Readme

This is the dataset accompanying the paper *If You Like Me, Please Don't "Like" Me: Inferring Vendor Bitcoin Addresses From Positive Reviews*, as submitted to PETS 2022. The uncompressed dataset has a total size of approx. 50 MB.

Please note: Due to privacy concerns, the publicly available data has been subset to only include the attributes that are required for replication. The number of examples (i.e. the number of vendors and products) is the same as in the original data. Potentially identifying information like vendor usernames, product IDs and PGP keys are either substituted by pseudonyms or salted and hashed using SHA-256. For the same reason, we do not provide any source code. We may share additional data as well as the source code with other researchers for scientific purposes. If you need access, please send an E-Mail to one of the authors and make sure to include:

- Your current affiliation.
- What information you need and what research question you intend to answer with it.
- A description of how you plan to ensure the privacy of the individuals included in the dataset.
- The potential ethical implications of your work and how they will be addressed.

The data is stored as comma separated values in a total of seven individual files. Each of these files corresponds to a table in a relational database. Please find a description of the individual files and values below. Missing values are coded as NA.

Cannazon_Vendors.csv

Contains information on 246 Cannazon vendors. The data has been scraped from the respective vendor profile pages.

vendor A randomly assigned, pseudonymous substitute for the username in the form *Cannazon Vendor 123*. Serves as a unique identifier within the Cannazon datasets.

orders_finalized An integer, showing the number of orders the vendor had finalized at the time of scraping.

Cannazon_Products.csv

Includes information on 6,598 products sold on Cannazon.

salted_product_hash A pseudonymous substitute for the product id. Calculated as $SHA256(product\_id||1024\_char\_random\_string)$. Serves as a unique identifier within the Cannazon datasets.

escrow The types of escrow available for the product. One of *Only Escrow*, *Escrow or Finalize Early*, or *Only Finalize Early*.

btc_accepted/xmr_accepted Dummy variables indicating whether the vendor accepts Bitcoin (BTC) and/or Monero (XMR) as payment for the product.

Cannazon_Reviews.csv

Contains data on 351,769 vendor reviews scraped from Cannazon.

vendor ID of the vendor the review is referring to (see Cannazon_Vendors.csv).

market The name of the market the review was published on. Always Cannazon.

salted_product_hash ID of the product the review is referring to (see Cannazon_Products.csv).

positive/neutral/negative Dummy variables indicating whether the review was positive, neutral or negative. Mutually exclusive.

time Timestamp of the review, formatted YYYY-MM-DD.

price_eur Order volume in Euro as stated in the review.

Cannazon_Keys.csv

Information on the PGP keys provided by the Cannazon vendors.

vendor ID of the vendor the key belongs to (see Cannazon_Vendors.csv).

salted_key_hash A pseudonymous substitute for the key hash. Calculated as $SHA256(key\_hash||1024\_char\_random\_string)$.

Cryptonia_Vendors.csv

Contains information on 745 Cryptonia vendors. The data has been scraped from the respective vendor profile pages.

vendor A randomly assigned, pseudonymous substitute for the username in the form *Cryptonia Vendor 123*. Serves as a unique identifier within the Cryptonia datasets.

orders_finalized An integer, showing the number of orders the vendor had finalized at the time of scraping.

finalize_early A dummy variable showing if a vendor is allowed to accept advance payments.

Cryptonia_Reviews.csv

Contains data on all 28,966 vendor reviews scraped from Cryptonia.

vendor ID of the vendor the review is referring to (see Cryptonia_Vendors.csv).

market The name of the market the review was published on. Always Cryptonia.

salted_product_hash ID of the product the review is referring to (see Cryptonia_Products.csv).

positive/neutral/negative Dummy variables indicating whether the review was positive, neutral or negative. Mutually exclusive.

time Timestamp of the review, formatted YYYY-MM-DD.

price_btc Order volume in Bitcoin as stated in the review. Can be zero if the review refers to an order paid for in Monero.

Cryptonia_Keys.csv

Information on the PGP keys provided by the Cryptonia vendors.

vendor ID of the vendor the key belongs to (see Cryptonia_Vendors.csv).

salted_key_hash A pseudonymous substitute for the key hash. Calculated as $SHA256(key\_hash||1024\_char\_random\_string)$.