Bitcoin transaction tracing and purchasing behavior characterization of online anonymous marketplaces using side channels

Submitted in partial fulfillment of the requirements for

the degree of

Master of Science

in

Information Technology - Information Security

Eugene Lemuel R. Garcia

B.S., Electronics and Communications Engineering,
University of the Philippines - Diliman

Carnegie Mellon University
Pittsburgh, PA

December, 2017

Acknowledgements

First, I would like to thank my thesis advisor, Dr. Nicolas Christin. From teaching the class that got me interested in the field, to accepting to be my advisor, to providing useful advice and directing me as well to other good resources, this work[1] definitely would not have been possible without his gracious help and support.

I am grateful to Kyle Soska, for the practical guidance he provided while deciding on the thesis topic to pursue, for the wealth of suggestions and direction particularly in the machine learning side of things, and for the tireless support through it all.

My appreciation goes out to Dr. Kilho Shin for the insightful questions and advice on the clustering aspect of this project. It has been a pleasure working with him in the past and I look forward to doing research with him again in Japan.

I would like to express my gratitude to Dr. Sarah Meiklejohn, for her group's excellent research and for taking time to explain the code she has generously shared.

I want to acknowledge my groupmates in cryptocurrencies class, Baptiste, Satin and Glen. It was great working with them in tracing Bitcoin transactions early on.

Many thanks goes to my friends in our small group at Northway Oakland for the prayers, laughter, reprieve from schoolwork, and all the random things that we do.

I am especially indebted to my family, for all the love and support I have been showered with regardless of being halfway around the world from them. Definitely would not have survived here without them backing me up.

Last but not the least, I want to thank God for creating a universe that could be explored, and for giving us minds to explore His creation and make our own inventions along the way.

[1] This study is self-funded.

Abstract

We investigate to which degree one could trace Bitcoin transactions and characterize purchasing behavior of online anonymous marketplaces by exploiting side channels. Using a list of addresses found by the FBI on Silk Road servers, and information on the marketplace's official guides, we infer the role played by each address in the list and classify them based on heuristics. We then attempt to trace Bitcoin transactions and show that the anonymity set size is greatly reduced using product review data and the address classification performed on the previous step. Finally, using clustering techniques based on transaction graph analysis, we assign addresses into user wallets, then group these wallets together based on spending patterns, to be able to characterize purchasing behavior.

Table of Contents

List of Tables

List of Figures

1

Introduction

Bitcoin is often viewed as an anonymous digital currency because creating an address which serves as an account does not require one's name. The creation process is also free and therefore a person can create as many addresses as he wants, strengthening the notion of untraceability. For this reason, online anonymous marketplaces—websites hidden through Tor such as the now defunct Silk Road and AlphaBay—used Bitcoin, in order to protect their users who mostly dealt in drugs and other illegal goods.

Side channels however, are available. These sources of external information could provide a clue, a means to track or narrow down the list of candidates that match a certain marketplace transaction. Indeed, previous efforts using side channels have been effective in characterizing the vendor side of the equation using product review information available from these websites. Additionally, with every law enforcement operation that successfully shuts down an online anonymous marketplace, data on the website are released which could potentially serve as new side channels.

This paper presents an investigation on the extent to which, 1) Bitcoin transactions could be traced, and 2) purchasing behavior could be characterized on online

anonymous marketplaces through the use of information from side channels. Previous studies either focused on clustering Bitcoin addresses based on ownership [13] or on characterizing the vendor side of the marketplace ecosystem [9, 17].

Starting from a list of addresses on Silk Road servers found by the FBI, along with information from the marketplace's forum and official guides, we infer the possible roles that an address could play, then classify the addresses according to heuristics. Next, we attempt to trace Bitcoin transactions based on product review data and the categorized addresses from the previous step. Finally, we cluster addresses into user wallets based on transaction graph analysis, then group the wallets together according to spending patterns, as a way to characterize purchasing behavior of users in the marketplace.

We find that using the product review data for tracing Bitcoin transactions reduces the anonymity set size by a factor of around 400 to 4,000. We also observed that using the address list as well further reduces the size by a factor of 20. These describe the degradation in anonymity of transactions when given side channel information.

For characterization, we find that users could be categorized into three main groups: high spending users, one-time customers, and regular users. These three behavior groups correspond to 0.2%, 59%, and 33% of users in the market, contributing 22%, 31%, and 26% of the market volume respectively.[1] These information provide policy makers insight into the marketplace which could aid in decision making.

[1] The remaining 7% of users and 21% of market volume belong to the outliers among the users, i.e. those whose behaviors are not close enough to be clustered with members of the other groups.

2

Background

The heuristics that we use to trace transactions and characterize purchasing behavior are based on how Bitcoin is used by an online anonymous marketplace, so we first discuss both Bitcoin and online anonymous marketplaces here. Additionally, side channels are used extensively in the analysis, so we further describe the concept, and present a list of the channels we use in the investigation.

2.1 Bitcoin

Bitcoin is the world's first decentralized digital currency [7]. It was created by a person (or groups of persons) under the name Satoshi Nakamoto in 2008 [14]. Its popularity and value has grown through the years, and in late 2017 for the first time the value of one bitcoin surpassed 10,000 US dollars, an impressive amount given that its value at the start of that year was less than a thousand US dollars.

To use Bitcoin, one first needs to generate a Bitcoin address, typically through one of the readily available desktop or mobile applications. Once a person has an address, he can receive bitcoin through that address, or send bitcoin from that address to another person's address.

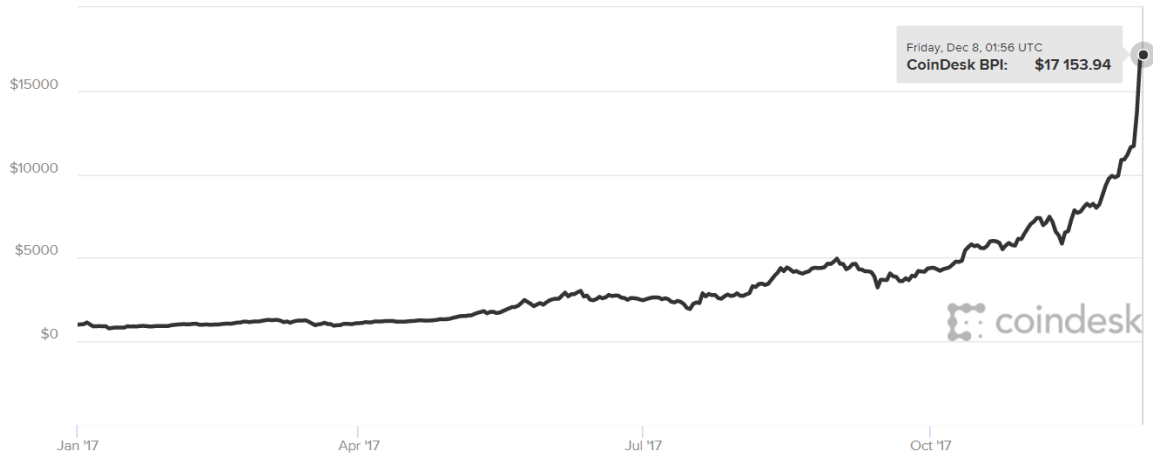


Figure 2.1: Bitcoin growth in 2017. Its value reached 10,000 US dollars for the first time in late November of that year [10].

Address generation is private, and free. There is no need to reveal one's name or register to a central authority to create a Bitcoin address. There is also no significant cost to this action, and as such, typically users have more than one address, either created manually, or automatically created by the application being used.

Transaction records from one Bitcoin address to another Bitcoin address however, are completely public. These are stored in a distributed ledger called a blockchain.

2.2 Online Anonymous Marketplaces

Online anonymous markets, also known as darknet markets or cryptomarkets, are websites similar to Amazon.com that operate on darknets such as Tor or I2P [12]. These are primarily used for drugs, weapons, and other illegal products. Examples of these markets include Silk Road, Agora, AlphaBay, among many others.

Transactions with online anonymous markets are typically facilitated using Bitcoin because of the expectation of privacy that the digital currency brings. Given the focus on privacy through the use of Tor and Bitcoin, to prevent this privacy from

being abused by would-be scammers, users of the marketplace are required to always leave a product review. Similar to clearnet marketplaces, these reviews contain a rating along with comments from buyer who left the review, and are visible to users of the website to guide them towards vendors with superior goods and/or vendor practices.

Online anonymous markets, given the often illegal nature of goods being transacted, are frequently targeted by international law enforcement. For example, in October 2013, the Federal Bureau of Investigation (FBI) shut down Silk Road and arrested its operator. In July 2017, AlphaBay was shut down as well by a multinational law enforcement operation.

2.3 Side Channels

Side channels are sources of external information not directly provided by the system under observation. The external information that side channels provide could be used or exploited in a way that would reveal more information about the system, without the side channel being maliciously designed to do so.

The concept may be more clearly illustrated using the following example: Let us say you are in a coffee shop, and there is this man that bought coffee using Bitcoin. If you are tasked to get this person's Bitcoin address you would be faced with a difficult endeavor, as there are a lot of Bitcoin transactions occurring at any point in time. Let us say however, you see by the counter the Bitcoin address of the coffee shop, along with the note that this is the address that they always use. In such a case, the number of candidate addresses that belong to the man who bought coffee would now be reduced, because we could eliminate those that do not send to that destination address. Additionally, let us say a friend you are with on the coffee shop shares in a casual remark his observation to you that this man always buys coffee every weekday in this same shop, at around this time. This information would effectively reduce

the number of candidate addresses further, possibly allowing you to pinpoint exactly the address in question. Notice that in both cases, the side channels represented by the note on the counter and your friend sharing observations are not explicitly designed to maliciously undermine the man's Bitcoin address privacy, the additional information they provide just happens to do so unintentionally.

In this paper, the side channels we use are the user-generated product reviews, the Silk Road address list from the FBI, the Silk Road forum website, and also the marketplace's official buyer's guide and seller's guide.

The product review data comes previous research by Christin [9] and Soska [17]. These contain the date that the review was posted, the value of the product in US dollars, vendor name, and comments left by the user regarding the product/service.

The Silk Road address list is a large list of addresses, 2.1 million in all, found by the FBI on Silk Road servers [11]. It is completely unlabeled, and no context was provided as to what the addresses were used for.

The Silk Road forum website is no longer operational, however an archive is available for download [1]. Saved copies of the official buyer's guide and seller's guide were used as evidence in the case following Silk Road's take-down. Likewise, these are also available for download from the same website [2].

3

Methodology

We begin by analyzing the Silk Road address list and classifying the addresses found within. The information we glean from this step would prove useful in both transaction tracing and purchasing behavior characterization.

3.1 Silk Road Addresses Classification

Given the lack of information about the nature of the addresses within the list, the initial goal is to reidentify the structure of the website/system from which the addresses were taken. From this analysis, we create heuristics to use for classification of the addresses. We then validate the resulting classification by comparing known figures about the marketplace with new data that could be derived from the address classification.

3.1.1 System Structure Reidentification

As stated previously, little is known about the addresses contained within the list aside from them being associated to Silk Road servers. At the start we were unaware of who had control over these addresses, what their function is in the system, and

whether the list is complete and consistent to known data.

To discover about these aspects, we first refer to the blockchain and Silk Road forum posts. From the blockchain, we observe that the addresses in the list are included among those that sent bitcoin to 1F1tAaz5x1HUXrCNLbtMDqcw6o5GNn4xqX, an FBI-controlled address that was used to hold the seized Silk Road funds [6]. This indicates that after the Silk Road was shut down, FBI had control over these addresses, implying that prior to the take-down Silk Road was the one controlling these addresses. Additionally, in the Silk Road forum, some users have stated that as they monitor the blockchain for transactions to and from their deposit address (which were also found in the list), they observe that bitcoin is being moved from these addresses even without them making a purchase or withdrawal, supporting the finding that it was the website itself that had control over the addresses.

To learn about the addresses' function, we turn to the Silk Road Buyer's Guide. In the said guide, it was stated that users are presented with a deposit address, which they are instructed to deposit bitcoin into prior to making purchases in the website. It was also made known in the guide that a tumbler system is being used, which "sends all payments through a complex, semi-random series of dummy transactions", implying that the system makes use of automated tumbler addresses that serve this purpose.

3.1.2 Classification

For user deposit addresses, we expect all transaction inputs to come from addresses that are outside Silk Road, i.e. those not within the address list. This is because users would be loading their accounts from either exchange services, or their personal Bitcoin addresses. On the other hand, for automated tumbler addresses, we presume that all inputs to these would come from within the address list, consistent to its declared behavior of simply mixing things up in the system.

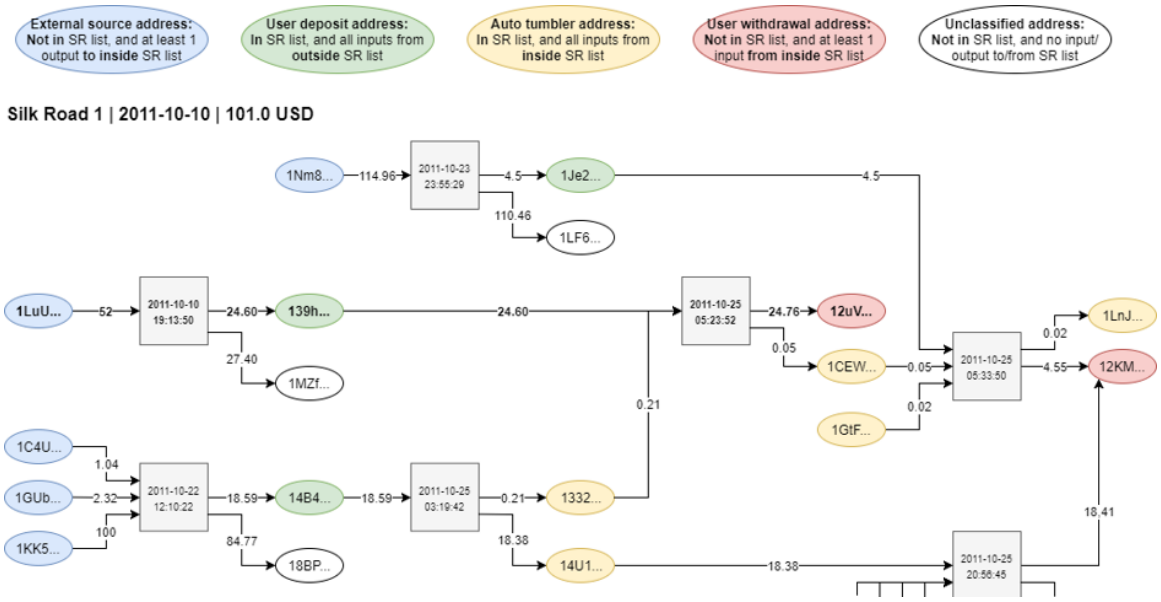


Figure 3.1: Sample address classification of addresses involved in a 101 USD transaction in Silk Road on 2011-10-10, including addresses from adjacent transactions.

Using these expectations, we classify each address in the list as either a deposit address or a tumbler address.

Additionally, we also classify addresses outside the list that are encountered along the way into either an external source address or a user withdrawal address. External source addresses are those that pour money into Silk Road, i.e. those that have at least one transaction output into an address in the list. User withdrawal addresses are the opposite, they are those that take money from Silk Road, i.e. those that have at least one transaction input coming from an address in the list.

Figure 3.1 shows a sample of the address classification. Addresses 1LuU..., 1C4U..., 1GUb..., 1KK5..., and 1Nm8..., being the first point of contact before entering Silk Road, we designate as external source addresses. Those addresses that they deposit to (139h..., 14B4..., 1JE2...) we classify as user deposit addresses. Addresses that have all transactions coming from within the Silk Road system (1332...,

14U1..., 1CEW..., 1GtF..., 1LnJ...) we assign as tumbler addresses. And those that receive money coming outside of Silk Road (12uV..., 12KM...) we categorize as user withdrawal addresses.

3.1.3 Validation

Validation of the classification can be performed by comparing known data from previous research with new information obtained given the addresses' classification.

Silk Road seized coins

It was reported in late 2013 that the FBI seized 29,655 bitcoins from Silk Road [3]. Theoretically, the maximum amount of bitcoins that can be seized from Silk Road is total amount of bitcoins in its addresses at the date of seizure, October 2, 2013. This is equivalent to the total number of bitcoins that were deposited into Silk Road, minus the total number of bitcoins that were withdrawn, on the said time frame. Relating this value with the newly classified addresses, it would be equal to the total bitcoin amount that has entered all user deposit addresses, minus the total bitcoin amount that exited from Silk Road into user withdrawal addresses. We could therefore express the maximum amount that could be seized as:

$$S_{max} = \sum_{n=1}^{N} \sum_{i=1}^{I_n} d(n,i) - \sum_{m=1}^{M} \sum_{j=1}^{J_m} w(m,j) \tag{3.1}$$

where N is the total number of user deposit addresses, I_n is the total number of input transactions to address n, and $d(n,i)$ is the amount of bitcoins deposited to address n on input transaction i. In the same way, M is the total number of user withdrawal addresses, J_m is the total number of output transactions to address m from a Silk Road address, and $w(m,j)$ is the amount of bitcoins withdrawn from Silk Road to address m on output transaction j.

Daily volume estimates

Because product reviews are mandatory, Christin [9] and Soska [17] have shown that daily volume estimates could be acquired by aggregating the US dollar values of each review for a given date and plotting them over time.

Another way to calculate these estimates using the addresses that have undergone classification is to sum all the inputs to all user deposit addresses for each day. More specifically, we could express volume as a function of time as:

$$V(t) = \sum_{n=1}^{N} d(n, t) \tag{3.2}$$

where again N is the total number of user deposit addresses, and $d(n, t)$ is the amount of bitcoins deposited to address n on day t.

A good classification would yield daily numbers that match the transaction volume pattern found in previous research. Additionally, we expect these numbers to always be greater than the figures from the product review data, given that the latter was described to be a lower bound because of the possibility of users depositing then withdrawing without making a purchase, and the chance that the review was for a purchase of more than one product.

3.2 Transaction Tracing

Given evidence of a marketplace transaction, the goal is to find the corresponding blockchain transactions for it. To be able to do so, we first investigate the relationship between the marketplace and the blockchain, specifically, what blockchain transactions would be observable for a given marketplace action by one of its users. Next, we perform a search on the blockchain based on these relationships. We then take note of the limitations of the approach and discuss a method of validating the technique.

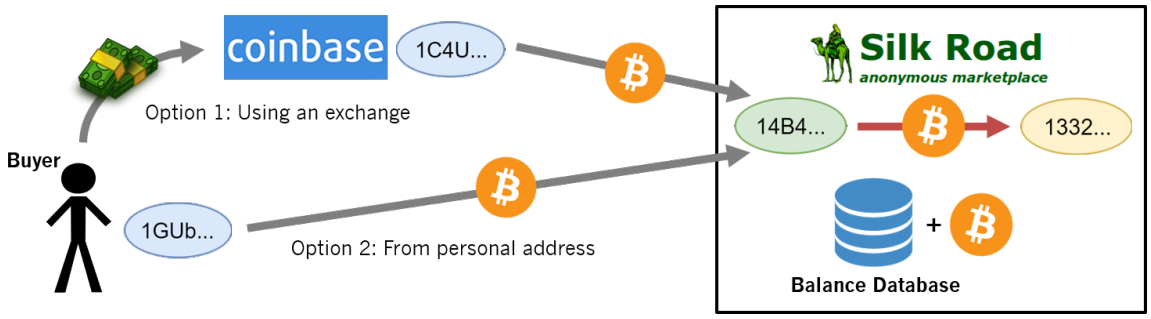


Figure 3.2: Blockchain activity when purchasing. The user has two options, either to use an exchange, or to load directly from his own address.

3.2.1 Marketplace-Blockchain Relationship

To determine what kind of blockchain transactions could be detected based on the occurrence of a marketplace transaction, we refer to both the Silk Road Buyer's Guide and Seller's Guide. Forum posts also provide additional information on the process.

Purchasing from the market

The buyer's guide and forum posts states that prior to purchasing, two options are available to load bitcoin into one's deposit address. Figure 3.2 presents the two options. The buyer could either 1) pay an exchange to transfer the corresponding amount of bitcoin to his user deposit address, or 2) transfer directly from his personal bitcoin address to his Silk Road deposit address.

On observant member of the forums also notes that after the transfer is made, bitcoins from the deposit addresses are automatically transferred to another address (presumably one of Silk Road's tumbler addresses) even without the user initiating a purchase or a withdrawal. The user also notes that the balance reflected in his account page remains the same, even if the bitcoins have been moved. This indicates that Silk Road has an off-blockchain mechanism of keeping track of each user's

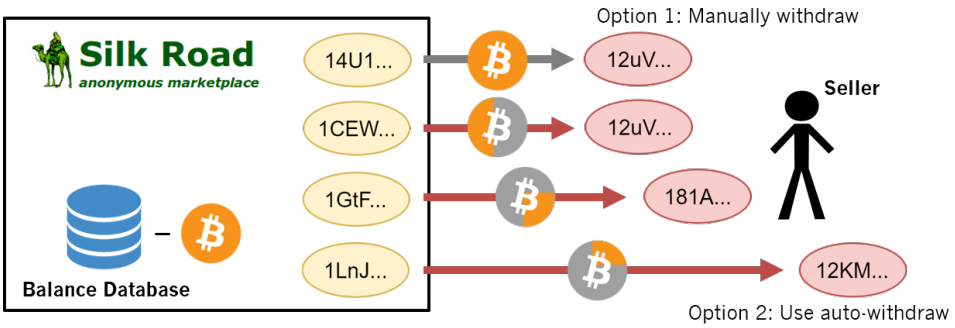


Figure 3.3: Blockchain activity when cashing out. The user also has two options, either to manually withdraw to one address, or to use the auto-withdrawal feature.

balance.

From a transaction tracing standpoint, we can see that purchases on the market has the possibility of being detected on the blockchain if the user deposits in one transaction an amount close to the value of the item being purchased. There are limitations to this approach of course, which will be discussed towards the end of this section.

Withdrawing money from the market

For cashing out, the seller's guide and forum posts state that there are also two options, both of which are presented on Figure 3.3. A user who wishes to convert his Silk Road balance into bitcoins he could spend in his own wallet could either 1) manually open his account page and direct the system to make a one-time transfer to an address he provides, or 2) enable the auto-withdrawal feature and provide three withdrawal addresses. What the auto-withdrawal feature does is to wait until the user's balance reaches 3 bitcoin or above, then once it reaches this threshold, a "withdrawal will then be made in 3 random sized chunks to these three addresses with a small, random delay in between them."

A forum moderator mentioned that most vendors use the auto-withdrawal fea-

ture. From a transaction tracing point of view, this makes correlating blockchain activity to vendor earnings difficult, as there are a large number of ways to group together separate auto-withdrawal transactions, with each possibly being as plausible as several other configurations. For this paper, we focus on the purchasing side of the equation.

3.2.2 Candidate Transaction Search

Now that we know what we may detect in the blockchain as a reflection of marketplace purchases, we could perform a search through the blockchain for transactions with characteristics matching what we would expect from a specific marketplace transaction. To measure the effectiveness of the technique, the metric we will use is the size of the "anonymity set". For the purposes of this paper, the term "anonymity set" pertains to the set of blockchain transactions consistent with an observed marketplace transaction. A smaller size of this set indicates that a marketplace transaction is less anonymous than another transaction with a larger anonymity set size.

To filter out the transactions in the blockchain, we search around the date and expected bitcoin value of the marketplace transaction. Additionally, we constrain the matches to only be those whose destination address is one of the user deposit addresses according to the classification performed earlier.

Date range

The range of dates we use for the search is either 1) up to two weeks before, or 2) up to one day before. The range we use will depend on the estimated maximum time difference between the blockchain transaction and the marketplace review. The reason that there are two cases for the estimated maximum time difference is because there are two main ways that purchases are performed on the market: 1) the "normal" process, and 2) the "early finalization" process.

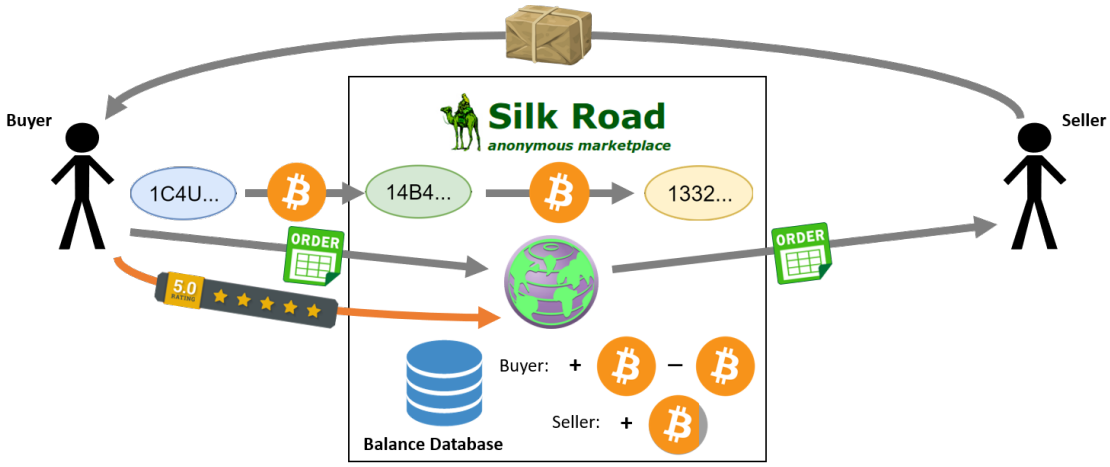


Figure 3.4: Normal process for purchases. Product review is only provided after receiving the item [5, 4].

Normal process Figure 3.4 presents the usual process recommended by Silk Road. First, the buyer deposits bitcoin to his user deposit address, and the balance database is updated with this value. The bitcoin then moves to a tumbler address. Next, the buyer sends an order for an item to the website and bitcoins are deducted from his balance. The seller receives the order, then ships the item to the buyer. This step anecdotally takes up to two weeks for international shipments. Once the buyer receives the item, he leaves a product review on the website. At this point the seller receives the payment, minus the cut taken by the marketplace.

For this normal process, the time difference between the blockchain transaction and the product review is largely dependent on the time it takes to ship the item. For this process the date range of blockchain transactions being searched is up to two weeks prior to the date of the product review.

Early finalization process Figure 3.5 presents the early finalization process. Same as the normal process, the buyer deposits bitcoin, balance database is updated, bitcoin

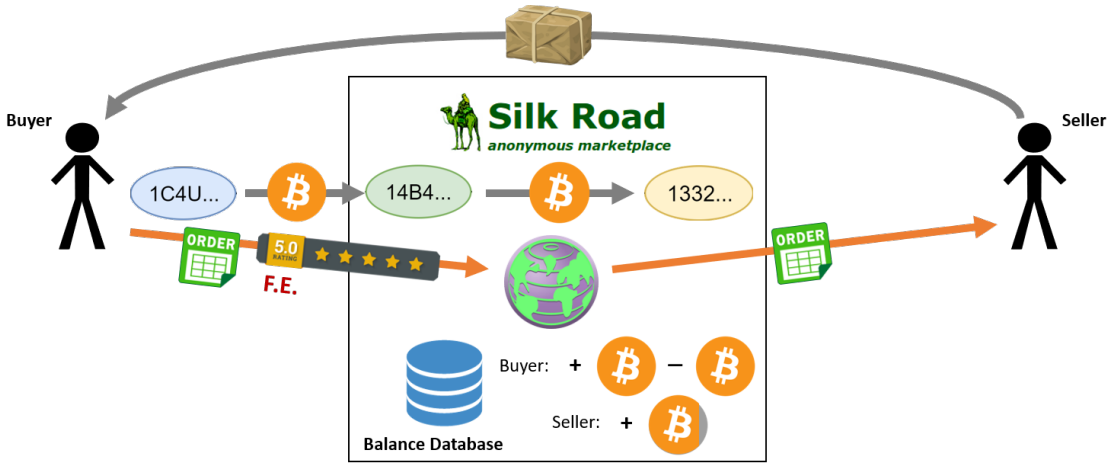


Figure 3.5: Early finalization process for purchases. Product review is provided right after the order is placed, before receiving the item.

moves to a tumbler address, the buyer orders an item, and bitcoins are deducted from his balance. The difference is that along with placing the order, even before the item is shipped, the user also immediately leaves a product review, usually with the comment "F.E." which stands for "finalized early". This is not the recommended practice of Silk Road because of the possibility of being scammed, however some sellers request this arrangement possibly to have the money to pay their real-life sources of the product and thus enable the shipment.

For this early finalization process, the time difference between the blockchain transaction and the product review is very small, because there is no need to wait for the time it takes to ship the item. For this process the date range of blockchain transactions being searched is only up to a day prior to the date of the product review.

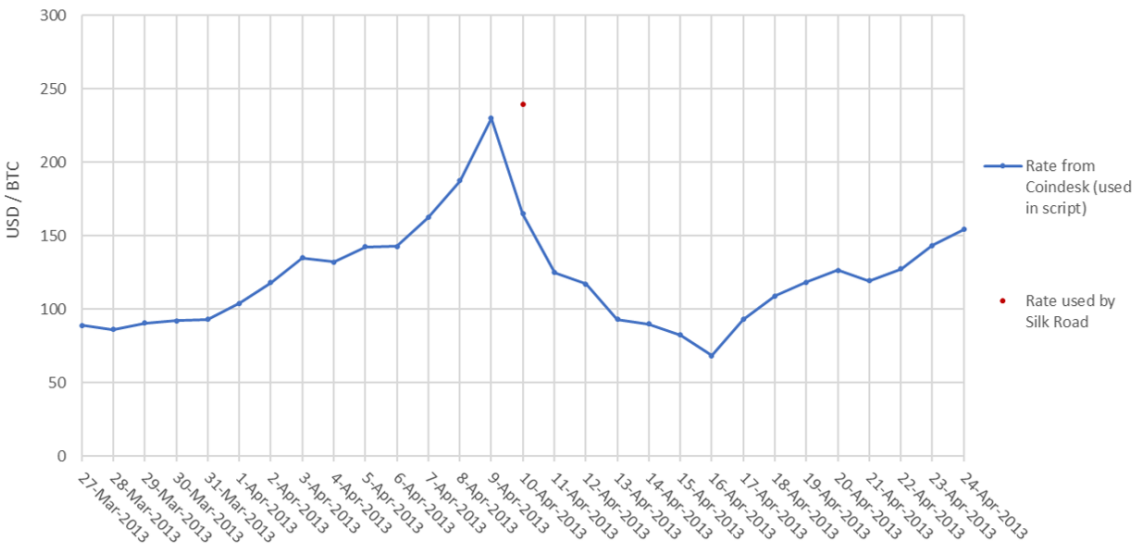


Figure 3.6: Exchange rate discrepancy. Expected rate of 170 USD/BTC, but actual rate used was 240 USD/BTC. Used range taken from 3-day window to take variation into account.

Bitcoin value range

For the bitcoin value range being searched, we use from 100% to 120% of the item's value. This is because users typically consider fluctuations in Bitcoin's exchange rate and therefore leave some margin.

Additionally, we also use a range for exchange rates. The exchange rates considered are from 95% of the minimum exchange rate in a 3-day window, to 105% of the maximum rate also in a 3-day window. The 3-day window is centered on the date of the marketplace transaction in question.

Figure 3.6 provides an example of when this exchange rate range is needed. For a known actual marketplace transaction traced, we find that the expected rate based on the date of the transaction should have been around 170 USD/BTC, however the actual rate used ended up to be 240 USD/BTC. The reason this is so could have been that the hour when the transaction happened is closer to the previous day, and

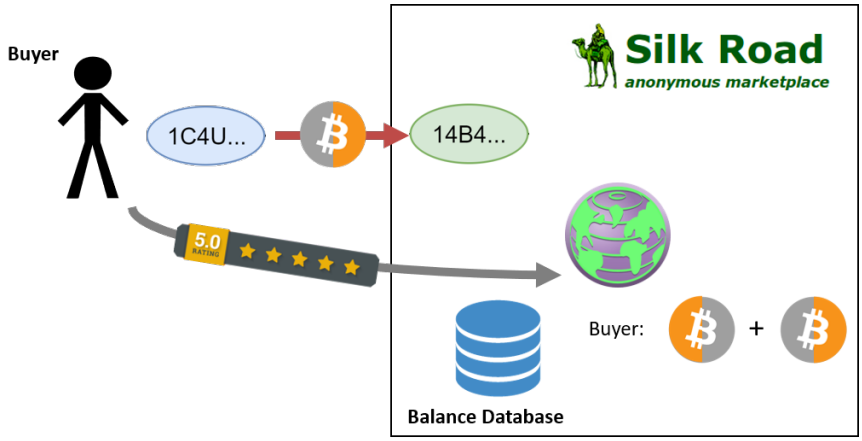


Figure 3.7: Partial deposit. Blockchain transaction is only a part of the value of the item indicated in the product review.

hence was closer to that day's exchange rate.

3.2.3 Limitations

This searching technique will not work for certain scenarios. One such situation is shown in Figure 3.7, wherein the user previously has some balance already, and thus only needs to deposit part of the amount. In this case the blockchain transaction will not match the value in the product review.

Another scenario is presented in Figure 3.8. Here, a user with good operational security practices could use two separate external source addresses to deposit to two different user deposit addresses, both associated to his Silk Road account (the system has a provision to request a new deposit address). In this situation, neither blockchain transaction would match the information in the product review.

3.2.4 Validation

Validation can be performed on the technique by applying it to the Spare Coins Thread of the Silk Road forum. The Spare Coins Thread is a forum post wherein

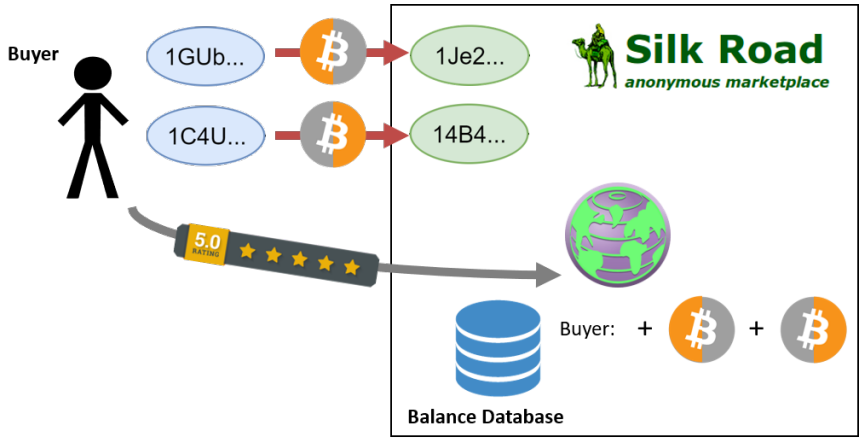


Figure 3.8: Split transaction. Neither of the two blockchain transactions would match the value of the item indicated in the product review.

users could request other Silk Road users for small amounts of bitcoin that they need to complete the amount needed for a purchase. There are two modes wherein this transfer could be facilitated. One is a slow, on-blockchain option and the other is a fast, off-blockchain option.

Figure 3.9 shows the on-blockchain option. Here the sender requests a withdrawal from the system, directed to the user deposit address of another user. The off-blockchain option accomplishes the same thing except that everything happens in the internal balance database and no activity is generated on the blockchain.

Given that the date, value, and Bitcoin address are indicated in the Spare Coins Thread, we could use the searching technique on the date and value instances, and verify if the associated Bitcoin address is found.

3.3 Purchasing Behavior Characterization

The Silk Road address classification discussed previously provides us a unique opportunity to analyze the behavior of user accounts. Since the user deposit addresses could be identified, we are able to look into the blockchain for just these addresses'

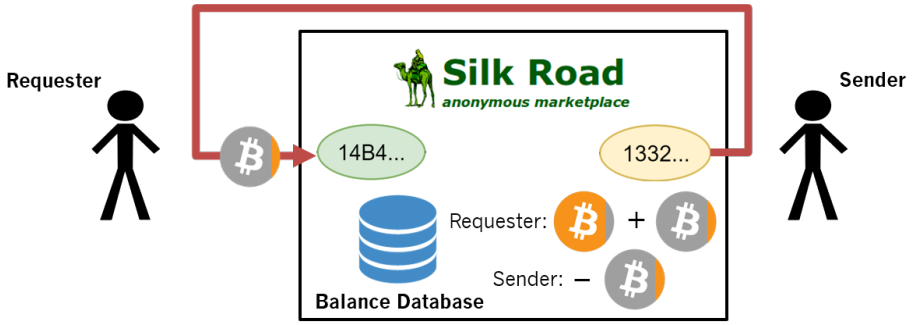


Figure 3.9: On-blockchain transfer option. The sender requests a withdrawal from the system, directed to the user deposit address of another user.

transactions, and characterize purchasing behavior as a result.

It should be noted though, that addresses are not 1:1 with real users. For privacy, some users have more than one Silk Road account. Also, within an account a user has the capability to request for a new deposit address associated to the account.

Our objective is to profile not only Silk Road addresses, but the real people behind those addresses. As such, we would need to look for a way to cluster addresses of one user together.

3.3.1 Clustering User Deposit Addresses to User Wallets

We conjecture that buyers use the same external source addresses to supply multiple deposit addresses. We also presume that a user's deposit address could be supplied by multiple external source addresses from the same user. Therefore, we could group together user deposit addresses that receive from the same external source address, and also group together external source addresses that move funds to the same user deposit address. Figure 3.10 illustrates this grouping.

However, it is also possible that an external source address is in fact an address controlled by an exchange. Such an address is expected to supply to thousands of addresses at a time, in and out of Silk Road. Therefore, for a case such as the one

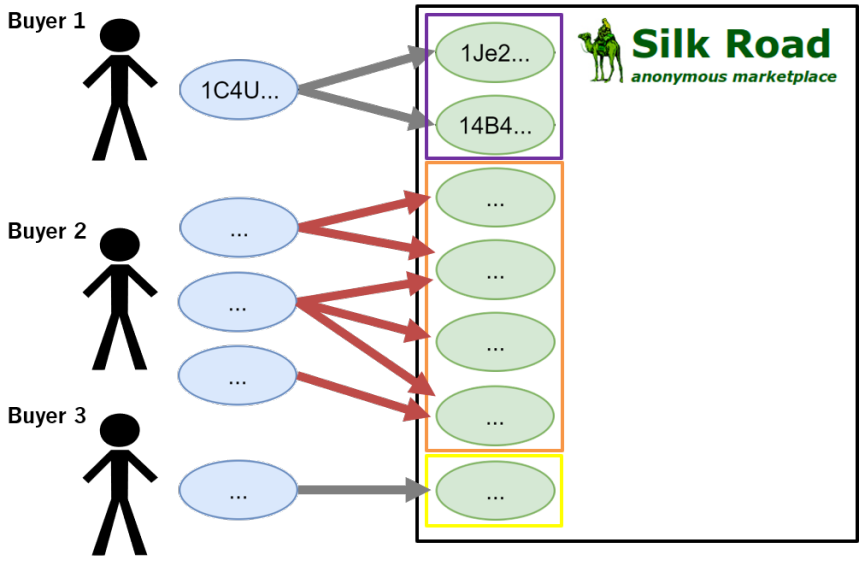


Figure 3.10: Address clustering into user wallets. User deposit addresses sharing external source addresses are grouped together. Likewise, external source addresses sharing user deposit addresses are also grouped together.

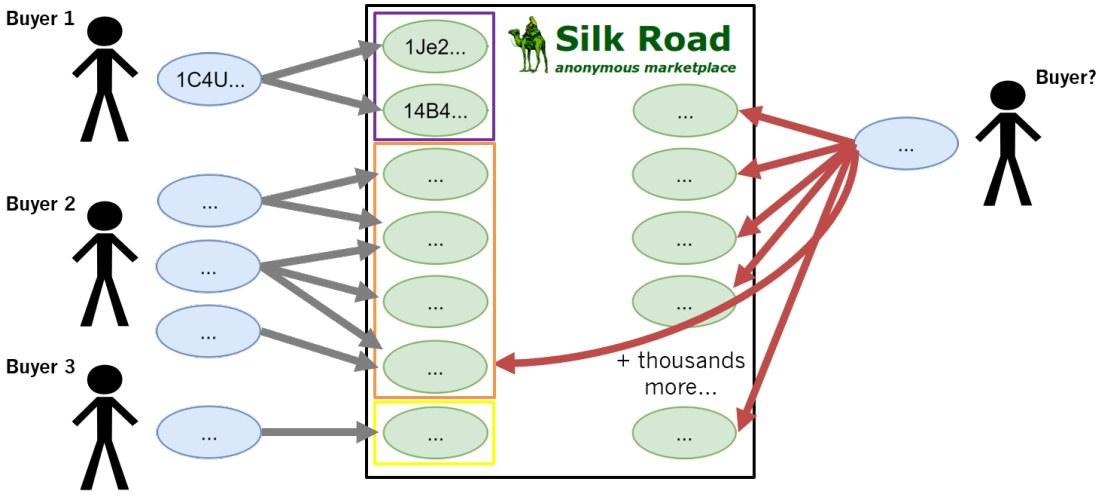


Figure 3.11: Exchange addresses excluded from clustering. Those with at least 1000 transaction outputs (whether to addresses inside Silk Road or not) are considered as exchanges and are not included in the clustering.

External Source Address	Exchange	No. of Tx Outputs
1BTC24yVKQdQNAa4vX71xLUC5A8Za7Rr71	Bitcoin-24.com	42k (4.2k to Silk Road)
1VayNert3x1KzbpzMGt2qdqrAThiRovi8	Deepbit	1.5M (3.7k to Silk Road)
1LNWw6yCxkUmkhArb2Nf2MPw6vG7u5WG7q	Mt.Gox	41k (3.2k to Silk Road)
1969VR2qCchXMW94tpcYirbVLUfFw4Pw7b	CoinAd.com	480k (1.5k to Silk Road)
1CDysWzQ5Z4hMLhsj4AKAEFwrgXRC8DqRN	Instawallet	9k (1.2k to Silk Road)

Table 3.1: Top external source address depositors to Silk Road.

shown in Figure 3.11, the address that supplies to a very large number of addresses is treated as an exchange address and is excluded from the clustering.

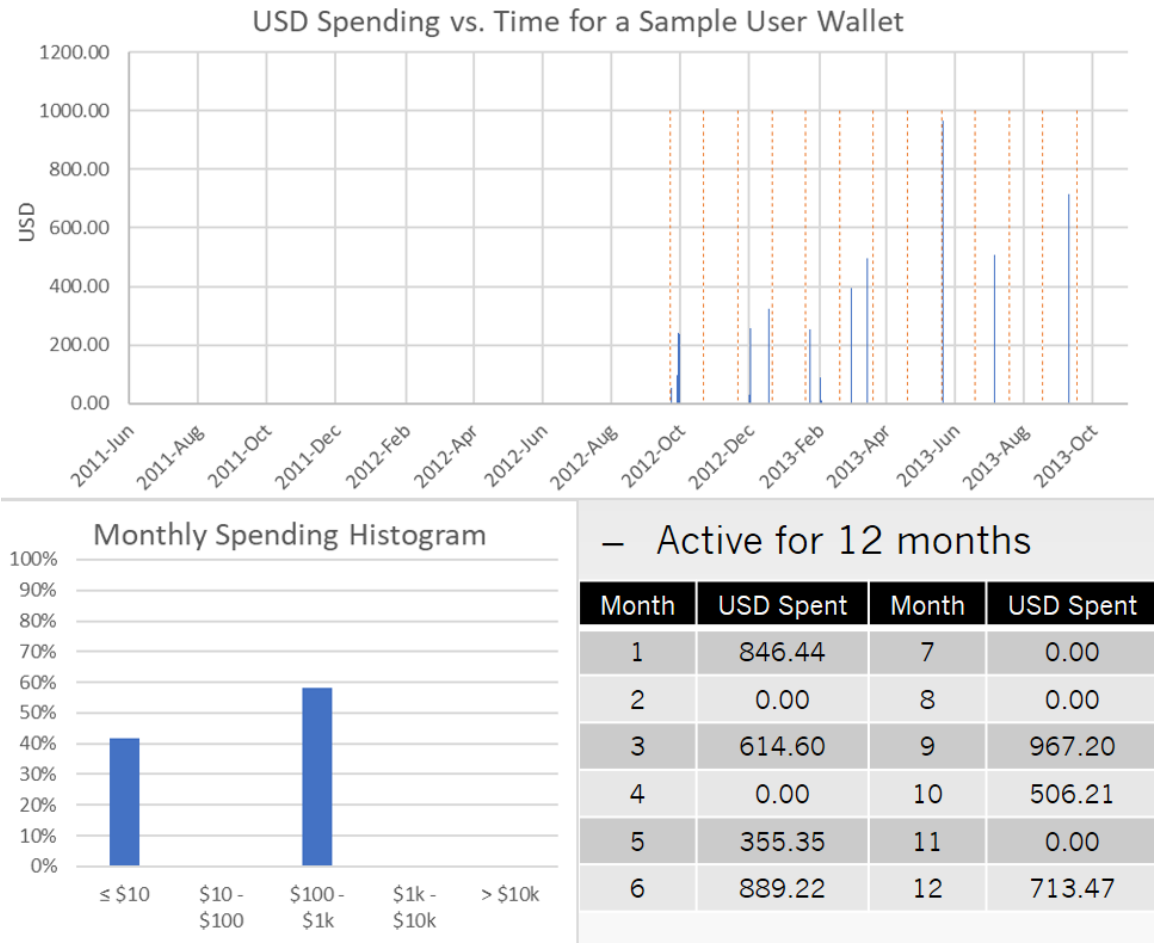
The threshold used for exclusion is 1,000. Those with at least 1,000 total transaction outputs (whether to addresses inside Silk Road or not) are considered as exchanges and are not included in the clustering. This number was chosen based on the values on Table 3.3.1. The external source addresses indicated are the top addresses that deposit into Silk Road. Running the addresses through an Internet search yields the exchange service that they are associated with. As it can be seen in the rightmost column of the table, the total number of transaction outputs on these addresses far exceed the threshold of 1,000 that we use.

3.3.2 Clustering User Wallets to Behavior Groups

With deposit addresses grouped into user wallets, the next thing we want to do is to cluster these user wallets based on their purchasing behavior. To do so, we profile the wallets according to two main criteria: 1) amount of money spent, and 2) frequency of transactions.

Feature generation based on transaction history

From the transaction history of the user wallets, we create two high-level features corresponding to the two criteria. First is the monthly USD spending histogram. This is the total USD spent on every 30-day period after the first deposit into the wallet. Second is the inter-arrival time of transactions histogram. This is simply the



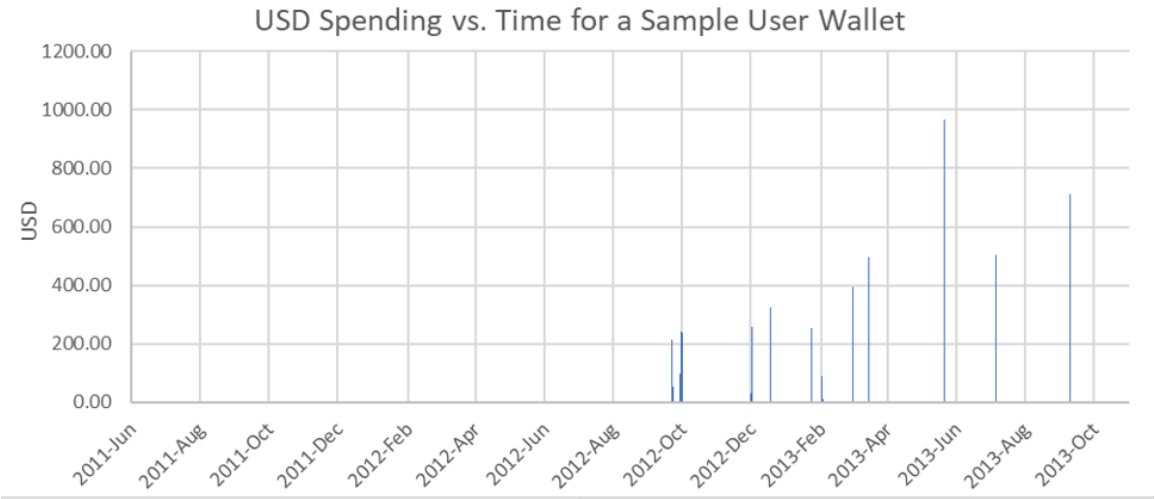
Month	USD Spent	Month	USD Spent
1	846.44	7	0.00
2	0.00	8	0.00
3	614.60	9	967.20
4	0.00	10	506.21
5	355.35	11	0.00
6	889.22	12	713.47

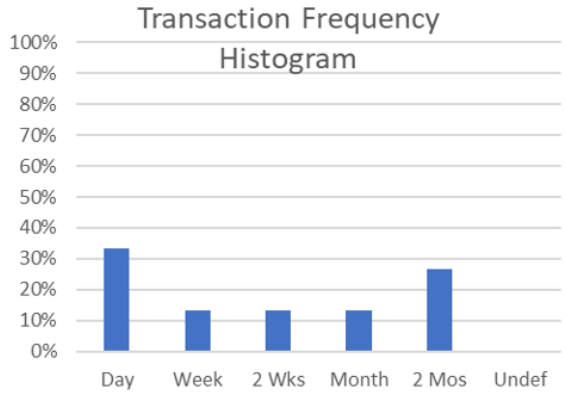
Figure 3.12: Monthly spending histogram. Total USD per month is assigned into bins.

number of days in between days with transactions.

Figure 3.12 illustrates how the monthly spending histogram is derived. First, the period the wallet is active is divided into 30-day periods. Then, for every period the total USD spent is calculated. These totals are then placed into bins for *leq*\$10, \$10-\$100, \$100-\$1k, \$1k-\$10k, and >\$10k.

From Figure 3.13, we could see how the transaction frequency histogram is obtained. For each day that a transaction is present, the number of days in between is





From Tx	# of Days	From Tx	# of Days
1-2	1	9-10	9
2-3	6	10-11	1
3-4	1	11-12	27
4-5	1	12-13	14
5-6	61	13-14	67
6-7	1	14-15	45
7-8	17	15-16	66
8-9	36		

Figure 3.13: Transaction frequency histogram. Number of days between transactions placed into bins.

tallied. These totals are then placed into bins for Day ($\leqslant$4 days), Week (5-9 days), 2 Weeks (10-20 days), Month (21-42 days), 2 Months ($\geqslant$43 days) and Undefined (only 1 day active).

Feature comparison

Now that we have distilled the two features into histograms, we need to have a measure of difference or distance between two user wallets with regards to these features. To do so, we calculate the distance between histograms using the histogram inter-

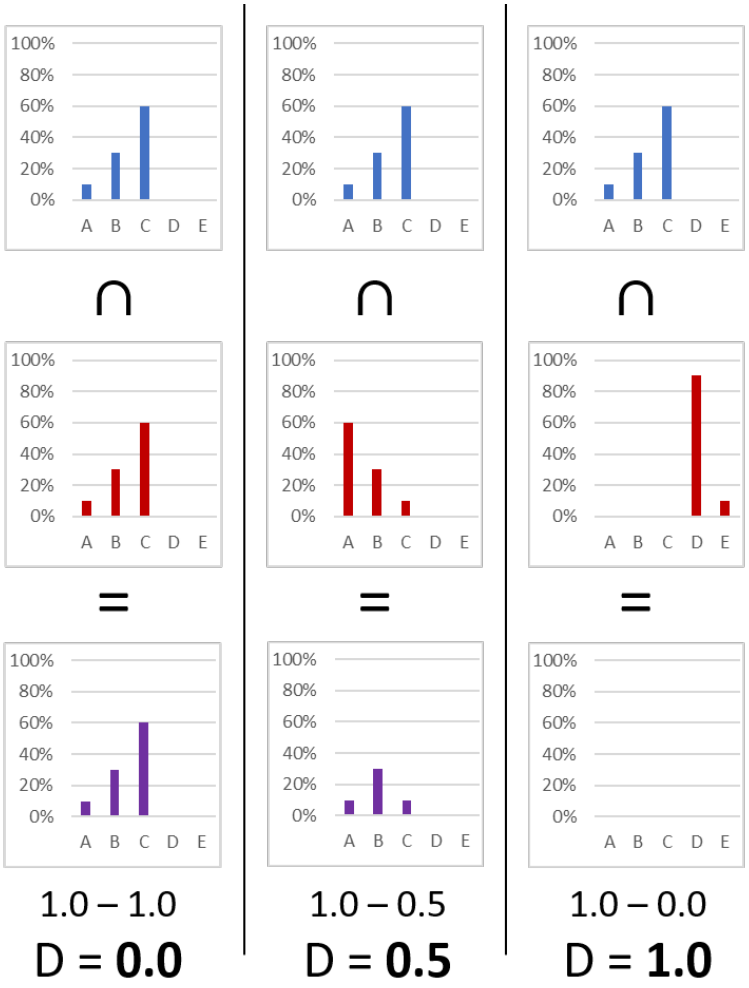


Figure 3.14: Sample distances using histogram intersection.

section algorithm introduced by Swain and Ballard [18]. Distance between wallets for a certain histogram could then be expressed as:

$$D = 1 - \sum_{b=1}^{B} min(H_{1b}, H_{2b}) \tag{3.3}$$

where B is the total number of bins, H_{1b} is the value of wallet 1's histogram for bin b, and H_{2b} is the value of wallet 2's histogram for bin b.

Figure 3.14 illustrates how the algorithm works on some sample histograms.

Overall distance function

Next, we create an overall distance function which includes the individual measures of distance for the two histogram types:

$$D_{overall} = \sqrt{D_{spending}^2 + D_{frequency}^2} \tag{3.4}$$

where $D_{spending}$ is the distance of the monthly spending histogram of the two wallets, and $D_{frequency}$ is the distance of the transaction frequency histogram of the two wallets.

Additionally, we add a special case in order to cluster the high-spending users together, regardless of transaction frequency. This changes the overall distance function to the following:

$$D_{overall} = \begin{cases} 0 & \text{\$10k bin of both histograms} > 0 \\ 1.4142 & \text{\$10k bin of only one histogram} > 0 \\ \sqrt{D_{spending}^2 + D_{frequency}^2} & \text{\$10k bin of both histograms} = 0 \end{cases} \tag{3.5}$$

Using this overall distance function, we cluster the user wallets into behavior groups using Hierarchical DBSCAN [8, 16]. For hyperparameters, we use minPts = 4, and minimum cluster size = 60. The minPts value is based on the rule of thumb of minPts = 2 * dim [15], while the minimum cluster size was derived empirically. It is large enough to limit the number of clusters, yet small enough not to negatively affect the high spending user cluster.

4

Results

We present our findings starting with the address classification, followed by the transaction tracing and purchasing behavior characterization.

4.1 Silk Road Addresses Classification

Among the 2.1M addresses on the Silk Road address list, we find that only 731k (35%) have been used, i.e. have at least one input transaction. The other 65% may have been for users who registered for an account to try out the market but ended up never purchasing anything nor even depositing bitcoin at least once.

From this 731k addresses, 304k (41.6%) are classified as user deposit addresses, because all transaction inputs come from outside Silk Road. 423k (57.9%) are categorized as automated tumbler addresses, because for these all their transaction inputs come from within the Silk Road system. A small number of addresses, around 4k in all (0.5%) are not initially classified as either, because they have deposits coming from both outside Silk Road, and also from within. For these addresses, we decide to treat them as deposit addresses for the purpose of the analysis. This is because we believe there is only a very low probability for a user to manually look through

the blockchain for a tumbler address then manually send funds to it and thus cause the address to have sources from both in and out of Silk Road. On the other hand, given the existence of the Spare Coins Thread, there is a higher probability that what actually happened is that a user posted his user deposit address to request for bitcoin, and someone obliged his request and performed a manual withdrawal with the requester's address as the destination.

Among the non-Silk Road addresses encountered along the way, 75k were classified as external source addresses, and 139k were classified as user withdrawal addresses.

Validation

For the Silk Road Seized Coins, we stated that the the maximum amount to be seized could be expressed as:

$$S_{max} = \sum_{n=1}^{N} \sum_{i=1}^{I_n} d(n,i) - \sum_{m=1}^{M} \sum_{j=1}^{J_m} w(m,j) \tag{4.1}$$

Plugging in the results from the script that we use to perform the summations, we get:

$$S_{max} = 10,757,805.80 - 10,726,669.78 \tag{4.2}$$

$$S_{max} = 31,136.02 \text{ bitcoins} \tag{4.3}$$

29,655 bitcoins, the amount reported to be seized by the FBI, falls within this expected maximum value of 31,136.02 bitcoins. The remaining 1481.02 bitcoins (5%) could have been withdrawn by users prior to the site being seized, as the total volume in the days leading to the take-down of Silk Road has withdrawals of that same order of magnitude as well.

For the comparison of daily volume estimates, Figure 4.1 displays the estimates from previous work and the estimates using the user deposit addresses from the

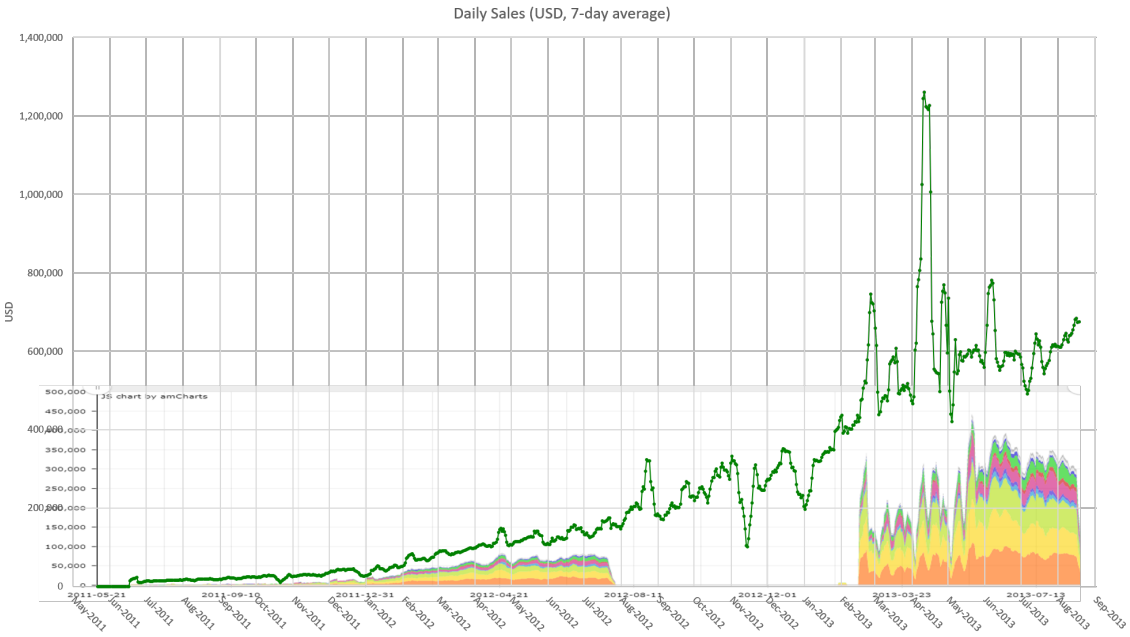


Figure 4.1: Daily volume estimates. Multi-colored part is the estimate from previous research based on the product review data. Green line is from the aggregate volume entering user deposit addresses per day.

classification combined in one graph. The multi-colored part is the estimate from previous research based on the product review data, while the green line is from the aggregate volume entering user deposit addresses per day. We can see that for every point in the graph the green line is above the lower bound set by the multi-colored plot. Additionally, the pattern tracks remarkably well, with a relatively stable percentage of difference between the two plots. This discrepancy could be explained by the existence of users withdrawing early without making a purchase. Alternatively, it could be that some users purchased more than 1 unit of the product indicated in the review, because while the deposited bitcoin should be a multiple of the cost of the item, the product review has a limitation of only being recorded once for each purchase, whether only one item was purchased or several were.

4.2 Transaction Tracing

Starting from the case of having no information at all aside from "a transaction was made in Silk Road", the size of the anonymity set would be 16.6M transactions. These are the number of transactions in the blockchain from June 17, 2011 until Oct 1, 2013, the days when Silk Road was operational.

When we include information the product review data side channel, we find that for normal, non-finalized-early (non-FE) transactions, the size of the anonymity set drops to 45k (0.26% of 16.6M) on average (median: 41k), when the search is conducted over a sample of 30k transactions out of all 541k instances of non-FE transactions. For finalized early (FE) transactions, the size of the anonymity set drops to 4.3k instead (0.026% of 16.6M) on average (median: 3.5k), over all 61k instances of FE transactions. FE transactions had a lower set size because we use a smaller range of one day for the search instead of two weeks for non-FE transactions.

When we apply the filter of having destination addresses as user deposit addresses only, for non-FE the number drops down to 1.7k (4% of 45k) on average (median: 1.2k). While for FE transactions the number decreases down to 215 (5% of 4.3k) on average (median: 134), representing a further reduction of a factor of 20 compared to the previous step.

Figure 4.2 and Figure 4.3 plot the number of matches when only using the product review data for non-FE and FE transactions respectively. Figure 4.4 and Figure 4.5 on the other hand show the number of matches when also including information from the Silk Road address list, also for non-FE and FE transactions respectively.

On these graphs, the vertical lines with high number of matches represent dates with relatively high fluctuations in the bitcoin exchange rate, causing the searching algorithm to user a larger range of exchange rate. Horizontal lines with relatively higher number of matches represent price points that are common in the market.

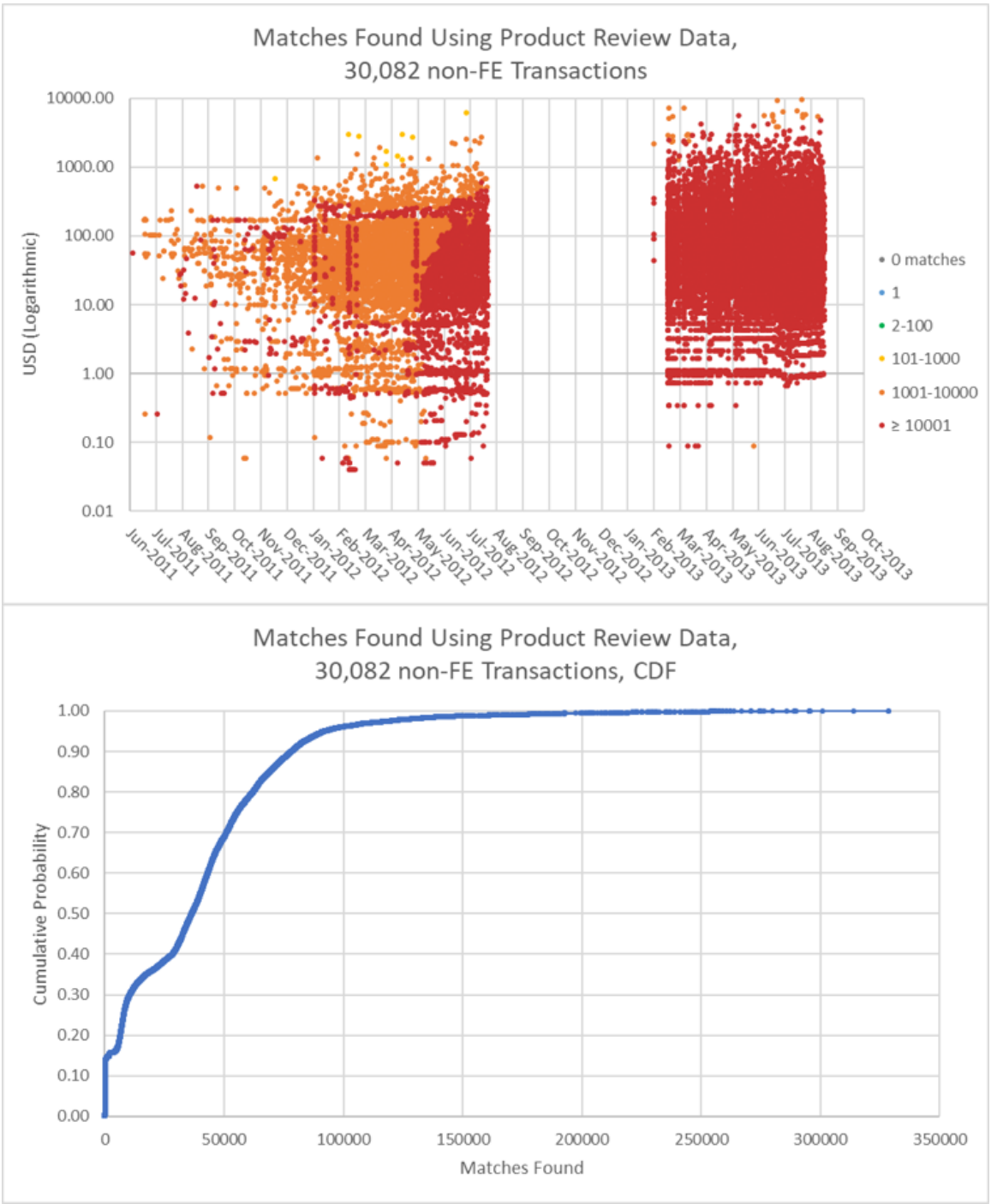


Figure 4.2: Matches found using product review data, non-FE.

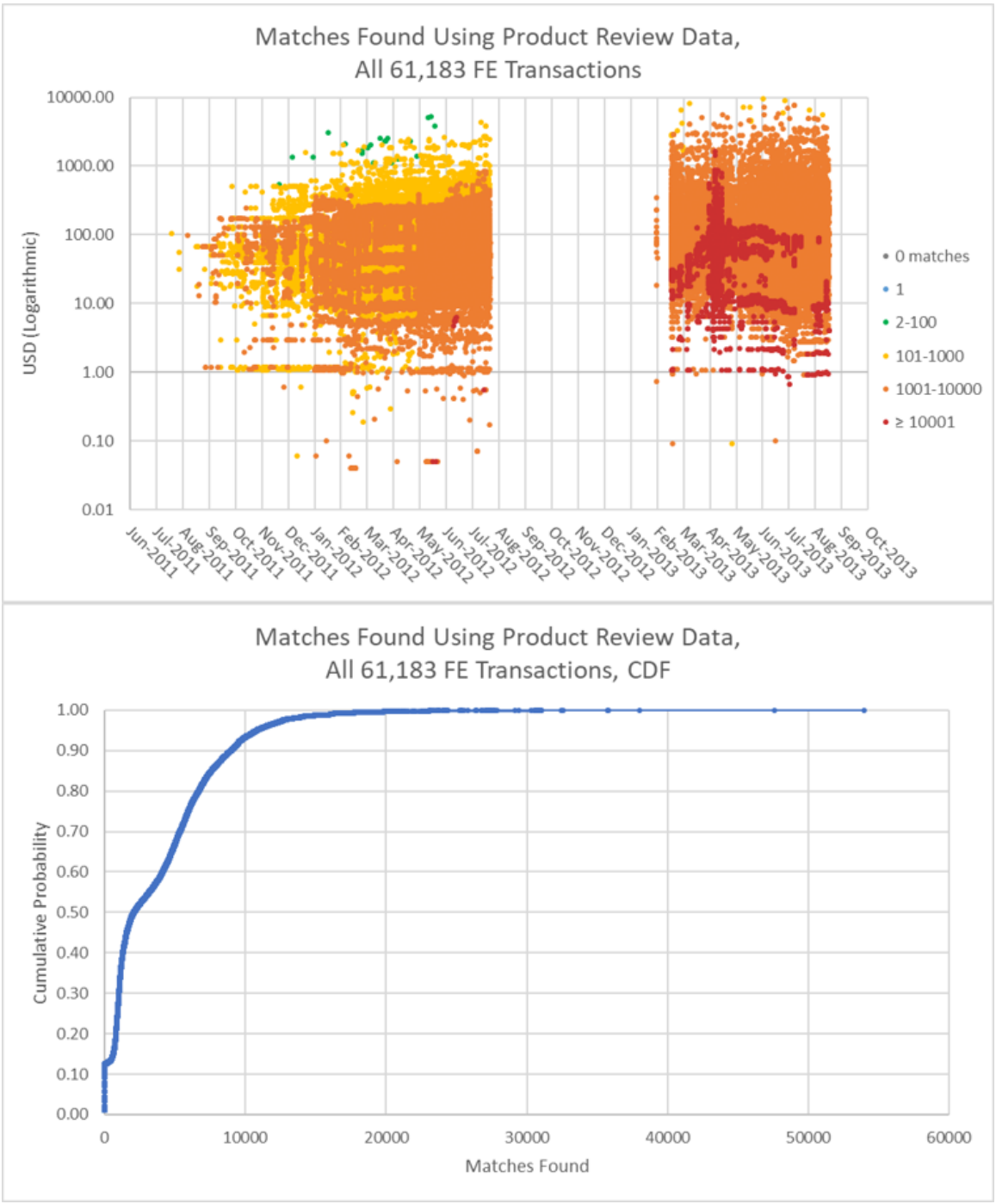


Figure 4.3: Matches found using product review data, FE.

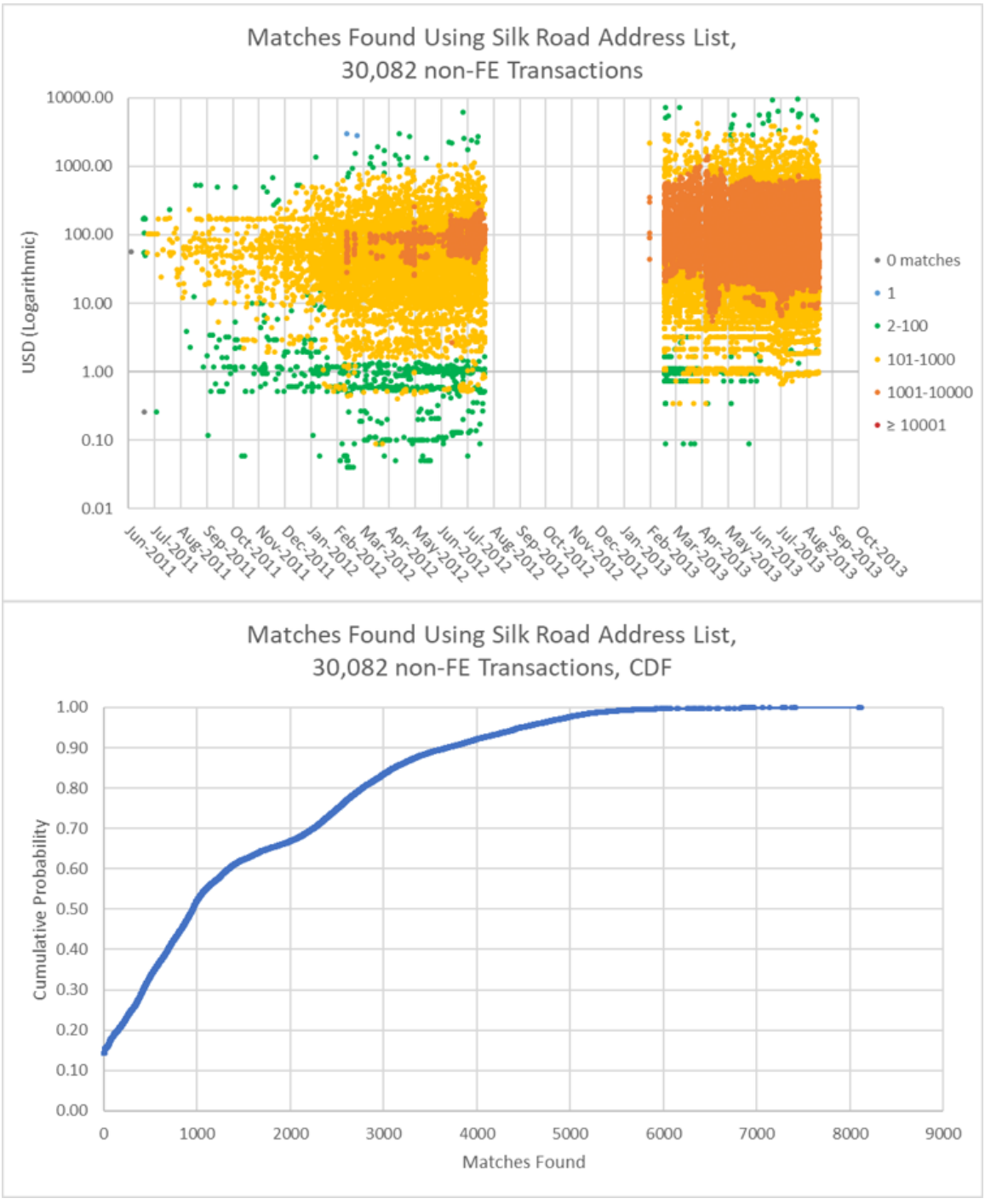


Figure 4.4: Matches found using Silk Road address list, non-FE.

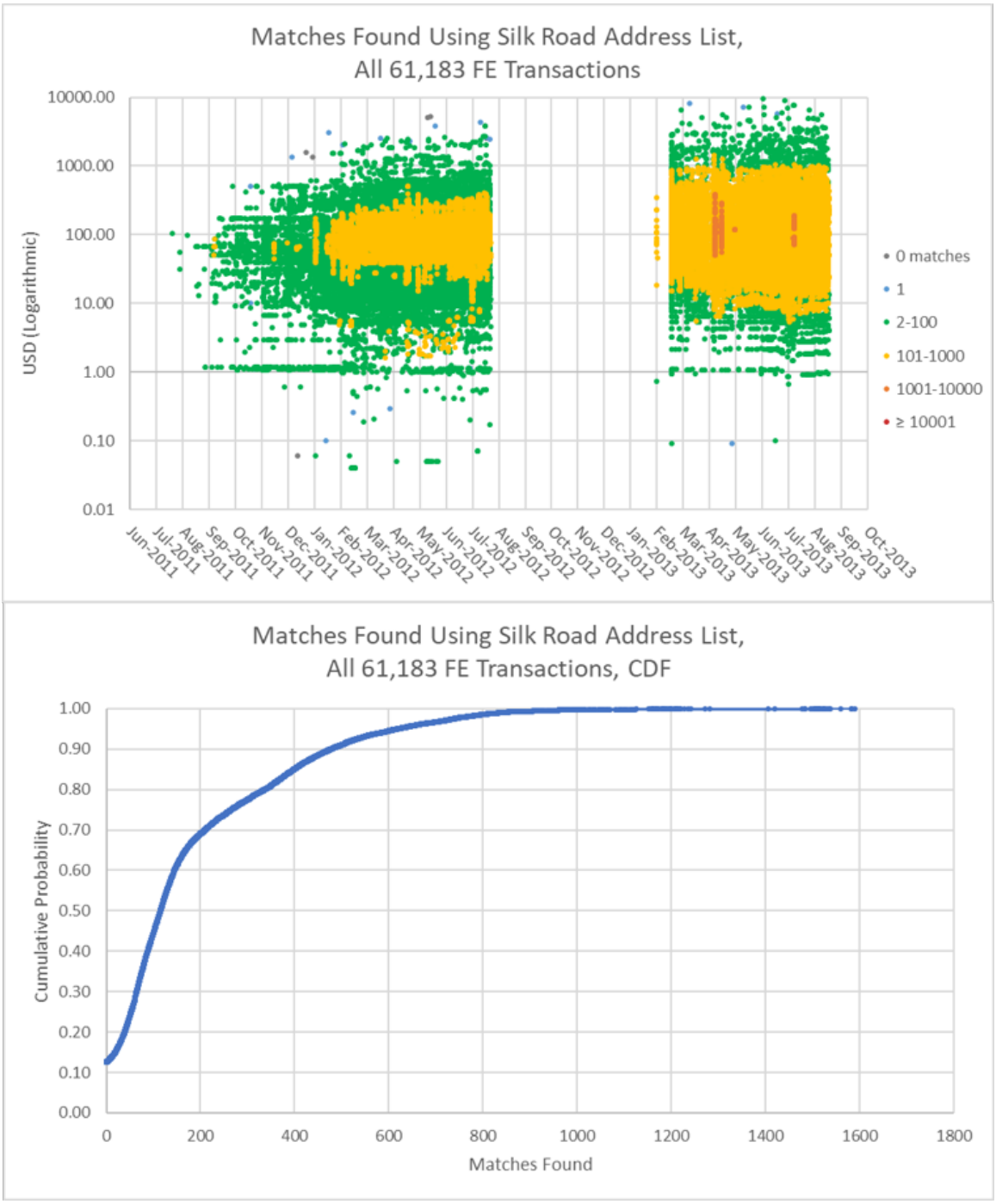


Figure 4.5: Matches found using Silk Road address list, FE.

Validation

Using the Spare Coins Thread to validate the technique, we find that for the cases when the sender explicitly states that he used the on-blockchain method to transfer funds, 15 out of 15 (100%) transactions are reidentified. Out of these, 13 are uniquely identified, i.e. anonymity set size of 1, while the remaining 2 transactions had a set size of 2.

For the case where the sender does not state the option used, but the requester provided only the Bitcoin address (suggesting a higher probability that the on-blockchain method would be used), we find that 33 out of 43 (77%) transactions are reidentified. Of these 32 are uniquely identified, while only 1 had an anonymity set size of 2. The transactions that are not reidentified could have been transfers that were performed using the off-blockchain method, as we find evidence also in that thread where senders have simply assumed that a user's forum name is the same as their Silk Road account name, and thus send funds with the off-blockchain method using that assumed name.

4.3 Purchasing Behavior Characterization

Among the 308k addresses we consider as user deposit addresses, using the heuristics mentioned previously, we cluster 126k (41%) into 13k user wallets. 182k (59%) addresses we treat as singletons with their own user wallet containing only one address, because there is no non-exchange external address linking them to another deposit address. All in all this results to 13k + 182k = 195k user wallets.

Of these 195k user wallets, we cluster 181k (93%) into 153 behavior groups. The remaining 14k (7%) are outliers in the dataset and are not categorized into a behavior group.

Among the 153 behavior groups, one group, composed of 410 (0.2%) user wallets

are high volume spenders. These are wallets that have deposited at least $10k in any monthly period. We find that this group contributes 22% of the total volume in the market.

Three groups, comprising of 116k (59%) user wallets in total are one-time spenders. These wallets have all their transactions on just one day. Spending ranges from low to high, but never reaching $10k in any monthly period (in which case it would be categorized as a high volume spender). This type of spenders contribute 31% of the total volume in the market.

149 behavior groups, consisting of 65k (33%) wallets all in all are the regular spenders. There is a high number of behavior groups falling under this category, representing the many ways to categorize these types of users. One of the largest groups within this category are those who spend $10-$100 per month, transacting on a monthly basis. Another large group have users who spend $100-$1000 in total per month, and transact typically once every two weeks. Many other groups can be formed for each combination of monthly spending and transaction frequency.

The 14k (7%) wallets belonging to the outlier group contribute the remaining 21% of the total volume, indicating that they on the average spend more than a regular spender, but not to the levels of the high volume spenders.

5

Conclusions

We set out to look into the extent to which we could trace Bitcoin transactions and characterize purchasing behavior of online anonymous marketplaces by harnessing and exploiting the information available to us from side channels.

Using the dates and US dollar values from product review data, we reduce the size of the anonymity set by a factor of around 400 to 4,000. Furthermore, we discover that including the information from the Silk Road address list further to reduce the size of this set by a factor of 20.

For purchasing behavior characterization, we find that classifying the Silk Road address list places us in a unique position to be able to perform analysis on user transaction patterns and spending habits. After clustering was performed, we see users being categorized into three main groups: 0.2% are high volume spenders, 59% are one-time customers, and 33% are regular users, contributing 22%, 31% and 26% of the market volume respectively.

Side channels were the key to uncovering all these. The information from product review data from a public website, Silk Road address list presented by the FBI, marketplace forum posts and official guides provide insight into the system and

extend the boundaries on what we can trace and characterize. As time goes by, more and more sources of side information become available. In fact, there may even be a lot more presently existing side channels that we have not yet considered!

Bibliography

[1] "Silk road — download." [Online]. Available: https://antilop.cc/sr/download/." [Accessed 2017-08-01].

[2] "Silk road timeline." [Online]. Available: https://antilop.cc/sr/#exhibit." [Accessed 2017-08-01].

[3] "Manhattan u.s. attorney announces seizure of additional $28 million worth of bitcoins belonging to ross william ulbricht, alleged owner and operator of silk road website," Oct 2013. [Online]. Available: https://archives.fbi.gov/archives/newyork/press-releases/2013/manhattan-u.s.-attorney-announces-seizure-of-additional-28-million-worth-of-bitcoins-belonging-to-ross-william-ulbricht-alleged-owner-and-operator-of-silk-road-website. [Accessed 2017-09-01].

[4] "Rating," Feb 2016. [Online]. Available: http://www.nic.org/blog/peering-under-the-hood-of-cms-five-star-quality-rating-system/. [Accessed 2017-11-27].

[5] "Order," Nov 2017. [Online]. Available: https://play.google.com/store/apps/details?id=com.eonsoft.OrderListV2. [Accessed 2017-11-27].

[6] J. Biggs, "As feds fumble with bitcoin, the internet trolls the fbis private wallet," Oct 2013. [Online]. Available: https://techcrunch.com/2013/10/07/as-feds-fumble-with-bitcoin-the-internet-trolls-the-fbis-private-wallet/. [Accessed 2017-12-08].

[7] J. Brito and A. Castillo, *Bitcoin: A primer for policymakers*. Mercatus Center at George Mason University, 2013.

[8] R. J. Campello, D. Moulavi, A. Zimek, and J. Sander, "Hierarchical density estimates for data clustering, visualization, and outlier detection," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 10, no. 1, p. 5, 2015.

[9] N. Christin, "Traveling the silk road: A measurement analysis of a large anonymous online marketplace," in *Proceedings of the 22nd international conference on World Wide Web*. ACM, 2013, pp. 213–224.

[10] Coindesk, "Bitcoin price index - real-time bitcoin price charts." [Online]. Available: https://www.coindesk.com/price/." [Accessed 2017-12-07].

[11] J. Jackson, "Fbi consultant: Silk road founder had $16-18m worth of bitcoins on laptop," Jan 2015. [Online]. Available: https://www.computerworld.com/article/2877772/malware-cybercrime/fbi-consultant-silk-road-founder-had-16-18m-worth-of-bitcoins-on-laptop.html. [Accessed 2017-09-01].

[12] J. Martin, "Lost on the silk road: Online drug distribution and the cryptomarket," *Criminology & Criminal Justice*, vol. 14, no. 3, pp. 351–367, 2014.

[13] S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelker, and S. Savage, "A fistful of bitcoins: characterizing payments among men with no names," in *Proceedings of the 2013 conference on Internet measurement conference*. ACM, 2013, pp. 127–140.

[14] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.

[15] J. Sander, M. Ester, H.-P. Kriegel, and X. Xu, "Density-based clustering in spatial databases: The algorithm gdbscan and its applications," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 169–194, 1998.

[16] E. Schubert, A. Koos, T. Emrich, A. Züfle, K. A. Schmid, and A. Zimek, "A framework for clustering uncertain data," *PVLDB*, vol. 8, no. 12, pp. 1976–1979, 2015. [Online]. Available: http://www.vldb.org/pvldb/vol8/p1976-schubert.pdf. [Accessed 2017-12-08].

[17] K. Soska and N. Christin, "Measuring the longitudinal evolution of the online anonymous marketplace ecosystem." in *USENIX Security Symposium*, 2015, pp. 33–48.

[18] M. J. Swain and D. H. Ballard, "Color indexing," *International journal of computer vision*, vol. 7, no. 1, pp. 11–32, 1991.