]. However, we are not aware of any studies that examine the mechanisms and the wider impact of airdrops.

We use common terminology from graph theory through-out the paper. Please refer to Diestel [3] or a similar reference for definitions.

3 The Four Blockchains

Due to their shared lineage, the Bitcoin, Litecoin, Dogecoin and Clam blockchains are schematically similar, i.e. they have similar block headers, transaction formats, etc. However, they have very different sizes. Table 1 compares the four blockchains based on their transaction, transaction output, address, address cluster and *non-trivial* address cluster counts. Non-trivial address clusters are those that contain more than one address; trivial address clusters contain exactly one. At the time of this analysis, the Clam blockchain is small ($\sim$ 16 million transaction outputs); the Litecoin and Dogecoin blockchains are medium-sized ($\sim$ 84 million and $\sim$ 117 million transaction outputs respectively); and the Bitcoin blockchain is large ($\sim$ 845 million transaction outputs).

Table 1: We analyse the Bitcoin, Litecoin, Dogecoin and Clam blockchains as of 1st May 2018. The last eight hexadecimal digits of the blocks at the tip of each blockchain are shown below. All analysis in this paper is based on these snapshots.

	Bitcoin	Litecoin	Dogecoin	Clam
Tip Hash Ending	`0x5bd3e36d`	`0x709b646d`	`0xe6f50d81`	`0x31873627`
# Transactions	313 522 772	23 844 704	36 371 905	5 036 940
# Transaction Outputs	845 351 818	83 998 454	117 403 235	15 648 326
# Addresses	389 757 330	28 696 848	30 573 438	3 775 072
# Address Clusters	183 312 914	15 876 469	19 521 831	3 435 622
# Non-Trivial Address Clusters	37 362 066	1 802 872	1 355 230	21 034

Clam is the least known of the four cryptocurrencies. Its source code is a fork of Blackcoin, which is a fork of Novacoin, which is a fork of Peercoin, which is a fork of Bitcoin. It derives much of its functionality from its ancestors. Clam was announced on the 24th March 2014.[2] An airdrop sent every address with a non-dust balance on the Bitcoin, Litecoin and Dogecoin blockchains, as of block heights 300 377, 565 693 and 218 556 respectively, 4.6 CLAM.

The impact of the airdrop on the sharing of addresses between the four blockchains is illustrated in Fig. 1. We observe that the Bitcoin, Litecoin and Dogecoin blockchains share a small number of addresses (see Fig. 1a). However, there is significant sharing of addresses with the Clam blockchain (see Fig. 1b). For example, of the 3 775 072 addresses on the Clam blockchain, only 566 642 are not shared with any of the other three blockchains.

3.1 Address Cluster Size and Coverage

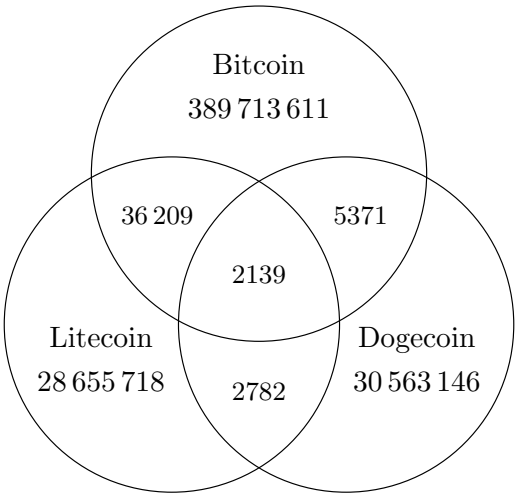
The address cluster sizes and coverage for each blockchain are visualised in Figs. 6a to 6d. The address cluster sizes are binned in the histograms. Both the horizontal and the vertical axes are log-scaled. The inset pie charts show the coverage, or total number of addresses in the address clusters, in each size range. The address clustering for the Clam blockchain differs from the others: 90% of it addresses are in trivial address clusters. This is primarily due to the airdrop; many of the addresses appear in a single transaction output that was assigned 4.6 CLAM. If we ignore this difference, then we can observe the significant coverage provided by the large address clusters: 30% of the addresses in the Bitcoin blockchain are in address clusters of size greater than 100; the comparable numbers for Litecoin and Dogecoin are 28% and 25%, respectively. These are significant when tagging address clusters with real-world identities: the large address clusters provide good coverage.

The largest address cluster on the Bitcoin blockchain is an anomaly: it originally belonged to the Mt. Gox exchange that, for a time, allowed users to import private-keys directly from their wallets. This feature causes the multi-input heuristic to produce false positives and requires additional heuristics and information not available on the blockchain in order to correct for it. For the purposes of this analysis we will ignore the impact of this false positive.

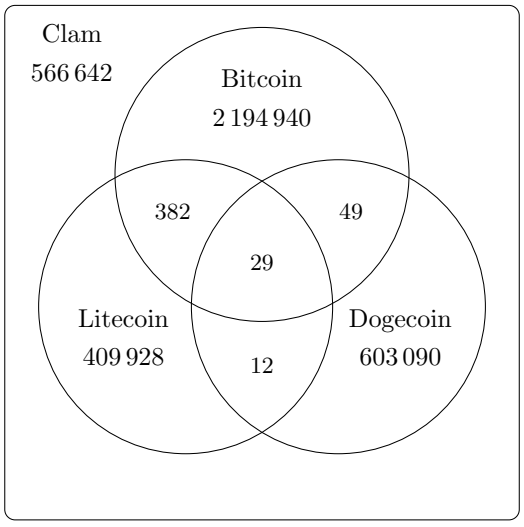
3.2 Address & Address Cluster Novelties

The levels of *address novelty* and *address cluster novelty* for each blockchain are plotted in Figs. 7a to 7d. We define the address novelty of a transaction to be its number of *new addresses* divided by its number of transaction outputs. A new address is one that has not yet been observed in that transaction's blockchain. Nakamoto [13] advised that "a new key pair should be used for each transaction." This is from the perspective of the payees only; if the payer requires additional transaction outputs, say, for change, they should also generate

[2] https://bitcointalk.org/index.php?topic=623147.0



(a) The number of addresses on the Bitcoin, Litecoin and Dogecoin blockchains and how they are shared. For example, 36 209 addresses are shared between the Bitcoin and Litecoin blockchains but are not on the Dogecoin blockchain. This Venn diagram does not consider the Clam blockchain.



(b) The number of addresses on the Clam blockchain and how they are shared with the Bitcoin, Litecoin, and Dogecoin blockchains. In this instance, the universe (the rounded rectangle) is the set of addresses on the Clam blockchain.

Figure 1: The sharing of addresses between the Bitcoin, Litecoin, Dogecoin and Clam blockchains.

a new key pair. For transaction outputs that contain `P2PK` and `P2PKH` scripts, the number of transaction outputs in a transaction is the potential number of new addresses. It can be adjusted for `OP_RETURN` scripts, multi-signature scripts and `P2SH` scripts where the redemption script is known. If everyone followed Nakamoto's advice, the address novelty for every transaction would be 1. However, due to address reuse, this is not the case.

We plot the address novelties for each blockchain using a simple moving average (SMA) that includes the last 1% of transactions (blue lines). To adjust for the variable rate of transactions across the blockchains, we place the ordinal transaction number, rather than time, along the horizontal axis. This compresses periods of low-activity and expands periods of high-activity. We observe that the address novelty in the Dogecoin blockchain (Fig. 7c) is much lower than that in the Bitcoin blockchain (Fig. 7a). The Clam blockchain has a high address novelty during the airdrop but low values elsewhere (Fig. 7d). If addresses are shared between blockchains then the low address novelties for the Dogecoin and Clam blockchains may be weak links.

Similarly, we define the address cluster novelty of a transaction whose transaction inputs reference at least two different addresses to be a Boolean value that is one if the transaction merges two or more trivial address clusters, and zero otherwise. The metric is only defined for transactions whose transaction inputs reference at least two different addresses since only those transactions can result in the merging of address clusters. It assigns a value of one to the transactions that merge two or more trivial address clusters since these may be avoidable: the payer may need to combine several transaction output values. However, any other instance of merging could have been avoided by generating new key pairs. Even in the presence of low address novelties, merge avoidance [8] can produce high address cluster novelties.

We plot the address cluster novelties for each blockchain using an SMA that includes the last 1% of transactions (red lines). We again observe that the address cluster novelties in the Dogecoin and Clam blockchains are much lower than that in the Bitcoin blockchain. We note that even if one blockchain maintains or increases its address and address cluster novelties on an individual basis, it can be adversely impacted by a blockchain with lower novelties with which it shares addresses.

4 The Impact of a Single Blockchain

Does the address clustering in the Clam blockchain provide us with additional information that could improve the address clusterings in the other blockchains? To answer this question, we define a graph called the *co-cluster graph* for a set of blockchains: Each vertex represents an address cluster in a blockchain. Each edge between two vertices represents the maximal set of addresses that are shared between the two corresponding address clusters. If the two address clusters do not share any addresses then there is no edge between the vertices.

Figure 2 illustrates a co-cluster graph for three blockchains. Blockchain A

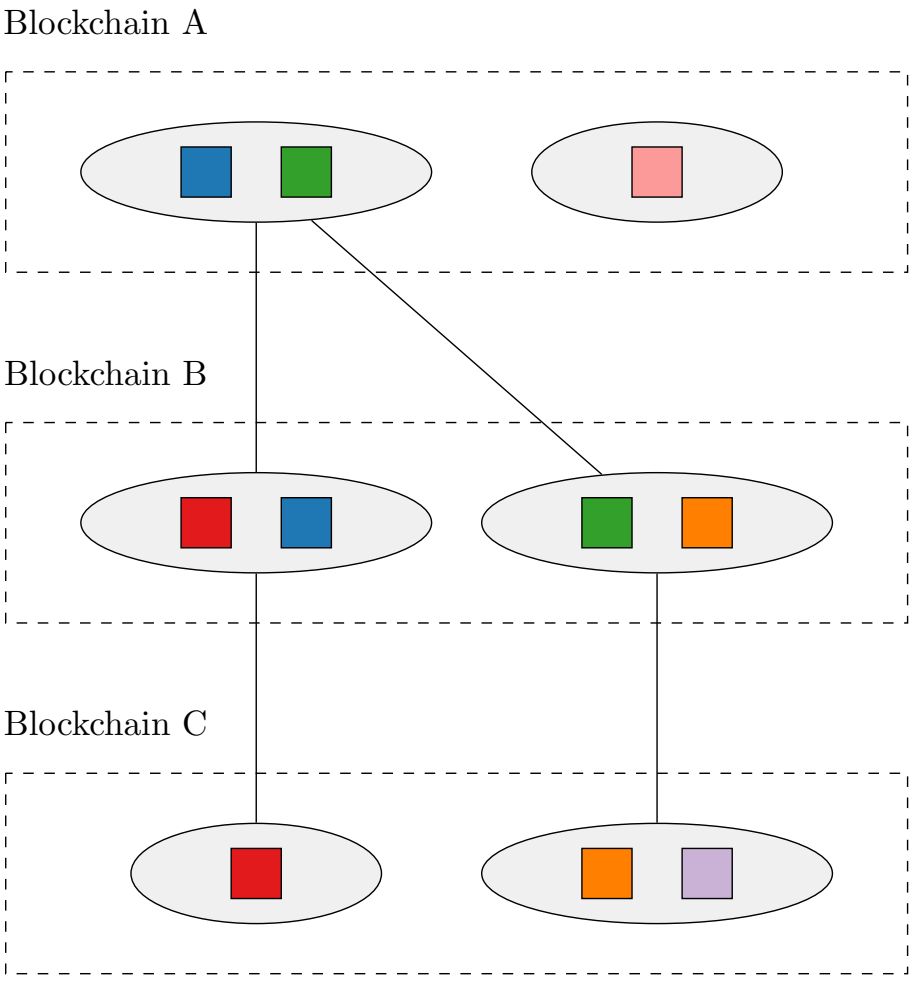


Figure 2: The co-cluster graph for Blockchains A, B and C. The six grey filled ellipses are the vertices and represent the address clusters in each blockchain. The four edges connect vertices whose address clusters share addresses (filled squares with the same colour) between the blockchains.

(the topmost dashed rectangle) contains three addresses (the blue, green and pink filled squares): two of the addresses (blue and green) are in one address cluster and one address (pink) is in another. The address clusters are represented by the grey filled ellipses. Blockchain B contains four addresses: two in one address cluster and two in another. Two of the addresses (blue and green) are shared between Blockchains A and B. Blockchain C contains three addresses: one in its own address cluster and two in another. Two of the addresses (red and orange) are shared between Blockchains B and C. No addresses are shared between Blockchains A and C.

The top-left vertex is incident with two vertices that correspond to address clusters in the same blockchain. This is significant because it shows that two addresses (blue and green) belong to the same address cluster in Blockchain A but different address clusters in Blockchain B. We can therefore merge the two address clusters in Blockchain B. In fact, it also indirectly shows that the two address clusters in Blockchain C can also be merged. We will return to this point at the end of the section.

We can identify additional information that the Clam address clustering provides over a counterpart by identifying the vertices corresponding to address clusters in the Clam blockchain that are incident with at least two vertices corresponding to address clusters in the counterpart's blockchain. In other words, we wish to identify address clusters in the Clam blockchain that contain addresses that are shared with another blockchain but are found in two or more address clusters in that blockchain.

We constructed the co-cluster graph for the Clam and Bitcoin blockchains. We extracted the subgraph that is the maximal induced subgraph on the set of vertices that represent Clam address clusters with a degree of at least two, and their neighbours that represent Bitcoin address clusters. This subgraph has 689 connected components — this number represents the number of 'new and improved' address clusters in the Bitcoin address clustering that are created by the Clam address clustering. It represents a tiny fraction of ($\sim 0.0004\%$) of the original number of Bitcoin address clusters.

Figure 3 is a visualisation of a portion of the extracted subgraph. The connected components can be divided into stars and non-stars. The stars account for the vast majority of the vertices. We hypothesise that stars represent an entity transferring their Clams in one sweeping transaction. The transaction merges the addresses on the Clam blockchain into a single address cluster. However, the same addresses on the Bitcoin blockchain belong to two or more address clusters. The non-stars are more difficult to explain but they can be examined on an individual basis. One possible explanation involves entities who purchased disused private-keys in order to retrieve the corresponding Clam balances.[3]

We repeated this analysis for the remaining two blockchains. The extracted subgraph for the Litecoin blockchain has 179 connected components. This represents a very small fraction ($\sim 0.0011\%$) of the total number of Litecoin address

[3] https://bitcointalk.org/index.php?topic=2247688

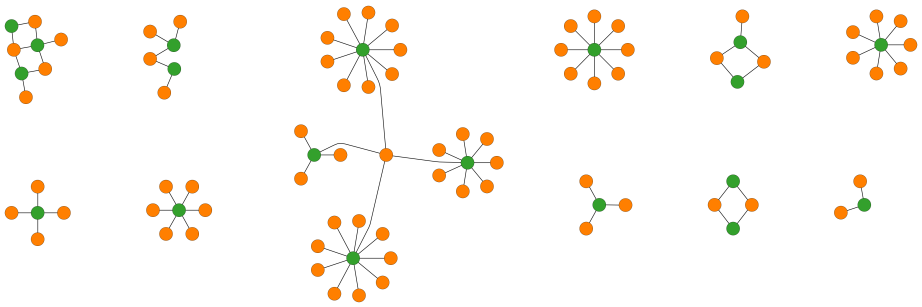


Figure 3: A visualisation of a portion of the subgraph of the co-cluster graph for the Clam and Bitcoin blockchains containing 689 connected components. The green and orange vertices represent address clusters on the Clam and Bitcoin blockchains, respectively.

clusters. In the case of Litecoin, there is one large connected component that contains most of the vertices (2223 out of the 3255 vertices). The extracted subgraph for the Dogecoin blockchain has 1181 connected components. This represents a small fraction ($\sim 0.0060\%$) of the total number of Dogecoin address clusters but it is far higher than that observed in the Bitcoin or Litecoin blockchains.

The analysis above measures the direct impact of the Clam address clustering on the Bitcoin, Litecoin and Dogecoin blockchains. It shows that the number of impacted address clusters is small but they can be easily identified and examined on an individual basis. The analysis can be repeated in the opposite direction, measuring the impact of the Bitcoin, Litecoin and Dogecoin blockchains on the Clam blockchain, by extracting the subgraphs in the opposite direction.

However, this method does not capture all interactions between a set of blockchains. In Fig. 2 the address cluster represented by the top-left vertex indirectly impacts Blockchain C's address clustering via Blockchain B. We can handle this by analysing A's impact on B and then the impact of this new address clustering on C. Alternatively, we can consider all three blockchains simultaneously. We describe this method in the next section.

5 The Impact of Multiple Blockchains

In Section 3 we considered the impact of the Clam address clustering on each blockchain in isolation. However, we can combine blockchains to produce a single ordering of all of their transactions. We can use the timestamps in the block headers to produce an ordering that is approximately temporal. If we have access to the times at which blocks and/or transactions were first broadcast, we can improve the temporal accuracy of the ordering, but this is not necessary for our purposes.

We constructed two combinations: Combination I included all transactions

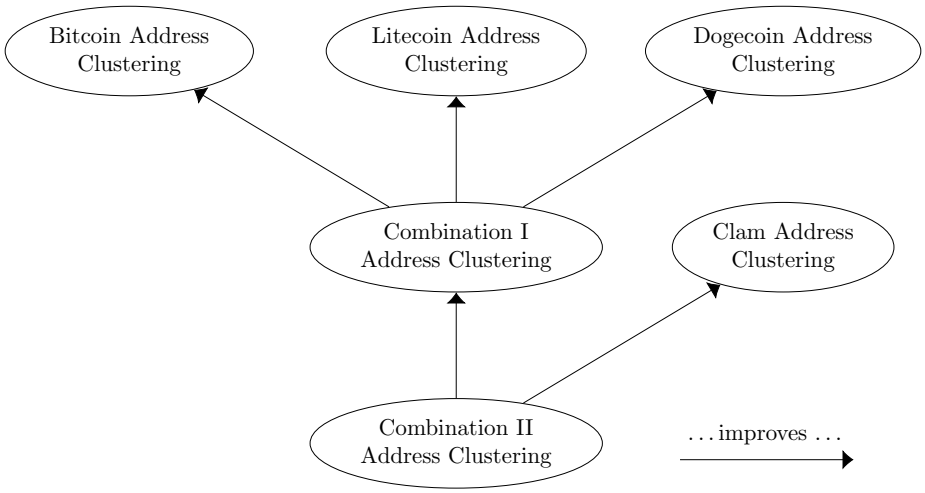


Figure 4: A Hasse Diagram showing the improvement relation between the address clusterings for the individual blockchains and the combinations. The vertices represent the address clusterings and there is a directed edge from a source to a target if the address clustering corresponding to the source is an improvement of the address clustering corresponding to the target.

in the Bitcoin, Litecoin and Dogecoin blockchains; Combination II included all transactions in the Bitcoin, Litecoin, Dogecoin and Clam blockchains. We computed the sizes and coverage of the address clusters (see Figs. 6e and 6f) and the novelties of the addresses and address clusters (see Figs. 7e and 7f) for the combinations as before. The dominant size of the Bitcoin blockchain is evident. The difference between the two histograms is marginal. However, the impact of the Clam airdrop on the address novelties during the six month period following the creation of the Clam genesis block is evident — see the grey rectangles in the two line charts. The airdrop decreases the address novelty in Combination II since the addresses were already observed in the other three blockchains. We can also observe a decrease in the address cluster novelty in Combination II that occurs shortly after a record price for Clam, as denominated in US dollars, was set in early 2018 — see the grey circles in the two line charts. This may indicate a renewed interest in the Clam airdrop and a resulting loss in privacy.

The address clusterings for the individual blockchains and the combinations are related as follows. A partition P_1 is an *improvement* of a partition P_2 if and only if every component of P_2 is a subset of some component of P_1 or is disjoint with every component of P_1. The address clustering for Combination I is an improvement of each of the address clusterings for the Bitcoin, Litecoin and Dogecoin blockchains. Similarly, the address clustering for Combination II is an improvement of both the address clustering for Combination I and the address clustering for the Clam blockchain. It is also an improvement of the

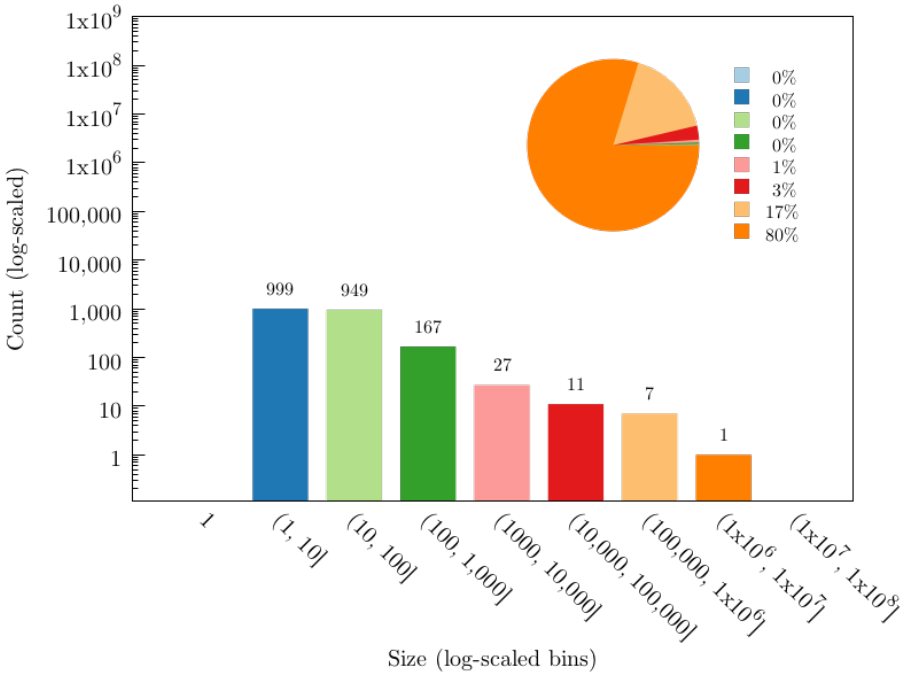


Figure 5: A histogram showing the address cluster size and coverage for the address clusters for Combination II that are the union of two or more address clusters for Combination I.

address clusterings for the remaining blockchains since the relation is transitive. In fact, the relation is a partial order and be visualised using a Hasse Diagram (see Fig. 4).

This relation provides a convenient method of measuring the impact of the address clustering for the Clam blockchain. We can identify the address clusters in Combination II that are the union of two or more address clusters in Combination I. These are the aggregated address clusters across the Bitcoin, Litecoin and Dogecoin blockchains that the address clustering for the Clam blockchain has improved or merged. We enumerated these address clusters: there are 2161 of them. Section 4 identified a total of 2049 address clusters on the individual blockchains that were directly impacted by the Clam address clustering. This new number includes address clusters that are directly *and indirectly* impacted by the Clam address clustering as described at the end of the previous section.

Figure 5 shows the address cluster size and coverage for the 2161 address clusters. Intuitively, this is the difference between the histograms in Figs. 6e and 6f. The Clam address clustering impacts both small and large address clusters. Small address clusters may correspond to individuals making their claims; large address clusters may correspond to centralised services such as

exchanges and mining pools making their claims. For a more thorough analysis, one would need to use address tagging to associate real-world identities with the address clusters. It is important to note, that in both cases, the entities have disclosed information about their address ownership on the Bitcoin, Litecoin and Dogecoin blockchains, exclusively via their activity on the Clam blockchain.

6 Conclusion and Future Work

We have examined a privacy risk in using existing blockchains as a basis for an airdrop's distribution. Specifically, we computed address clusterings for the four blockchains involved in the Clam airdrop: Bitcoin, Litecoin, Dogecoin and Clam. We showed that the sharing of addresses between the blockchains leads to instances where an entity discloses information about their address ownership on one blockchain, exclusively via their activity on another. In the case of the Clam airdrop, we identified ~ 2000 such instances.

The results can help blockchain analysts gather information regarding the beneficiaries of airdrops. The opposing camp, those seeking to hinder blockchain analysts, can also use the results to improve coin control features that consider the implications of using addresses that are shared across blockchains.

Our future work centres around applying this analysis to other airdrops. These include *holder airdrops* similar to the airdrop discussed in this paper, e.g. Bitcoin Private, BitCore and United Bitcoin, and *forked airdrops*, e.g. Bitcoin Cash, Bitcoin Gold and Bitcoin Diamond. A fork creates an airdrop since the entire set of addresses at the point of the fork are shared between the resulting blockchains. A user's activity on one side of the fork can expose information about their address ownership on the other. We would also like to generalise this approach to study airdrops involving Ethereum-based tokens.

References

[1] Elli Androulaki, Ghassan Karame, Marc Roeschlin, Tobias Scherer, and Srdjan Capkun. Evaluating user privacy in Bitcoin. In *Proceedings of the 17th International Conference on Financial Cryptography and Data Security (FC'13)*, pages 34–51, 2013.

[2] Christian Catalini and Catherine Tucker. Seeding the S-Curve? The role of early adopters in diffusion. *SSRN Electronic Journal*, 2016.

[3] Richard Diestel. *Graph Theory*. Springer Graduate Texts in Mathematics. 5 edition, 2017.

[4] Abeer ElBahrawy, Laura Alessandretti, Anne Kandler, Romualdo Pastor-Satorras, and Andrea Baronchelli. Evolutionary dynamics of the cryptocurrency market. *Royal Society Open Science*, 4(11), 2017.

[5] Dmitry Ermilov, Maxim Panov, and Yury Yanovich. Automatic Bitcoin address clustering. In *Proceedings of the 16th IEEE International Conference on Machine Learning and Applications (ICMLA'17)*, pages 461–466, 2017.

[6] Steven Goldfeder, Harry Kalodner, Dillon Reisman, and Arvind Narayanan. When the cookie meets the blockchain: Privacy risks of web payments via cryptocurrencies. *CoRR*, abs/1708.04748, 2017.

[7] Martin Harrigan and Christoph Fretter. The unreasonable effectiveness of address clustering. In *Proceedings of the 13th IEEE International Conference on Advanced and Trusted Computing (ATC'16)*, 2016.

[8] Mike Hearn. Merge avoidance. `https://medium.com/p/7f95a386692f`. Accessed: 2018-08-01.

[9] Harry Kalodner, Steven Goldfeder, Alishah Chator, Malte Möser, and Arvind Narayanan. BlockSci: Design and applications of a blockchain analysis platform. *CoRR*, abs/1709.02489, 2017.

[10] Matthias Lischke and Benjamin Fabian. Analyzing the Bitcoin network: The first four years. *Future Internet*, 8(1):7, 2016.

[11] Gregory Maxwell. CoinJoin: Bitcoin privacy for the real world. `https://bitcointalk.org/index.php?topic=279249`. Accessed: 2018-08-01.

[12] Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon McCoy, Geoffrey M. Voelker, and Stefan Savage. A fistful of Bitcoins: Characterizing payments among men with no names. In *Proceedings of the 2013 Internet Measurement Conference (IMC'13)*, pages 127–140, 2013.

[13] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. 2008.

[14] Jonas Nick. Data-driven de-anonymization in Bitcoin. Master's thesis, ETH Zürich, 8 2015.

[15] Patrick Nieves. Identification of cross-blockchain transactions: A feasibility study. Master's thesis, Technical University of Munich, 2018.

[16] Micha Ober, Stefan Katzenbeisser, and Kay Hamacher. Structure and anonymity of the Bitcoin transaction graph. *Future Internet*, 5(2):237–250, 2013.

[17] Manoj Popuri and Mehmet Gunes. Empirical analysis of crypto currencies. In *Proceedings of the 7th Workshop on Complex Networks*, pages 280–282, 2016.

[18] Fergal Reid and Martin Harrigan. An analysis of anonymity in the Bitcoin system. In Yaniv Altshuler, Yuval Elovici, Armin Cremers, Nadav Aharony, and Alex Pentland, editors, *Security and Privacy in Social Networks*, pages 197–223. Springer New York, 2013.

[19] Jeremy Rubin and Dan Elitzer. Announcing the MIT Bitcoin project. `http://bitcoin.mit.edu/announcing-the-mit-bitcoin-project`. Accessed: 2018-08-01.

[20] Michele Spagnuolo, Federico Maggi, and Stefano Zanero. BitIodine: Extracting intelligence from the Bitcoin network. In *Proceedings of the 18th International Conference on Financial Cryptography and Data Security (FC'14)*, pages 457–468, 2014.

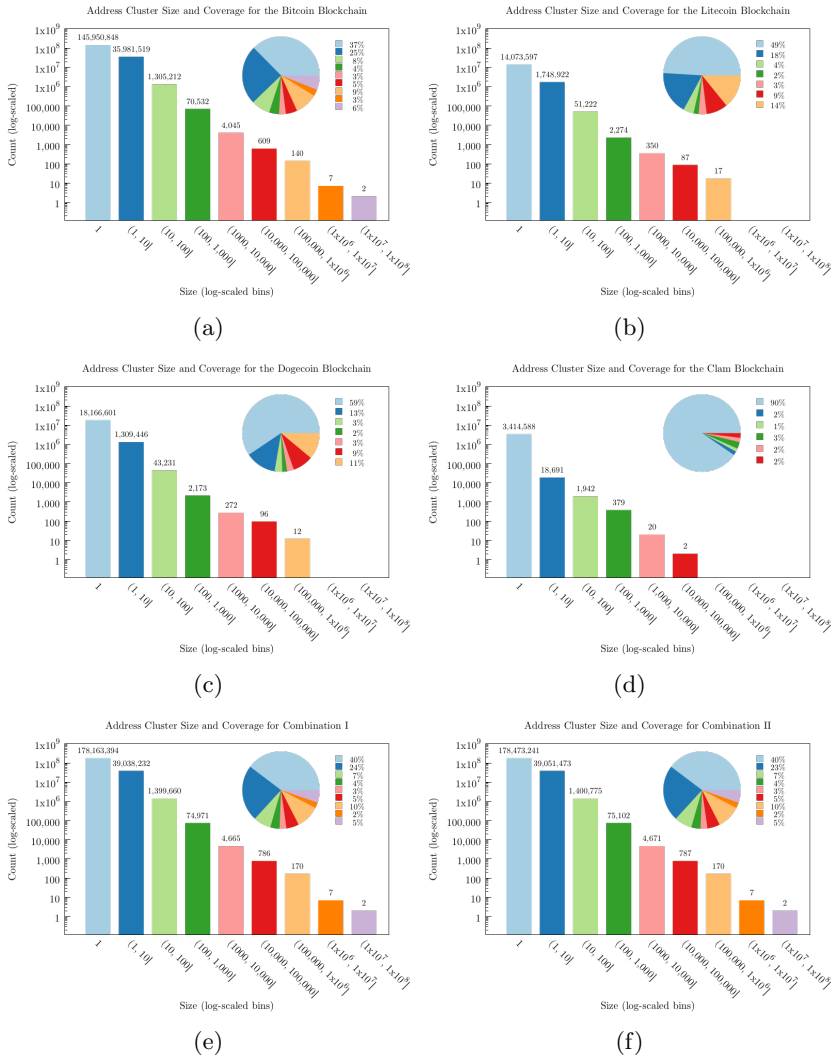


Figure 6: Histograms showing the number of address clusters in each size range for the four blockchains and two combinations. The inset pie charts show the coverage, or total number of addresses in the address clusters, in each size range.

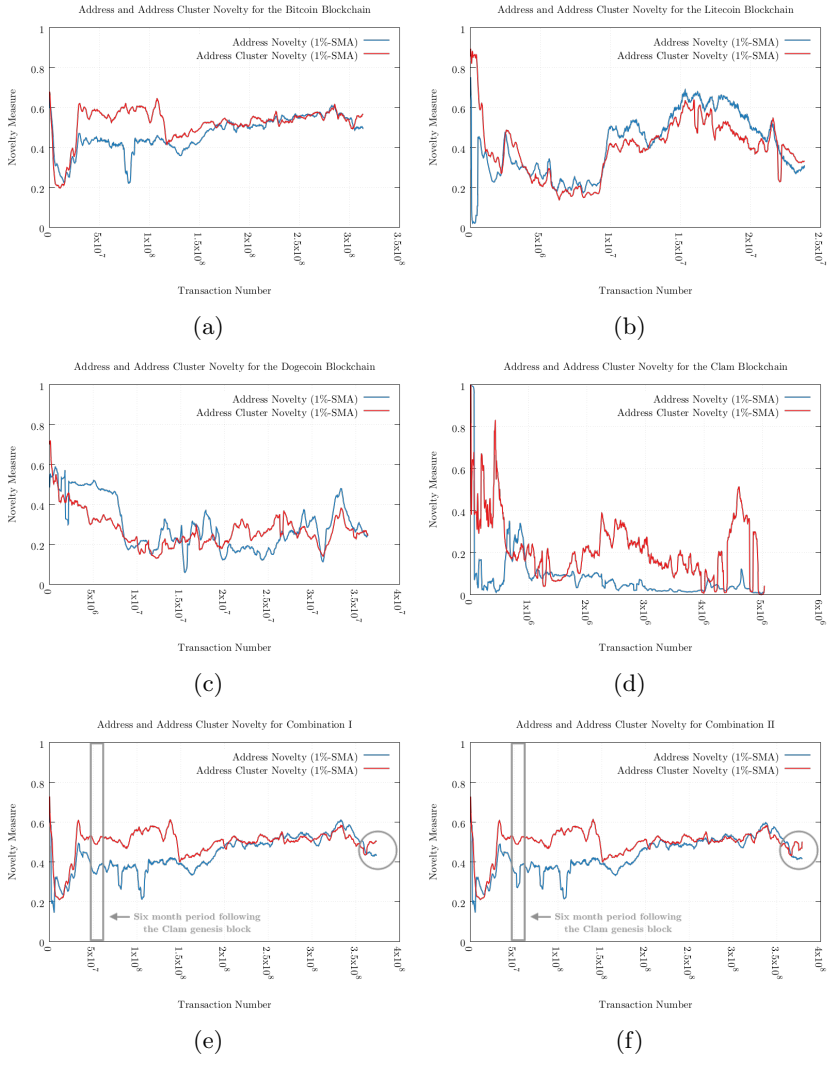


Figure 7: Line charts showing the address and address cluster novelties for the four blockchains and two combinations.