Blockchain analysis

Heuristics

- **"Clustering addresses"** refers to the process of attributing numerous addresses to the same wallet/controlling entity through the use of transaction behaviour heuristics.

- There are a number of factors which go into these heuristics. We will cover some of the well known ones in the following slide. It is important to note however that none of the heuristics are definite. They can be wrong and as investigators it is necessary to corroborate the results.

- Cryptocurrencies focused on principles of self sovereignty recognise blockchain analysis as an attack on the network. **Small minorities within these communities are working to break the heuristics used and minimise the ability to undermine privacy within the protocol**.

- The key point here is that the methods being used to provide Blockchain Forensic Tools is likely to evolve in line with efforts to break the heuristics. It could become more difficult to identify how results are being provided and this makes it important to keep informed on the subject.

Heuristic 1 – Common Input Ownership

Assumption: All inputs in a transaction belong to the same entity as they reside in the same wallet.

- ❑ Vast majority of bitcoin transactions are simple in nature. Very few collaborative transactions.

- ❑ As a result one wallet controlled by one entity will have provided all of the transaction inputs to send funds.

Heuristic 2 – Change address detection:

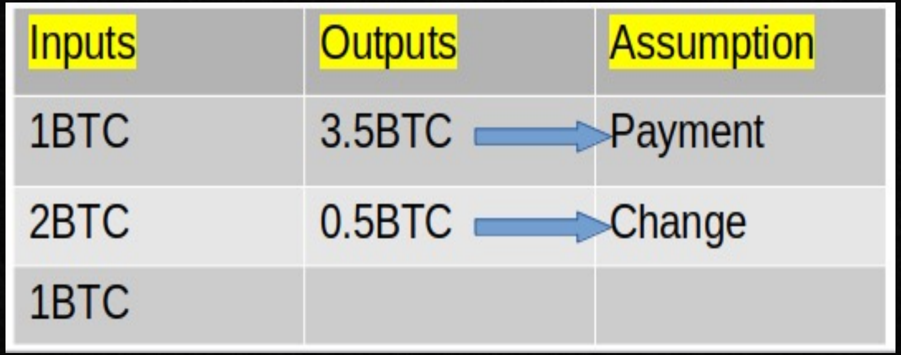
- Change amounts are linked to addresses never previously seen in the blockchain.

- If an output address is the same as an input address it is the change.

- Wallet fingerprinting can be used to detect change outputs because a change output is the one spent with the same wallet fingerprint.

- Round numbers as an output are payments not change.

- If the values of the inputs are more than one of the outputs but less than another, the lower figure output is change (Unnecessary input heuristic) e.g.



Inputs	Outputs	Assumption
1BTC	3.5BTC →	Payment
2BTC	0.5BTC →	Change
1BTC		

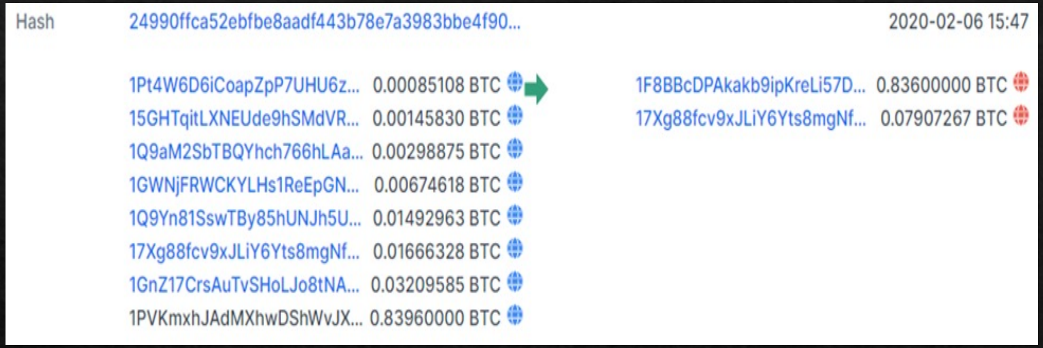
Assumptions:

- **If an output address has been reused it is very likely to be a payment output**, not a change output. This is because change addresses are created automatically by wallet software but payment addresses are manually sent between humans.

- Entities utilise wallet defaults for coin selection and fee payments.

- **Many payment amounts are round numbers**, for example 1 BTC or 0.1 BTC. The leftover change amount would then be a non-round number (e.g. 1.78213974 BTC). This potentially useful for finding the change address. The amount may be a round number in another currency. The amount 2.24159873 BTC isn't round in bitcoin but when converted to USD it may be close to an exact dollar value.

Heuristic examples

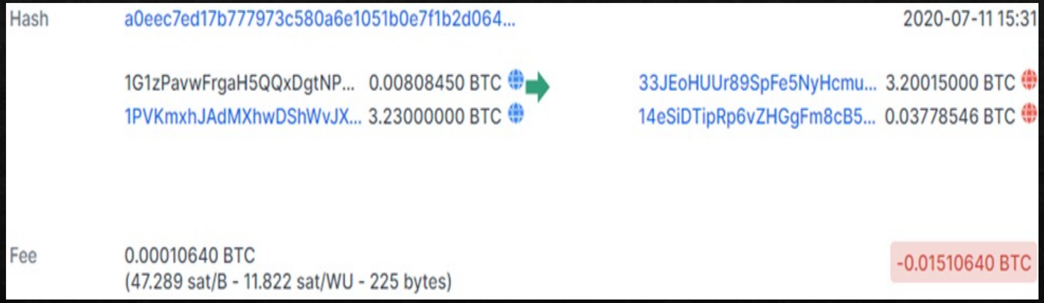


Hash 24990ffca52ebfbe8aadf443b78e7a3983bbe4f90... 2020-02-06 15:47

Input	Amount	Output	Amount
1Pt4W6D6iCoapZpP7UHU6z...	0.00085108 BTC	1F8BBcDPAkakb9ipKreLi57D...	0.83600000 BTC
15GHTqitLXNEUde9hSMdVR...	0.00145830 BTC	17Xg88fcv9xJLiY6Yts8mgNf...	0.07907267 BTC
1Q9aM2SbTBQYhch766hLAa...	0.00298875 BTC		
1GWNjFRWCKYLHs1ReEpGN...	0.00674618 BTC		
1Q9Yn81SswTBy85hUNJh5U...	0.01492963 BTC		
17Xg88fcv9xJLiY6Yts8mgNf...	0.01666328 BTC		
1GnZ17CrsAuTvSHoLJo8tNA...	0.03209585 BTC		
1PVKmxhJAdMXhwDShWvJX...	0.83960000 BTC		

Common input heuristic

Change heuristic: Outputs are different script types.

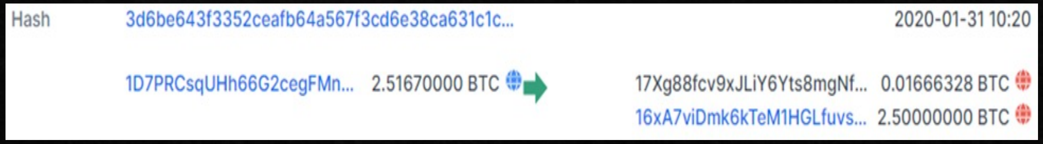


Hash a0eec7ed17b777973c580a6e1051b0e7f1b2d064... 2020-07-11 15:31

Input	Amount	Output	Amount
1G1zPavwFrgaH5QQxDgtNP...	0.00808450 BTC	33JEoHUUr89SpFe5NyHcmu...	3.20015000 BTC
1PVKmxhJAdMXhwDShWvJX...	3.23000000 BTC	14eSiDTipRp6vZHGgFm8cB5...	0.03778546 BTC

Fee 0.00010640 BTC
(47.289 sat/B - 11.822 sat/WU - 225 bytes)

-0.01510640 BTC



Hash 3d6be643f3352ceafb64a567f3cd6e38ca631c1c... 2020-01-31 10:20

Input	Amount	Output	Amount
1D7PRCsqUHh66G2cegFMn...	2.51670000 BTC	17Xg88fcv9xJLiY6Yts8mgNf...	0.01666328 BTC
		16xA7viDmk6kTeM1HGLfuvs...	2.50000000 BTC

Change heuristic: The address 16xA7 was active prior to this transaction. The address 17Xg88 was first active as part of this transaction. Round payment made to 16xA7 address

Private intelligence

Further to these heuristics Blockchain Forensic Tools will **utilise industry intelligence and covert surveillance tactics to attribute entities to addresses and build clusters**. This will mean it is often **opaque as to how an identification or cluster has been developed**.

It is possible to try and find connections. You can manually check for the heuristics and carry out open source research. This may however become unpractical (if significant amounts of data) or turn out to be inconclusive. In these instances it would be worth noting down the efforts made and the negative result. Such process will at least show efforts have been made to understand the intelligence and quantify it's origins.

Practical example

- https://www.reddit.com/r/Electrum/comments/a9x374/my_electrum_just_got_hacked/
- Chain swapping example