jroyer2 Jared Royer

tbarka2 Timur Barkaev

yuxuanz6 Yuxuan Zhou

Our clustering algorithm was very straightforward. We have 2 primary maps, one has user id as a key and a list of addresses as a value, the other has an address as a key and a user id as a value. We iterate through all the transactions, for each transaction, we store all input addresses, and the first address that we have already encountered. If such an address exists, we join the inputs with the already known joint addresses, otherwise, we create a new user id with these addresses. If we encounter known addresses corresponding to different users, we join those users together as well.

The assumption of joint control is reasonable if you know what you’re looking for in advance. It is not unreasonable, but we believe that the number of false positives and false negatives could be very significant if this method was applied to a different dataset. The false positives could be easily encountered if, for instance, several users decided to make a joint purchase, putting both of their addresses as inputs to a transaction. Joint control would tag them as the same user, which might be reasonable for that single transaction, but it would also link all their future purchases together. We think that false negatives would be almost impossible, since every address in the same transaction’s input would map to the same user. The only way some addresses would get left out is if it the address only receives Bitcoin (is only in transaction outputs), but it happens to be owned by a user who also owns other addresses.

The bitcoin address of the FBI is 1FfmbHfnpaZjKFvyi1okTjJJusN455paPH. We identified it by first clustering the addresses and then identifying the cluster that received the most bitcoins on the date of seizure. Checking online, <https://99bitcoins.com/price-chart-history/> said the FBI took 144,000 from Ross, which is the number that we had them receiving.

The Ross Ulbricht addresses are 19uruykinejqWxnMsusvw4Ftkq4rxJEHjK, 1Ap873XpqgPSJLfLELJ5SCKWZpGthnGuTY, 18SKdJHHGLFrD8KNgFj4Mb9tC3JNGmsAR. After clustering, we found the user that sent the most bitcoin that day and assumed it was Ross Ulbricht. To confirm, we used one of the addresses that we found (the first listed here), went back to transactions.txt, and searched for it. That and 6 other addresses were inputs to a transaction and the only output was to the FBI address. Hence, from joint-control, it is a reasonable assumption that this was Ulbricht, especially since one of the “in’s” was for 12600000000 Satoshi (126,000 BTC).