Open Wallet Explorer

Degree programme: BSc in Computer Science | Specialisation: Digital Business Systems
Thesis advisor: Prof. Dr. Kai Brünnler
Expert: Dr. Andreas Spichiger (Eidg. Finanzdepartement)

Open source variant of the Wallet Explorer created using Python and the BlockSci platform for clustering.

Wallet Explorer

A bitcoin wallet consists of as many addresses as you wish. There are multiple ways to manage a wallet and it is completely up to you how you do it. From an observer point of view it would be perfect to see which addresses belong to the same identity, but the bitcoin protocol does not allow so directly.

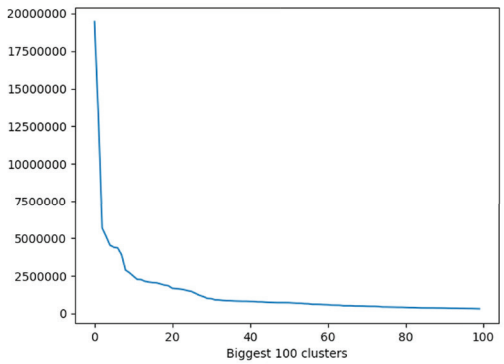
The main goal of this study is to create an open source variant of the Wallet Explorer. The creator of the old site has decided against public access of his work. While still working and providing results, the site has not been updated for a long time and feels out of date. Furthermore, it is questionable whether or not results will hold against a case in court. The Open Wallet Explorer is able to find addresses belonging to one wallet, therefore the same identity, with providing it with one address. The API also provides a way to see all transactions made of an address. It is easily expandable to include more meta-data or other analysis work. The written program relies on a big server with at least 60 GB of ram, as it calculates wallet clusters on the fly.

Lessons Learned

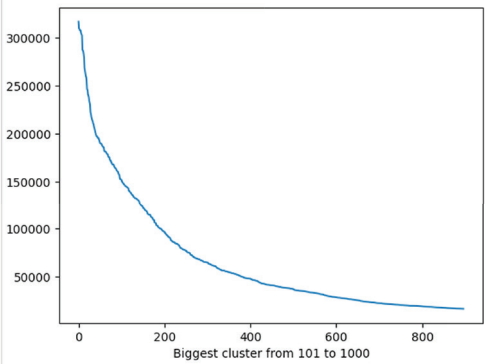
BlockSci was written in C++ and is optimized for speed. To help researchers analyse the blockchain faster and easier it has a python interface as well. Most setups consist of a jupyter Notebook on which multiple people have access on. In this work we wrote a REST API which accesses the BlockSci library directly via Python. First, we wrote a script which saved the result in a database for later analysis. The script was too slow: it took us about nine hours, while skipping clusters bigger than 30'000 addresses, to parse the Bitcoin blockchain. We spent a lot of time optimizing this code, which could have been used on the live version (current). Even now, clustering a wallet with over 500'000 addresses can take up to 5 minutes. Multiple issues on the main BlockSci repository complain about the slow clustering process or other slow methods using the python interface. If we could restart the project from scratch, we would definitely look into the C++ version.

To make the live version workable we developed the pre-downloader which downloads the biggest 1000 clusters. All clusters can now be queried within a reasonable time.

Nicolo Claudio Singer
nicolo.singer@gmail.com



Sizes of the biggest 100 clusters.



Sizes of the 101 to 1000 biggest clusters.