QATAR UNIVERSITY

COLLEGE OF ENGINEERING

DEANONYMIZING TOR HIDDEN SERVICE USERS THROUGH BITCOIN TRANSACTIONS ANALYSIS

BY

HUSAM BASIL AL JAWAHERI

A Thesis Submitted to the Faculty of the
COLLEGE OF ENGINEERING
in Partial Fulfillment
of the Requirements
for the Degree of
Master of Computing

June 2017

COMMITTEE PAGE

The members of the committee approve the Thesis of HUSAM BASIL AL JAWAHERI defended on May 25, 2016:

Qutaibah Malluhi

Thesis Supervisor

Mashael Al Sabah

Committee Member

Aiman Erbad

Committee Member

Approved:

Khalifa Al Khalifa, Dean, COLLEGE OF ENGINEERING

Abstract

Al Jawaheri, Husam, B, Masters:

June: 2017, Master of Computing

Title: DEANONYMIZING TOR HIDDEN SERVICE USERS THROUGH BITCOIN TRANSACTIONS ANALYSIS

Supervisor of Thesis: Qutaibah Malluhi

With the rapid increase of threats on the Internet, people are continuously seeking privacy and anonymity. Services such as Bitcoin and Tor were introduced to provide anonymity for online transactions and Web browsing. Due to its pseudonymity model, Bitcoin lacks retroactive operational security, which means historical pieces of information could be used to identify a certain user. We investigate the feasibility of deanonymizing users of Tor hidden services who rely on Bitcoin as a method of payment. In particular, we correlate the public Bitcoin addresses of users and services with their corresponding transactions in the Blockchain. In other words, we establish a provable link between a Tor hidden service and its user by simply showing a transaction between their two corresponding addresses. This subtle information leakage breaks the anonymity of users and may have serious privacy consequences, depending on the sensitivity of the use case.

To demonstrate how an adversary can deanonymize hidden service users by exploiting leaked information from Bitcoin over Tor, we carried out a real-world experiment as a proof-of-concept. First, we collected public Bitcoin addresses of Tor hidden services from their .onion landing pages. Out of 1.5K hidden services we crawled, we found 88 unique Bitcoin addresses that have a healthy economic activity in 2017. Next, we collected public Bitcoin addresses from two channels of online social networks, namely, Twitter and the BitcoinTalk forum. Out of 5B tweets and 1M forum pages, we found 4.2K and 41K unique

online identities, respectively, along with their public personal information and Bitcoin addresses. We then expanded the lists of Bitcoin addresses using closure analysis, where a Bitcoin address is used to identify a set of other addresses that are highly likely to be controlled by the same user. This allowed us to collect thousands more Bitcoin addresses for the users. By analyzing the transactions in the Blockchain, we were able to link up to 125 unique users to various hidden services, including sensitive ones, such as The Pirate Bay, Silk Road, and WikiLeaks. Finally, we traced concrete case studies to demonstrate the privacy implications of information leakage and user deanonymization. In particular, we show that Bitcoin addresses should always be assumed as compromised and can be used to deanonymize users.

Dedication

To my mom and dad, who supported me throughout my life to reach this point. I dedicate this work for you. Thank you for everything.

Acknowledgements

In the first place, I would like to acknowledge my co-supervisor Dr. Mashael Al Sabah for her constant support and help with different parts of this thesis. I believe without her help this would have been much harder. Second I would like to thank Dr. Aiman, other co-supervisor for his amazing advices and support. Thank goes to Yazan Boshmaf from QCRI for his amazing contributions in the data collection part. Big thanks goes also to Dr. Ryan Riley for his technical support in the early stages of the implementation. My brother Hasan also deserves a huge thank for his help with programming bugs here and there. Finally, I would like to thank anyone who supported me through this work in any way, without you all this would not have been possible.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Anonymity and privacy over the Internet are becoming more critical than ever. For that, many solutions are being deployed to improve the anonymity of users while browsing the web or doing online transactions. The most famous of these being the Bitcoin digital payment network and Tor anonymity network [20]. Bitcoin [34] is a decentralized digital currency network that provides users with the ability to perform online transactions anonymously. Tor [20] is the most widely used anonymous communication network with millions of daily users [5]. In addition to client-side privacy and anonymity, Tor also enables server-side anonymity through the design of hidden services. The goal of hidden services is to safely enable online freedom, anticensorship, and end-to-end anonymity and security [19]. Indeed, for those reasons, hidden services are operated by whistleblowing websites such as WikiLeaks, search engines such as DuckDuckGo, and social media providers such as Facebook. Hidden services have also become breeding grounds for the Dark Web vendors, such as Silk Road and Agora, which offer illicit merchandise and services [13, 33].

As discussed by Vincent and Johan [31], Tor and Bitcoin represent the main components for achieving anonymous online purchases with exhaustive operational security. In this context, operational security is the process of protecting individual pieces of information that could be used to identify a certain user. Unfortunately, Bitcoin lacks retroactive operational security due to its pseudonymity model [34]. This model has an important limitation because of

the linkability of Bitcoin transactions that are stored in the Blockchain and their public availability.

Problem A serious threat to the anonymity of Tor hidden services is their reliance on Bitcoin as a main channel of payment, which could lead to possible information leaks. Yet, Bitcoin is the most popular choice for Tor's hidden services for accepting donations or for selling merchandise [13]. Moore and Rid [33] recently studied how hidden services are used in practice, and noted that Bitcoin was the dominant choice for accepting payments for these services. While multiple studies [21, 22, 30] demonstrated that Bitcoin transactions are not as anonymous as previously thought, Bitcoin remains the most popular digital currency in the Dark Web [15], and many users still choose to use it despite its false sense of anonymity. Biryukov et al. [12] recently showed that even if users use Bitcoin over an anonymity network such as Tor, they are still vulnerable to deanonmization and man-in-the-middle attacks at the network level. While previous studies analyze the vulnerabilities that result from using Bitcoin over Tor [12], mostly at the network level, we provide the first study that sheds light on the information leakage resulting from combining public data from online social networks, Bitcoin transactions, and Tor's hidden services.

Hidden service users are one class of Bitcoin users whose anonymity is particularly important. The reason is that, by using the Tor network, hidden service operators and users are actively seeking to maintain their anonymity. However, those users are under the risk of deanonymization simply by revealing their Bitcoin addresses. By studying the transactions associated with these addresses, a significant amount of information can be leaked and used to gather sensitive information about hidden services and their customers, where a user can provably be linked to a hidden service.

In this thesis, we seek to understand the privacy and anonymity risks that Tor hidden service users expose themselves to by using Bitcoin as a payment channel. We also seek to study the implications of information leaks through Bitcoin transactions over Tor due to its lack of retroactive operational security.

Approach By browsing various hidden service landing pages, we observed that it is possible to extract the Bitcoin addresses of these services with minimal effort. Accordingly, we crawled 1.5K hidden service pages, and compiled a list of 105 Bitcoin addresses operated by those hidden services, including few ransomware addresses. We also crawled online social networks for public Bitcoin addresses, namely, Twitter and the BitcoinTalk forum. Out of 5B tweets and 1M forum pages, we found 4.2K and 41K unique online identities, respectively, along with their public personal information and Bitcoin addresses. We then analyzed the transactions in the Blockchain using the collected Bitcoin addresses in order to identify links between Bitcoin users, as online identities with public profiles, and Tor hidden services. This enabled us to provably link identities with hidden services and access their transaction history.

Using a simple heuristic proposed by Meiklejohn et al. [30], we extended the transaction analysis with a wallet-closure technique that allowed us to expand the collected Bitcoin addresses per user. So, for each address in our compiled lists, we were able to identify other addresses belonging to the same user who owns the address. This closure analysis approximates a user's wallet, which is the set of addresses that are controlled by the user. As a result, we were able to increase the number of identified links between users and hidden services, and thereby increase the number of deanonymized users. One problem with closure analysis is that the closure can over-approximate the size of the wallet, as a consequence of mixing [28] and CoinJoin [41] services. Therefore, we excluded closures that have common addresses from the analysis. This ensures that users

are not double-counted and reported results are lower-bounds estimates, as each remaining closure represents a (partial) wallet whose addresses are controlled by a unique user. To demonstrate the impact of deanonymization, we traced and described two case studies from linked users that show Bitcoin addresses should always be assumed compromised, as they can be used to deanonymize users. It is important to note that deanonymization depends on data that is publicly available.

Finally, to gain insights about the economic activity of the hidden services that were linked with deanonymized users, we analyzed the corresponding transaction history, focusing on number of transactions, the amount of money being exchanged, and the lifetime of these hidden services.

Results We were able to link 81 unique users to various sensitive hidden services, including The Pirate Bay and WikiLeaks. By performing closure analysis, we were able to increase the number of deanonymized users to 125. Digging deeper with case studies, we unmasked multiple users of The Pirate Bay hidden service, along with their personally identifiable information, such as location and age, where such services are illegal in their country. Another case study shows that users from multiple countries and different ages had links with the Silk Road address in our hidden service list. Interestingly, one of these users is a 13-year old boy who has many social media accounts showing his real identity.

The economic activity analysis of the linked Tor hidden services shows that Wikileaks and the Darknet Bitcoin mixer are among the highest receivers of payments. We observed that the flow of money coming in and out of hidden services is almost similar. This could mean that such services do not keep their Bitcoins on the address they use for receiving payments, but rather distributed

the coins to other addresses instead. Finally, we found that 34% of the hidden services we included in the analysis are still active in 2017.

Contributions In this work we show the implications of Bitcoin's pseudonymity model, which lacks retroactive operational security, on Tor hidden service users. Our contributions are the following:

1. A method that provably links online user identities with Tor hidden services through Bitcoin transactions analysis. The method improves linking results by using closure analysis techniques and by significantly eliminating the noise from mixing and CoinJoin services.

2. The first real-world experiment showing the feasibility of deanonymizing Tor hidden service users by exploiting a subtle information leakage from public data sources, namely, online social networks, Bitcoin's blockchain, and Tor hidden services.

3. Insights into the transaction history of various hidden services that were used by a number of deanonymized users. This includes statistics on their transactions, flow of money, and time activity.

4. Two datasets representing Twitter and BitcoinTalk forum online identities and their corresponding Bitcoin addresses.[1]

[1]Datasets available here: https://goo.gl/ZXtJWy

Chapter 2: Background

We now present the necessary background on Bitcoin and Tor.

2.1 Bitcoin

Bitcoin [34] is a decentralized digital crypto-currency system which eliminates the need for a central bank authority to manage the transfer of funds. The Bitcoin network is maintained by a peer-to-peer network of miners who validate transactions without relaying on trust. Due to its popularity, more than 100K merchants worldwide accept Bitcoin payments [17]. One of the reasons of Bitcoin's popularity is its presumed anonymity. The identities of users on Bitcoin are hidden using pseudonyms, derived from public/private key pairs, which are used as user addresses to perform transactions. To increase anonymity, users are encouraged to create new addresses for each transaction.

2.1.1 Transactions

In Bitcoin, Alice makes a payment to Bob by creating a new transaction. She uses one or more Bitcoin addresses that she controls as inputs. She also includes the amount to be transferred, and chooses Bob's address(es) as a transaction output. To protect the transaction, she signs it using her private key, and then broadcasts it to the whole network. In order to verify transactions and be rewarded with new generated coins, miners collect the broadcast transactions, embed them in a well-defined data structure called a block, and then attempt

to solve a hashing computational puzzle involving the block. When the block is solved, it is attached to the Blockchain, which is a hash-chain that maintains all solved blocks, and thereby all embedded transactions ever created and verified in the Bitcoin network.

The Blockchain is publicly maintained and can be downloaded using BitTorrent, Bitcoin's core client, or explored using centralized servers, such as BlockchainInfo.[1] Every transaction in the Blockchain has a list of inputs and outputs, where each includes addresses that were used in the transaction and the amount of coins spent in that transaction. Transactions downloaded from BlockchainInfo include more information, such as the relay IP address and the transaction timestamp that records the time at which the transaction was made. Figure 2.1 depicts the Blockchain and a simplified transaction data structure of a block.

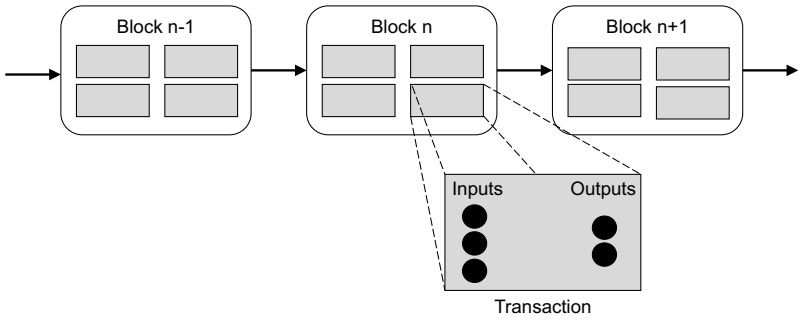

Figure 2.1: Bitcoin's blockchain and an example transaction.

2.1.2 Anonymity

While transactions in Bitcoin are presumed to be anonymous, linkability between addresses is possible due the nature of the Blockchain [34]. For example, one can verify if Alice and Bob have a transaction between them. Furthermore, if Alice owns multiple addresses, one may be able to link them as belonging to the same person.

[1] `https://blockchain.info`

Meiklejohn et al. [30] observed that two Bitcoin addresses, A and B, belong to the same user if both A and B have been used as inputs for the same transaction, or A receives, as an output, the unspent change of a transaction where B is used as input. The authors used this observation to define a heuristic for mapping multiple addresses to an entity representing a unique user. Specifically, the heuristic is based on the idea that since the private keys of the user are used to sign the inputs A and B, then both A and B are controlled by the same person. As the addresses or the underlying public/private keys that are owned by a user represents a wallet, the heuristic tries to induce the wallet of a user given a subset of the addresses in the wallet. The authors also define a second heuristic based on another observation. When an address is used as an input in a transaction, all of its associated Bitcoins have to be spent at once. If those coins exceeds what the sender wants to spend, then the sender has to reference two outputs, one to the receiver with the intended amount, and another for the change. The sender typically controls the change address within the transaction. Both heuristics represent wallet-closure techniques that are used in Bitcoin transaction analysis.

It is important to note that wallet-closure techniques are noisy and can result in addresses that do not belong to the same user or wallet. One reason for this is the use of mixing [28] and CoinJoin [41] services. Given a set of input addresses of multiple users, these services generate a sequence of transactions that effectively mixes the coins to enhance anonymity, as described in Section 8. We adapted the first wallet-closure technique to handle Bitcoin mixing for transaction analysis, as described in Section 4.3.1.

2.2 Tor

Tor [20] is the most widely used anonymous communication network available online. Tor enables server-side anonymity through the design of hidden services, also known as onion services. To achieve their anonymity goal, a hidden service client and operator establish a communication tunnel, known as a circuit, between each other over multiple intermediate routers. Anonymity is maintained as long as the intermediate routers at the two ends of the tunnel are not controlled by an adversary who can use time or traffic analysis to link the source to the destination. Hidden services have also been subjected to active attacks in the wild [18, 29]. For these reasons, the Tor project is actively working on addressing the security weaknesses of hidden services [1].

To ensure transaction anonymity, Bitcoin has become the most popular choice by Tor hidden services for accepting donations or selling merchandise [13]. Unfortunately, this has contributed to the rise of illegal hidden services, such as Silk Road and Agora, which offer illicit merchandises and services [13, 33].

Chapter 3: Deanonymization Approach

While the goal of using Bitcoin for Tor hidden services is to provide transaction and browsing anonymity, we show that this usages typically leaks information that can be used to deanonymize hidden service users. In particular, the adversary can link users, who publicly share their Bitcoin addresses on online social networks, with hidden services, which publicly share their Bitcoin addresses on .onion landing pages. This is achieved by inspecting historical transactions involving these two addresses in the Blockchain. In doing so, the adversary only relies on data that is publicly available online.

3.1 Adversary Model

We assume a passive limited adversary. The adversary has access to Bitcoin addresses of a subset of censored hidden services. This attacker does not need to control network resources, but exploits easily accessible public information, such as the Bitcoin address of a user, to deanonymize users' activities. Note that obtaining user Bitcoin addresses can be straightforward using social engineering or using metadata. For example, if Eve knows that Alice booked a ticket on Expedia at a certain time with a certain amount, Eve can easily deduce Alice's Bitcoin address through the Blockchain.

The scenario for the passive limited adversary goes as follows: Eve suspects that Alice is donating to a whistleblowing forum, which is operated by a hidden service B. Eve visits B, and tries to obtain its Bitcoin address. She can also

obtain Alice's bitcoin address using social engineering. Eve can then confirm if Alice indeed donated to B by inspecting the Blockchain. She can also reveal other metadata such as the time and amount of the donations. We note that one challenge facing Eve is that hidden services currently require various levels of involvement (sending emails, filling forms, etc) from their users before they reveal their addresses. Another example for this attacker is a network or a forum admin who is interested in spying on the activities of his users/members who publically publish their Bitcoin addresses.

3.2 Data Collection

3.2.1 Hidden Service Bitcoin Addresses

Hidden services on Tor are not indexed by normal search engines and are not straight forward to find. These hidden services can be found using specific search engines such as Ahmia [1], which is accessible from the normal web. Others are available but require Tor browser in order to access. These search engines are used to access the website of many hidden services. Hidden services publish their Bitcoin addresses on their front pages for receiving payments. Bitcoin addresses of hidden services can be collected by simply downloading these front pages and crawling them for using the following Regex: `*[13][a-km-zA-HJ-NP-Z1-9]{25,34}`. However, some hidden services require their customers to create an account on their website and use that account as an intermediary to transfer Bitcoins from their addresses to the hidden service. And to collect these addresses an attacker would need to use an active adversary model, which is discussed in Section 7.

[1] `https://ahmia.fi/`

3.2.2 Public User's Bitcoin Addresses and Information

Users of Bitcoin often post their addresses on various social network websites and forums for different purposes such as receiving donations, offering services or showing they are part of the community. Public Bitcoin addresses exposed online could potentially put these users at the risk of transactions history tracing and linkage. Not only do users reveal their public Bitcoin addresses, but they also reveal Personally Identifiable Information (PII) such as their contact information (email, website, etc), gender, age, location and various other depending on the social platform being used. In addition to public Bitcoin discussion forums, Bitcoin addresses of users can be found on different social media networks such as Facebook, Reddit and Twitter.

Bitcoin addresses of these users can be collected by web crawling through these social media, or by using APIs such as Twitter API or Twitter Decahose stream data [40]. Then by simply downloading web pages or profiles, and by parsing and Regex matching, Bitcoin addresses along with a large pool of information can be collected from publicly available data.

3.3 Wallet-Closure Analysis

The goal of closure analysis is to enumerate more Bitcoin addresses controlled by users whose addresses exist in the first phase of data collection. Expanding on the number of Bitcoin addresses allows us to identify more links between users and hidden services. Using the first heuristic from Meiklejohn et al. [30], we define the *closure* of a Bitcoin address as follows. If addresses A and B are in a closure, then there exists a transaction where addresses A and B appear as inputs. The motivation for this is that if two addresses appear in the same transaction as inputs, then they are likely to be controlled by the same user

since they are signed by the private keys of the owner, who performed the transaction. However, this heuristic is noisy when users utilize mixing services or use CoinJoin transactions, as mixing results in closures that include addresses belonging to multiple identities. This is evident when running such analysis, one will end up with some closures with huge number of addresses. Mixing services are third party services that receive Bitcoins from one user's transactions, mixes them with another user's coins and sends back the transactions using coins from different users to their destinations. CoinJoin on the other hand is a P2P mixing protocol, which achieves a similar goal as mixing services, but it uses a different approach. These services are used to improve the anonymity of transactions and reduce linkability. More elaboration on these services in Section 8.

The algorithm for calculating the closure of an address is as follows in the pesudo-code:

Therefore, in order to eliminate the possibility of having common closures resulting from mixing services, we developed an algorithm that can find intersections between closures and consequently merges these closures. That is, if at least one address is common between two closures, then these closures are merged. This results is merged wallets which contain addresses for multiple identities and unique wallets that have no intersections. This ensures that closures which belong to different users, who used the same mixing service or CoinJoin and thus have common closure, are merged together and are not double counted.

The closure for a given address can be calculated using the following algorithm. Briefly, it works as follows. First, it takes an address as an input and retrieves the list of transactions for which that address appeared as an input, from the Blockchain. Next, for each transaction, search in the list of inputs, if the given address is found as an input within that transaction, then add all of

Algorithm 1 Compute the closure of a Bitcoin address

```
1: procedure computeClosure(A)
2:     closure = []
3:     toBeProcessed = [A]
4:     txsProcessed = []
5:     while toBeProcessed ≠ φ do
6:         currAddr = toBeProcessed.pop()
7:         Add currAddr to closure
8:         currAddrTxs = getTxs(A)
9:         for all tx ∈ currAddrTxs do
10:            if tx.txid ∉ txsProcessed then
11:                txsProcessed.add(tx.txid)
12:            else
13:                for all input ∈ tx.inputs do
14:                    if input.addr = A then
15:                        txsProcessed.add(tx.txid)
16:                        for all input ∈ tx.inputs do
17:                            if input.addr ∉ closure & input.addr ∉
   toBeProcessed then
18:                                toBeProcessed.add(input.addr)
19:                            end if
20:                        end for
21:                    end if
22:                end for
23:            end if
24:        end for
25:    end while
26:    Return closure
27: end procedure
```

the addresses (except the given address) contained in that transaction's input to the closure.

3.4 Bitcoin Transactions Analysis

In order to find links between Bitcoin addresses belonging to users and hidden services, we need to search for such links through the Blockchain. The whole Blockchain can be downloaded using Bitcoin software. The size of the whole Blockchain is over 230 GB and takes around 2 days to fully download on an

average Internet connection. Unfortunately, the Bitcoin client does not provide an easy, native way to access Blochchain transactions. For that, an API has to be built on top of the Blockchain and Bitcoin Core. This can be implemented by deploying a local API through a set of available platforms Bitcore Node [6] and Insight API [7]. Bitcore Node provides an interface to the Bitcoin Core with additional indexing. Bitcoin core is the client that comes with the Bitcoin software and responsible for managing Bitcoin node and transactions. These tools allow a Bitcoin Node to run more advanced queries to the Blockchain. Insight API provides a flexible way to query the Blockchain using local HTTP requests.

The linking process is performed as follows. Using the list of hidden service addresses crawled earlier, we query the the Blockchain for the transactions history for each address. The query takes an address as an input and returns a list of transactions where that specific address appeared either as an input or as an output. Then, for each address found for social media users we perform the same query to the Blockchain to obtain their transactions history. Now we have two datasets, one that contains hidden service transactions and the other has users' transactions. Using these two datasets we perform a cross matching between the transactions of hidden services and users. If an address of any user is found as an input in a transaction where a hidden service address appears as an output, then this user has a relationship with this hidden service and thus a link is established. This link includes details about that transaction such as the addresses participating in the transaction, transaction id and a reference to the user profile, for instance, the username. Using this method, we are able to identify users who had transactions with hidden services, how many transactions they made, and how much they have paid.

4.1 Ethical Considerations

Our results are obtained by correlating the public Bitcoin addresses of users, with the transactions revealed by the Blockchain. Many prior studies performed similar analysis based on crawled public Bitcoin addresses, and data obtained from the Blockchain [22, 30, 35]. While our study narrows this analysis down to the scope of hidden services and their users, we stress that even hidden service Bitcoin addresses were readily available to us just by visiting the hidden websites. We have not tried to obtain the Bitcoin addresses of hidden services which required any sort of authentication, payment, or exchange of emails. We have also obtained approval from the IRB in Qatar University to do such analysis.

We believe the data we used is easily available to attackers, so they can use them for malicious purposes or for breaking the anonymity of users. For our study, we use data available in two sources, merely Bitcointalk forums and a subset of Twitter data. An entity such as Google or any other big organization that has access to significantly larger amount of data and resources could do our analysis on a larger scale and potentially leak a lot more of information about users. Ignoring the existence of the data, or the security implications of using Bitcoin as a payment channel for hidden services can leave the users and security community unaware of the possible privacy leaks.

4.2 Data Collection

4.2.1 Hidden Service Bitcoin Addresses

We first compiled a list of Tor's *.onion* addresses from Ahmia. We downloaded the front pages of more than 1500 hidden services. Our goal was to automate the process of collecting Bitcoin addresses; however, while Ahmia lists thousands of onion addresses, many were often unavailable or offline while we ran our scripts. A simple search on the front pages allowed us to extract a very small number of Bitcoin addresses (less than 20 addresses).

Furthermore, by manually visiting various hidden services, we observed that the majority of services do not expose their Bitcoin addresses on their front pages, and would require users to attempt purchasing items before a Bitcoin address is shown to the user.[1]

Both our automated and manual searches, by crawling through the downloaded front pages resulted in a total of 105 Bitcoin addresses. We verified that those addresses were active by downloading their transactions. We removed all addresses that contained no transactions and the ones that had very low amount of Bitcoins (of value less than 0.5 USD), which are likely to be inactive. This resulted in 88 unique Bitcoin addresses that we used to construct our *seedHS* dataset. While the number we ended up with might seem relatively small compared to the total number of hidden services, we are still able to observe user and transaction information leakage, that are possibly considered sensitive information, as we show in Section 6.1. It is also important to note that it is now hard to obtain addresses for hidden services in an ethical manner due to their usage of intermediary accounts, as mentioned earlier. More about this is discussed in Section 7.

[1]Services we manually visited offered variety of different content ranging from dark markets (e.g. drug, stolen card, and arms) and including services such as Wikileaks.

Known ransomware addresses. Ransomware is a malware category that limits the access of the user to his files by encrypting them, for example [3]. Ransomware requires victims to pay in order to get access to the decryption keys. In order to remain anonymous, ransomware requires victims to pay through the bitcoin network. Ransomware operators are known to use Tor hidden services as a place to hide their malicious activities [26]. We collected a small set of addresses that belong to ransom cryptolockers, some were found in web resources [8]. Other interesting ransom addresses were found while searching through the Blockchain [2]. We added these addresses to our set of hidden services.

4.2.2 Public Bitcoin Addresses

We next describe how we compiled public Bitcoin addresses of users from social media, specifically BitcoinTalk forum and Twitter.

Forums. BitcoinTalk is one of the most popular forums for Bitcoin users' interactions, with nearly 900,000 members. Users exchange their interests, technical expertise, and experiences in the development of the Bitcoin software. The forum also has several different sections for coins mining, technical support and economy of Bitcoin. It is the first forum of its kind that discusses topics related to Bitcoin and has reached its billionth post in July, 2012 and as of 2017 it has around 1.8 billion posts. Based on its popularity, we sought to use it as a resource to extract public addresses of Bitcoin users.

Figure 4.1 depicts the structure of a sample user profile from BitcoinTalk. A profile contains its user's name, corresponding Bitcoin address, contact information (email, website, etc), and various other metadata such as the gender, registration date, activity (number of posts), and local time. Interestingly, we observed that while some users choose to hide their email addresses, the

[2]https://blockchain.info/address/1AEoiHY23fbBn8QiJ5y6oAjrhRY1Fb85uc

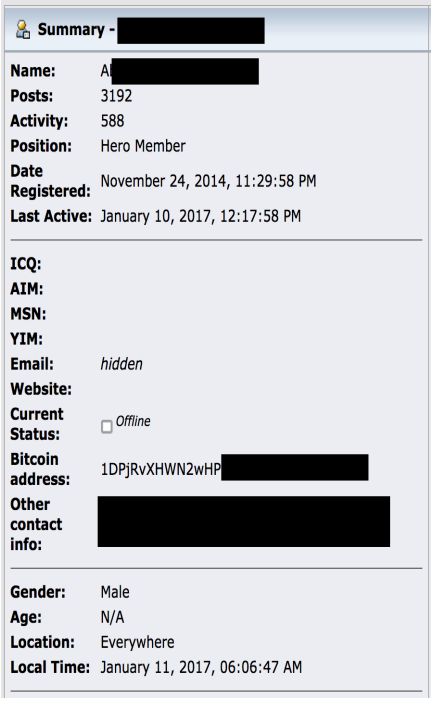

Figure 4.1: Example user profile revealing his Bitcoin address

page source still showed the email address of users. To carry out our analysis, we downloaded all 900,000 user profiles on the forum simply by retrieving each profile page using its distinctive URL index. Overall, we had 22 GB of user profiles. Having the profiles downloaded, we crawled them for the Bitcoin address field and using the regex `* [13][a-km-zA-HJ-NP-Z1-9]{25,34}`, we were able to extract addresses for 40,970 users, along with their profiles as shown in figure 4.1). We compiled a list of { *userBitcoinAddress*, *profileName* } pairs in order to identify each unique user. We refer to the list of public Bitcoin addresses belonging to BitcoinTalk, forum users as the *forumUsers* dataset.

Twitter. In addition to the data collected from forums, we noticed that some Bitcoin users share their Bitcoin addresses on Twitter. Accordingly, we used Twitter Decahose stream data [40] that we previously collected from Dec 11, 2013 to April 7, 2014. Decahose provides a 10% realtime random sampling of all public tweets through a streaming connection. Overall, data collection resulted in 10TB of JSON-formatted data representing 5 billion public tweets. To

extract tweets that contained Bitcoin addresses, we scanned the whole dataset and kept the tweets that matched a Bitcoin address regex described above, resulting in 509,173 tweets. Next, we ran another pass on these matched tweets and extracted unique Bitcoin addresses that are valid `base58`-encoded hash values. From 509,173 tweets, we found 4,183 unique Bitcoin addresses, where an address appeared in 165 different tweets, on average. We refer to this list as the *twitterUsers* dataset. Table 4.1 summarizes all of our datasets used in the experiment.

4.3 Wallet-Closure Analysis

From our initial datasets(*forumUsers* and *twitterUsers*), we used closure analysis in order to expand our datasets and gather more addresses that are controlled by the same owners of the addresses that comprise our initial datasets. We call our expanded datasets *expandedForumUsers* and *expandedTwitterUsers*, respectively.

After applying closure analysis on our initial set of addresses, we were able to expand our *twitterUsers* and *forumUsers* datasets significantly. By applying the closure analysis on *twitterUsers* dataset, we were able to find closures for 1,322 users out of 4,183, for a total of more than 600 thousand additional addresses. These results were even more significant for *forumUsers* dataset with closures found for 22,843 out of 40,970. Figures 4.2 and 4.3 depict the distribution of the number of additional addresses that have been found through closure analysis per user before and after cleaning.

4.3.1 Closure Results

We can notice from Figure 4.2 that the number of addresses found per user after closure increased significantly, and upon investigation we found out that

Table 4.1: Summary of datasets created

Dataset Name	Details
seedHS	List of Bitcoin public addresses of hidden services collected from Ahmia
forumUsers	List of published Bitcoin public addresses for users on Bitcointalk forums.
twitterUsers	List of Bitcoin public addresses extracted from Twitter Decahose stream data.
expandedForumUsers	List of Bitcoin addresses obtained by applying the closure tool on the *forumUsers* dataset.
expandedTwitterUsers	List of Bitcoin addresses obtained by applying the closure tool on the *twitterUsers* dataset.

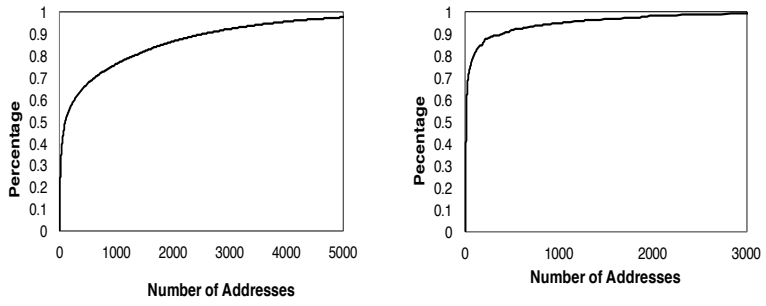

Figure 4.2: The distribution of the number of addresses owned by users from forumUsers and twitterUsers datasets before cleaning noise

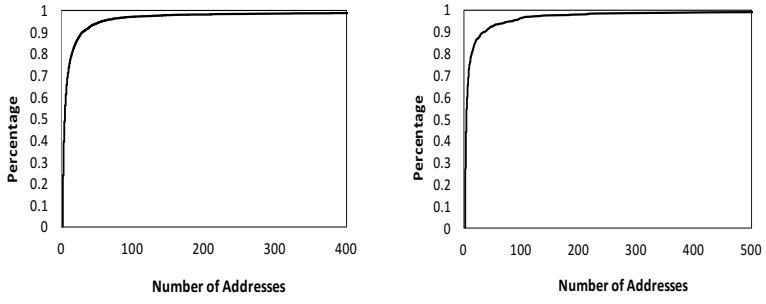

Figure 4.3: The distribution of the number of addresses owned by users from forumUsers and twitterUsers dataset after cleaning.

some users had intersections between their addresses in closures. As we have discussed earlier in Section 3, this is due to mixing and we have used our proposed cleaning algorithm in order to improve our results.

For the sake of correctness of results, we removed all of the wallets that had common addresses to eliminate the possibility of counting multiple users from one closure, which could cause inflation in the results. However, this also means that our results are a lower bound of the actual numbers that could be revealed. Hence, after applying the algorithm and cleaning our results, we were able to find 5440 unique wallets out of 22,843 found in the previous step for *forumUsers* dataset. We also found 779 unique wallets out of 4,183 for *twitterUsers* dataset. We use these subsets to perform the linking for closure analysis.

From Figure 4.3, we can clearly notice the significant drop in the number of addresses found per user. The median of addresses per user went down from 103 address to 5 addresses for *forumUsers* dataset and from 8 addresses to 4 addresses for *twitterUsers* dataset with medians of 5 and 4, respectively, after cleaning. From the figures, we can also clearly see that more than 90% of users now have 50 or less addresses in their revealed wallets. This most likely indicate that these wallets contain addresses for a unique user.

4.4 Deanonymization Through Bitcoin Transactions Analysis

We performed Bitcoin transactions analysis to our social media users datasets (*forumUsers* and *twitterUsers*), and the *seedHS* dataset. We have also performed the same analysis to our expanded datasets, *expandedForumsUsers* and *exapandedTwitterUsers*. Note that before closure analysis, we have removed

users that had more than 50,000 transactions from our *forumUsers* dataset as they are less likely to be normal users. The total number of cross-matched users before closure-analysis became 34,331 for *forumUsers* and remained the same, 4,183 for *twitterUsers*. We also cross-matched the two datasets after closure analysis, and the number of users were 5440 and 779, respectively. We discuss the results of our analysis, and concrete case studies in Section 5. We also provide a study to the list of hidden services we collected, *seedHS*, to show the economic activity and lifetime of these hidden services.

Chapter 5: Results

Given our seed datasets (*forumUsers*, *twitterUsers* and *seedHS*), and following the experiment we performed, we found transactions between 62 unique users from *forumUsers* and 17 different hidden services for a total of 84 transactions. We also found links between 17 users to 7 different hidden services for *twitterUsers* with a total of 127 transactions.

Moreover, using our expanded lists, *expandedForumUsers* and *expandedTwitterUsers*, we are able to increase the number of users who were de-anonymized from our original *forumUsers* and *twitterUsers* datasets. After performing the analysis on the *expandedForumUsers* list with our *HSTrans* list from the *seedHS*, we were able to observe an increase in the number of users who used hidden services. The number of users who used hidden services went from 62 to 97 and to 20 different hidden services, up from 17. We also performed the analysis on the *expandedTwitterUsers* list with the *seedHS* list and we found an increase from 18 to 28 users and with 17 different hidden services, up from 7. The total number of transactions was 115 and 167, up from 84 and 127, respectively.

Some of those hidden services along with the number of unique users that have links with these hidden services are summarized in Table 5.1. These services were topped by Wikileaks, which received transactions from 36 different users. We were also able to find links between 10, and 22 unique users which interacted with The Pirate Bay, a famous service known for copyright infringement, and Silk Road Seized Coins, as an address that contains the seized coins

Table 5.1: Number of users that had a link to some interesting hidden services

	forumUsers	twitterUsers	expandedForumUsers	expandedTwitterUsers	Total
Wikileaks	16	6	10	4	36
The Pirate Bay	5	0	2	3	10
DarkWallet	7	0	1	1	9
Snowden Defense Fund	3	2	3	1	9
Silkroad Sized Coins	12	1	6	3	22
Ahmia HS Search Engine	1	0	4	0	5

of the most famous drug market on the Dark Web, respectively. There were also a variety of users that used other services, ranging from mixing services such as DarkWallet, VEscudero (provides escrow services [42]), Bitcoin Lottery and the Internet Archive. More interesting services can be seen in table 5.1.

We can notice from the numbers that Twitter users had more transactions in total than BitcoinTalk, although they were smaller in number, but they had more activities with hidden services. Our results also suggest that there have been recurring users, since we noticed that some users performed multiple transactions to the same hidden services, this is evident also from the total number of transactions being larger than the total number of users.

In the next Section, we trace two specific users as concrete case studies of the amount of privacy leaks possible due to our analysis.

5.1 Case Studies

While we were analyzing our data, we were keen on finding interesting causes that would show the significant impact of such attack on the anonymity of users. In this Section, we use the user profiles we downloaded from `BitcoinTalk.org` and twitter to dig more information about the users de-anonymized in Table 5.1.

A very interesting case was from a number of users who used The Pirate Bay hidden service. As we analyzed their transactions we found that they had transactions to The Pirate Bay hidden service address. We were able to extract

the location that these users shared on their profile and found out that they were from Poland, Johannesburg and Sweden. On their profiles, they claimed to be 26 and 36 years old. The 36 years old male user from Sweden was of particular interest because according to [36], The Pirate Bay website was founded by a Swedish organization called Piratbyrån. Furthermore, the original founders of the website were found guilty in the Swedish court for copyright infringement activities. Since then, the website has been changing its domain constantly and eventually operated as a Tor hidden service. Therefore, the existing link between that user and The Pirate Bay can potentially be incriminating for him.

A similar interesting case we found while searching for information related to Silkroad seized coins address. As we summarized in Table 5.1, we found 11 users from the *forumUsers* dataset that had a direct link with Silkroad seized coins. Silkroad was known to be one of the largest drugs market on Tor. These users included 4 males and 6 females of different ages that ranged between 13 and 42. These users showed forum activity between 2013, and 2015 and 3 were active on BitcoinTalk forums in 2016. Some of these users had also posted their locations including India, Canada, South Africa and Milwaukee. One case claimed an age of 16 years on his profile, and set the location of Crossville, Tennessee. Moreover, the user has been a registered member since 2013 and his transaction to Silkroad was performed on October 2013 (same year as takedown), when he was around 13 years old. By viewing the profile of this user, we observed that he had posted his personal website that had most of his social media accounts including Facebook, Twitter, Youtube, Soundcloud and more. Using his Twitter profile, and his picture appeared to confirm that he is a 16 years old teenager.

Chapter 6: Studying Hidden Services

Biryukov *et al.* [13] have found that hidden services devoted to anonymity and security, human rights, and freedom of speech are as popular as the potentially illegal services. To get insights on the economic activities of hidden services, we embarked on an experiment on the list of hidden services collected in Section 4 *seedHS*.

6.1 Hidden Service Transactions

Using our *seedHS* dataset, we downloaded the whole transaction history for each hidden service to get a glimpse on the economic activity of our list of

Table 6.1: Chosen services from the list of HS we analyzed

Hidden Service	No. Txs	Description	No.
The Pirate Bay	1192	Famous torrents distributing website operating as a hidden service	6
DarkWallet	1084	Community of projects developing a wallet with privacy, scalability and integrity	7
Silkroad Seized Coins	979	Probably belongs to Silkroad market, found on `blockchain.info`	8
Ahmia Search Engine	403	A search engine that operates on Tor to search for hidden services	11
Wikileaks 2	232	Another address for Wikileaks where they accept donations	14
Ransomeware Cryptolocker	115	Malware that locks data on victims computer and receives payments through Bitcoin	21
Loli Advocacy Server	122	Service that provides a platform for freedom of speech	19
Fake Paypal EZ Cashouts	39	Service that provides fake Paypal accounts	34
Liberty Hacks	10	A hacking service for several types of hacks per customer request	-

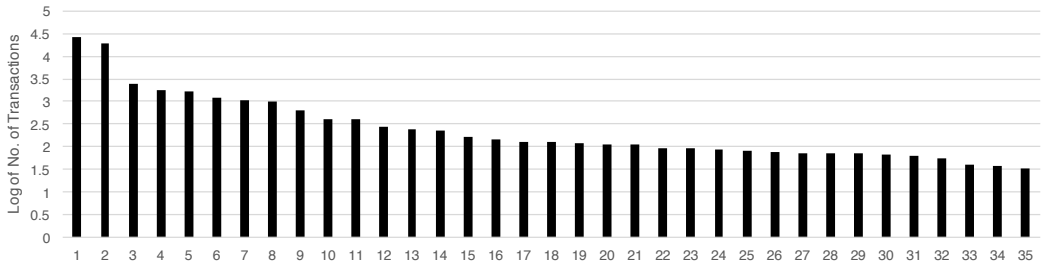

Figure 6.1: The total number of all transactions going to each hidden service

hidden services. In particular, we seek to find how many transactions were made to these hidden services, and how much money are these services receiving (or keeping), and transferring.

Transactions and popularity. Our results indicate that our *seedHS* dataset, while relatively small, consists of an active set of hidden services. Figure 6.1 shows the top 35 (among 88) hidden services in *seedHS* in terms of the total number of transactions. Some of these hidden services are described in Table 6.1, which also summarizes their total number of transactions. We observe that Wikileaks and Darknet Bitcoin Mixer hidden service, top the list with 25,730, and 19,784 transactions, respectively. One explanation for the popularity of the Darknet Bitcoin Mixer hidden service is that users are actually aware of the possibility of transactions linking and try to use mixing services in order to reduce the rate of traceability and improve their anonymity.

While the number of transactions drops fast for the subsequent hidden services, one can still observe the popularity for various services such as The Pirate Bay and Silkroad Seized Coins (numbers 3 and 6), which received more than 1200 and 970 transactions. In general, there is a significant number of transactions going to the rest of services we studied in our dataset, ranging from 50 to slightly over 1800 transactions with an average number of transactions of 145 per hidden service. We numbered a number of hidden services from 1 to 33 as seen in the Figure 6.1 to improve its readability.

Flow of money. We have analyzed the results of the total income and outgoing money for each hidden service in order to get a sense of how much money is actually flowing in and out of Bitcoin addresses controlled by hidden services. We notice the following upon analyzing the data. First, almost the same amount of money incoming is flowing out from most hidden services, and this indicates that the money is being distributed to other addresses and is not stored on their payment-receiving addresses. Second, we observe that

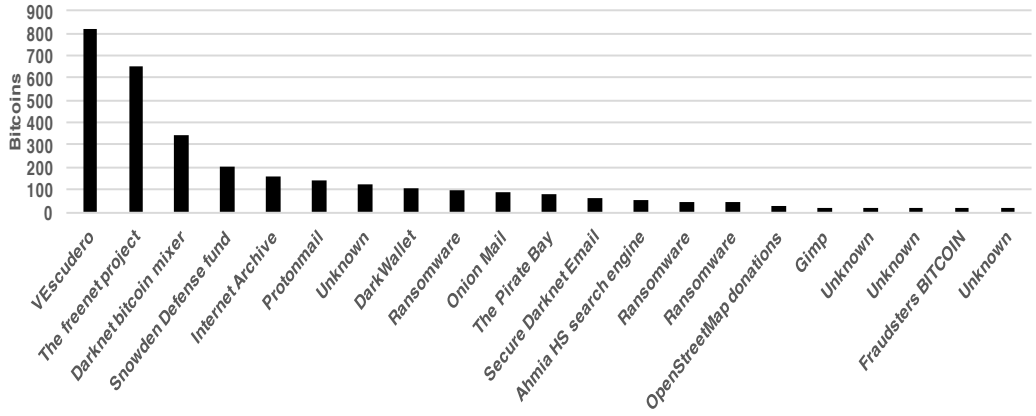

Figure 6.2: Top hidden service in terms of incoming Bitcoins.

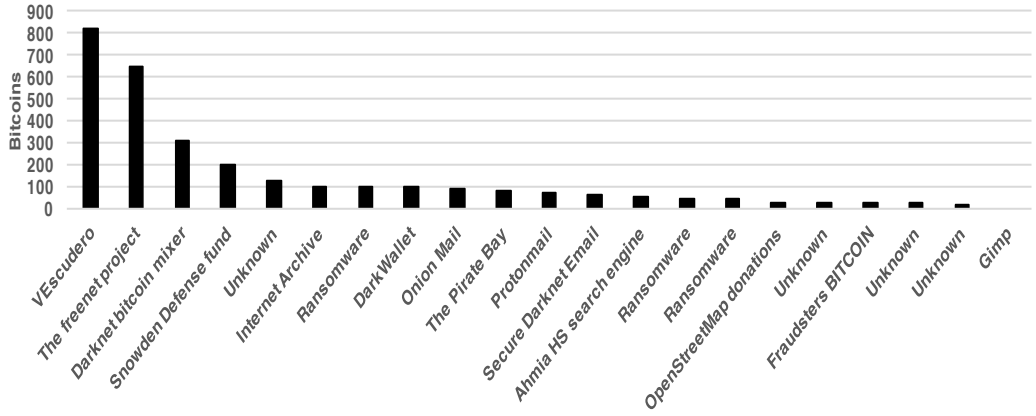

Figure 6.3: Top hidden service in terms of outgoing Bitcoins.

multiple hidden services have a revenue of more than 4,000 Bitcoins and up to 29,000 Bitcoins found on Silkroad Seized Coins Bitcoin address, followed by Ransomware Cryptolocker address of 5,332 Bitcoins. Note that at the time of writing, the value of exchange for 1 Bitcoin peaked at 1,959 USD. Therefore, if we take the example of Silkroad, one can see that the value available on that address exceeds 56 million USD.

Figures 6.2 depicts the next top 25 hidden services in the total amount of Bitcoins received. Observe that we have 3 hidden services that have amounts larger than 1,000 Bitcoins, these services are Silkroad Seized Coins, Ransomeware Cryptolocker and Wikileaks. After these top 3, the values fall to around

818 Bitcoins for VEscudero (provides escrow services [42]). Some hidden services had less than 1 Bitcoin in total. The figures only show services which are at or above 20 Bitcoins.

We have also analyzed the total amount of money that is being sent out from these hidden services and summarized our results in Figure 6.3. Again, one can observe that the total income is almost equal to the total outgoing. One possible explanation is that by distributing funds to other addresses, a hidden service can reduce traceability. Another possible explanation is that hidden services need to distribute the funds among the owners, sellers, etc.

6.2 Time Analysis

Tracking the activity of our hidden services over time allows us to understand the freshness of the hidden services in our dataset, and their activity duration or lifetime. To achieve this goal, we analyzed the timestamps of the transactions history for the list of hidden services *seedHS*. Recall that a Bitcoin transaction contains a timestamp field that represents when a transaction is made. We used that timestamp in our analysis in order to obtain the dates of transactions. Furthermore, by calculating the difference between the timestamp of the first and the last transactions published for each address, we were able to to estimate their activity period and determine if they are still active.

In Figure 6.4, we can see that hidden services vary in their lifetime, some have been operating for more than 5 consecutive years, while others have a lifetime of couple of days. From our analysis, we observe that the average lifetime of hidden services is around 16 months. The oldest address was created in 2011 and the latest was in early 2015. Note that the creation date does not imply that the hidden service began its services on that date. However, it shows that they started receiving payments on that date.

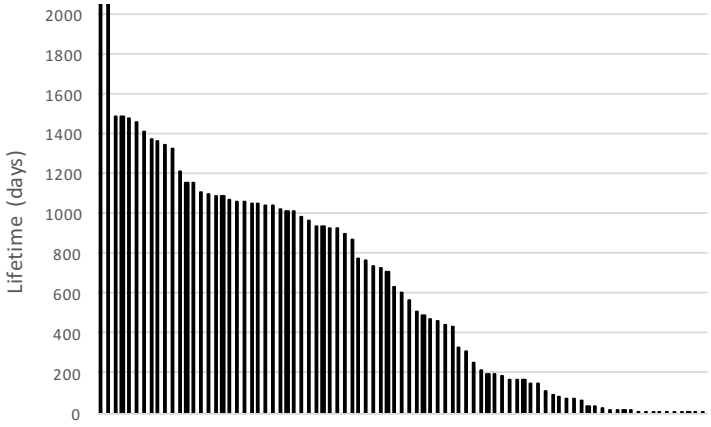

Figure 6.4: Life time of hidden services in days

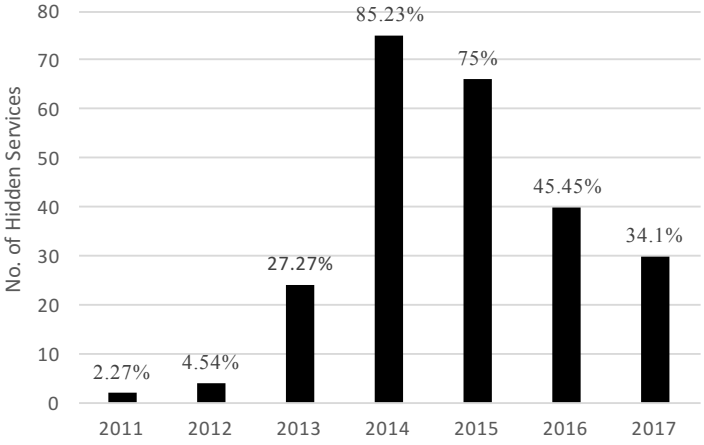

Figure 6.5: The percentage of hidden services in which they were active at a specific year from 2011 until 2017

Furthermore, from Figure 6.5 we can see the percentages of hidden services which have been active during each year starting from 2011 until 2017. Most of these services were operating during 2014 and 2015. We can also notice that very few hidden services used to operate in years 2011 and 2012, or most of our collected ones were more recent. Interestingly, note that around 34% of these hidden service addresses have been active in 2017. However, we can also observe that after 2015 there was a significant decrease in the number of hidden services that are still active. This might suggest that these hidden services started to change their addresses for payments or were shutdown and stopped receiving payments.

Chapter 7: Discussion

In this section we discuss the implications our analysis, results, and insights for future work that could be used to gain more knowledge and expand the space of deanonymization for Bitcoin users.

7.1 Anonymity attacks

Tor is expected to maintain its anonymity guarantees in the face of an active local adversary who can control no more than a fraction of the network resources (20% of the routers). The goal of an adversary in a system like Tor is to link both the source and the destination, which in our context is the hidden service client and operator. By observing both ends of a circuit, passive attackers can confirm a connection is happening between Alice and Bob using time or traffic analysis [20]. If the attacker only controls a small fraction of the network, the chances of such end-to-end compromise is very small. However, side-channel attacks can be used to increase the compromise rate [10].

We showed in our analysis that it is possible to deanonymize users' activities without the need to control any resources or inspect parts of the traffic. We considered passive attacker model that is within the accepted threat model of Tor. Our attacker model relies on the information leaks possible by correlating Bitcoin addresses of hidden services obtained by simply visiting those services, and the Bitcoin addresses of Tor users that perform transactions to those hid-

den services. Such analysis is possible due to the transparency of Bitcoin's Blockchain.

While collecting hidden service Bitcoin addresses, we observed that various hidden services used to expose their Bitcoin public addresses online on their front pages, which made them vulnerable to information leaks and tracing. However, with the increased awareness about the possibility of transactions linkage, we observed that more and more hidden services started to hide their addresses from front pages. Instead, they let users register an account on their website as an intermediary and use that account to perform transactions to the hidden service, without exposing the address they used to receive Bitcoins. The way a this works is that if Alice wants to perform a transaction with a hidden service, she starts by creating a personal account on the hidden service. The hidden service creates and controls a new key pair to which Alice makes a transaction from her personal Bitcoin addresses. This can lead to another type of attack.

Active funded adversary. A more serious threat is posed by a more resourceful adversary who can compile a larger set of hidden service bitcoin addresses by (1) impersonating a hidden service, and receiving payments on an adversary-controlled Bitcoin address, or (2) performs transactions to hidden services in order to reveal their Bitcoin addresses. Such attacks have been observed in the wild against hidden services by governments [16].

Due to ethical concerns, we only simulated the passive limited adversary in our experiments. However, it is important to define the active funded adversary as such adversary is very likely to exist in practice as various governments actively try to censor their citizens. Nonetheless, note that what the passive limited attacker observes constitutes a lower bound of what an active funded adversary can observe.

7.2 Implications

7.2.1 Identities of victims and social media

We showed that it is possible to deanonymize forum users and study their behavior and transactions with hidden services, and other ordinary Bitcoin services is possible using the same method. Furthermore, this analysis could be extended to other social media platforms by collecting all of the information related to that user. Some users explicitly revealed their name, age, nationality and other information in their bio or through posts. This information can be further taken to find the user's social media account on Facebook or Twitter for example to gain extra information about that user. This is considered a very serious threat to the privacy of these users, since hidden services may be associated with sensitive transactions. It can be also used as a tool for the official authorities to track users or suspects.

7.2.2 Gathering Bitcoin-IP addresses pairs

While we chose to gather public Bitcoin addresses available online, one can gather more user addresses and map them to IP addresses to increase the severity of surveillance attack. An interesting study was carried out to demonstrate the ability to map Bitcoin addresses to IP addresses [27]. These mappings were possible using solely real-time transaction traffic. Koshy *et al.* ran a Bitcoin client that actively listens and collects relayed data within the Bitcoin P2P network for over 5 months. This includes IP address information of the nodes relaying the traffic. Then, traffic is processed in multiple stages to eventually map hundreds of Bitcoin addresses to IP addresses that had a very large probability of ownership.

Combining this method of Bitcoin address to IP address mapping with our analysis could open doors to a more serious privacy threats. Note that such attack can be improved by implanting nodes in multiple locations to improve the gathered traffic. Worse, such attack can be executed by a moderately budgeted attacker.

7.3 Limitations

Our study has few limitations that have to be pointed out. In the analysis we assumed that linking is done between users as online identities as they were found on their corresponding social media accounts, and hidden service addresses. This does not necessarily mean an actual user is being deanonymized because they might be using a fake account to hide their real identity. These users could also be accessing these social media over Tor, which hides their physical or actual identity. However, our aim was to make use of publicly available data associated with published Bitcoin addresses on different social media. Moreover, the number of hidden services was relatively small and it is even harder now to collect more. This is due to the increased awareness by these hidden services and their usage of more anonymous methods of receiving payments. This was discussed in Section 7.1.

The other limitation was related to the approach, and is regarding the fact that mixing services causing noise in the data being linked. This is inevitable, once user coins are mixed with coins of another user, the linking history and tracing becomes harder and it is what these services aim to provide. Due to this fact, closure analysis without cleaning can not accurately result in linking unique users or identities to hidden services, as various users utilize mixing services. Our cleaning approach eliminates those users as we choose to err on the side of aggressiveness in eliminating outliers.

Chapter 8: Related Work

8.1 User anonymity in Bitcoin

Recently, several research papers discussed the anonymity and privacy of users on Bitcoin [35] [30] [21] [22]. Fergal and Martin [35] demonstrated that using passive analysis of the publicly available Bitcoin information can lead to information leakages. They constructed two networks representing transactions and users from the public ledger. Integrating these networks with off-network information, such as forum data, and context discovery and flow analysis techniques, it was possible to study the flow of Bitcoins between addresses and investigate thefts. Fleder *et al.* [22] explored the level of anonymity in the Bitcoin system. The study annotates the transactions' graph by linking users' pseudonyms to online identities. They developed a graph-analysis framework to summarize and cluster the activity of users. The information was collected from readily available online forums. The analysis links identities of users to their transactions. These studies form the base for our approach and we use some of their techniques to build our platform. However, the difference in our study is that we target a specific portion of Bitcoin users, which are hidden service users over Tor. We also provide analysis for these hidden services through transactions analysis.

Meiklejohn *et al.* [30] used clustering techniques to link related addresses belonging to the same entity and get insights on the flow of betcoins. The clustering is based on two heuristics, which we also use in our work. Two addresses

A and B belong to the same user if both A and B have been used as inputs for the same transaction (signed by the same user), if A receives, as an output, the unspent change of a transaction where B is used as input. The second heuristic is based on the following observation: When an address receives a number of Bitcoins, these Bitcoins have to be spent at once, if the number exceeds what the sender wants to spend, then he references two outputs, one to the receiver with the intended amount and the rest will be reference to a one-time address owned by the user. Using these two heuristics, they were able to perform clustering analysis and they were able to identify 1.9 million public keys with real-world services or identities (e.g., user names on forums for example).

DuPont and Squicciarini [21] proposed a technique to determine a Bitcoin user's physical location by examining the user spending habits and linking it to the user's time zone. Androulaki *et al.* [9] studied the privacy provisions in Bitcoin through a simulation mimicking the use of Bitcoin as the digital currency for daily transactions in a typical university setting. The study shows that behavior-based clustering can unveil the profiles of 40% of Bitcoin users even if they are using recommended privacy measures. Such method can be used in conjunction with our techniques to increase the deanonymization level from the online identity to the physical identity.

8.2 Enhancing user privacy in Bitcoin

To improve the anonymity of Bitcoin a number of solutions have been proposed. We shed some light on some of the existing and proposed solutions.

Mixing Services. The idea of mixing services is to provide more anonymity to Bitcoin transactions by making them harder to be linked. The mixing service acts as a third party that takes the coins from your transactions, mixes

them with another user's coins and send back the transactions using coins from different users to their destinations [28].

CoinJoin . [41] is a set peer-to-peer mixing protocols where Bitcoin users create transactions that permute ownership of their coins creating an anonymity set. It was originally proposed by Gregory Maxwell (forum name gmaxwell) [23]. The idea behind CoinJoin is to gather a group of users who would like to perform transactions at the same time and mix their inputs to correspond to different outputs without needing to share their private key. So far, there have been few attempts that implemented CoinJoin on the Bitcoin including Shared-Coin [4] , which went down due to reports of stuck transactions and privacy limitations [39], Dark Wallet [2], which is still alive and used, CoinShuffle [37] and JoinMarket [25].

There were other solutions that have been proposed such as Mixcoin [14], which is a solution where Bitcoin users can send transactions to a third-party mixing service and get in return the equivalent amount of coins from other users of the same service. Mixcoin added accountability mechanisms to expose thefts improving upon traditional mixes.

There were also different implementations that attempted to provide more security and anonymity to Bitcoin such as, Zerocash [38] and ZeroCoin [32], but have not been deployed due to their performance overhead. Tumblebit [24] is the latest addition to the family of solutions that were proposed to enhance Bitcoin anonymity. However, Tumblebit is still in its early stages and have been shown in a proof-of-concept, but not yet deployed.

8.3 Bitcoin over Tor

One critical attack vector on hidden services is the information leaks possible through their reliance on the Bitcoin network. Bitcoin is a popular choice for

accepting payments in Tor's hidden services [13]. Our work contributes to the line of studies tackling the intersection of Bitcoin and Tor [11] [12]. Biryukov *et al.* [11] presents an efficient method to deanonymize Bitcoin users by linking user pseudonyms to their IP addresses. The method deanonymizes users behind NATs and shows that using Tor anonymous network to protect your network identity can be compromised. This type of attack is performed at the network level. However, our work studies the information leakage from hidden services as a consequence of using Bitcoin for payments. Which takes the analysis up to the application level.

Chapter 9: Conclusion

We observe that Bitcoin transactions is an alarming threat that could hinder the anonymity of Bitcoin and, specifically, hidden service users. That is mainly due to the lack of retroactive operational security present in its pseudonymity model. In our study, we provide an approach that provable links a users, from different social media, to a hidden service. We start our study by gathering public Bitcoin addresses for a set of active hidden services. Next, to understand the possible information leakage for hidden service users through Bitcoin, We crawl famous Bitcoin forums and the social media for Bitcoin addresses that have been posted by their owners. We build and deploy a platform to perform information leakage analysis where we study the availability of links between forum users, social media users, and hidden services. We next perform closure analysis in order to expand on our initial datasets of social media users, and increase the number of de-anonymized users. Our results indicate that 125 users from both BitcoinTalk Forums and Twitter indeed engaged in transactions with various hidden services from our sample dataset of hidden services. We also show two case studies of how significant such information leak could be on the privacy and anonymity of users. Moreover, we use Bitcoin transactions analysis to study the activities of those hidden services. We inspect the amount of funds received and sent from our dataset of hidden services. We observed that the highest grossing list of hidden services is topped by human rights and whistle-blowing organizations such as Wikileaks and Snowden Fund, followed by illicit services. Other mixing and wallet services, and copyright infringement such

as The Pirate Bay also made it to the list. We also performed time analysis where we study the duration of activity of hidden services and observe that their lifetime ranges from a few days to more than 5 years with more than 34% are active in 2017. A valuable lesson learned is that, a Bitcoin's public addresses should always be assumed compromised and can be used as a seed for deanonymization of Bitcoin and hidden service users.

Bibliography

[1] A Hidden Service Hackfest: The Arlington Accords. https://blog.torproject.org/blog/33. Accessed May 2017.

[2] Dark wallet. https://darkwallet.is/.

[3] Ransomware. `http://www.trendmicro.com/vinfo/us/security/definition/ransomware`. Accessed May 2017.

[4] Sharedcoin. https://github.com/blockchain/Sharedcoin.

[5] Tor. Tor Metrics Portal. https://metrics.torproject.org/. Accessed May 2017.

[6] Bitpay: Bitcore node. https://github.com/bitpay/bitcore-node, Oct 2016.

[7] Bitpay: Insight-api. https://github.com/bitpay/insight-api, Jul 2016.

[8] Lawrence Abrams. Cryptolocker ransomware information guide and faq filed. http://www.bleepingcomputer.com/virus-removal/cryptolocker-ransomware-information. Accessed May 2017.

[9] Elli Androulaki, Ghassan O. Karame, Marc Roeschlin, Tobias Scherer, and Srdjan Capkun. *Evaluating User Privacy in Bitcoin*, pages 34–51. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[10] Daniel Arp, Fabian Yamaguchi, and Konrad Rieck. Torben: A practical side-channel attack for deanonymizing tor communication. In *Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security*, pages 597–602. ACM, 2015.

[11] Alex Biryukov, Dmitry Khovratovich, and Ivan Pustogarov. Deanonymisation of clients in bitcoin p2p network. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, CCS '14, pages 15–29, New York, NY, USA, 2014. ACM.

[12] Alex Biryukov and Ivan Pustogarov. Bitcoin over tor isn't a good idea. In *2015 IEEE Symposium on Security and Privacy, SP 2015, San Jose, CA, USA, May 17-21, 2015*, pages 122–134, 2015.

[13] Alex Biryukov, Ivan Pustogarov, Fabrice Thill, and Ralf-Philipp Weinmann. Content and popularity analysis of tor hidden services. In *34th*

International Conference on Distributed Computing Systems Workshops (ICDCS 2014 Workshops), Madrid, Spain, June 30 - July 3, 2014, pages 188–193, 2014.

[14] Joseph Bonneau, Arvind Narayanan, Andrew Miller, Jeremy Clark, Joshua A. Kroll, and Edward W. Felten. *Mixcoin: Anonymity for Bitcoin with Accountable Mixes*, pages 486–504. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

[15] Michael del Castillo. Bitcoin remains most popular digital currency on dark web. http://www.coindesk.com/bitcoin-remains-most-popular-digital-currency-on-dark-web. Accessed May 2017.

[16] Amanda Cochran. What was alleged silk road mastermind's "fatal flaw"? find out how fbi tracked him down. http://www.cbsnews.com/news/what-was-alleged-silk-road-masterminds-fatal-flaw-find-out-how-fbi-tracked-him-down/, Oct 2013.

[17] Anthony Cuthbertson. Bitcoin now accepted by 100,000 merchants worldwide. http://www.ibtimes.co.uk/bitcoin-now-accepted-by-100000-merchants-worldwide-1486613. Accessed May 2017.

[18] Roger Dingledine. Tor security advisory: "relay early" traffic confirmation attack. https://blog.torproject.org/blog/tor-security-advisory-relay-early-traffic-confirmation-attack. Accessed May 2017.

[19] Roger Dingledine. Using Tor Hidden Services for Good. https://blog.torproject.org/blog/using-tor-good. Accessed May 2017.

[20] Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: The Second-Generation Onion Router. In *Proceedings of the 13th USENIX Security Symposium*, pages 303–320, August 2004.

[21] Jules DuPont and Anna Cinzia Squicciarini. Toward de-anonymizing bitcoin by mapping users location. In *Proceedings of the 5th ACM Conference on Data and Application Security and Privacy*, CODASPY '15, pages 139–141, New York, NY, USA, 2015. ACM.

[22] Michael Fleder, Michael S Kester, and Sudeep Pillai. Bitcoin transaction graph analysis. *arXiv preprint arXiv:1502.01657*, 2015.

[23] gmaxwell. Coinjoin: Bitcoin privacy for the real world. https://bitcointalk.org/index.php?topic=279249, Aug 2013.

[24] Ethan Heilman, Leen Alshenibr, Foteini Baldimtsi, Alessandra Scafuro, and Sharon Goldberg. Tumblebit: An untrusted bitcoin-compatible anonymous payment hub. Technical report, 2017.

[25] JoinMarket-Org. Joinmarket. https://github.com/JoinMarket-Org/joinmarket.

[26] Paul Kimayong. New family of ransom locker found, uses tor hidden service. https://www.cyphort.com/new-family-of-ransom-locker-found-uses-tor-hidden-service/. Accessed May 2017.

[27] Philip Koshy, Diana Koshy, and Patrick Mcdaniel. An analysis of anonymity in bitcoin using p2p network traffic. *Financial Cryptography and Data Security Lecture Notes in Computer Science*, pages 469–485, 2014.

[28] M¨oser M. Anonymity of bitcoin transactions: An analysis of mixing services. *Munster Bitcoin Conference*, 2013.

[29] Nick Mathewson. Some Thoughts on Hidden Services. https://blog.torproject.org/category/tags/hidden-services. Accessed May 2017.

[30] Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon McCoy, Geoffrey M. Voelker, and Stefan Savage. A fistful of bitcoins: Characterizing payments among men with no names. In *Proceedings of the 2013 Conference on Internet Measurement Conference*, IMC '13, pages 127–140, New York, NY, USA, 2013. ACM.

[31] Vincent Van Mieghem and Johan Pouwelse. Anonymous online purchases with exhaustive operational security. *CoRR*, abs/1505.07370, 2015.

[32] Ian Miers, Christina Garman, Matthew Green, and Aviel D. Rubin. Zerocoin: Anonymous distributed e-cash from bitcoin. In *Proceedings of the 2013 IEEE Symposium on Security and Privacy*, SP '13, pages 397–411, Washington, DC, USA, 2013. IEEE Computer Society.

[33] Daniel Moore and Thomas Rid. Cryptopolitik and the darknet. *Survival*, 58(1):7–38, 2016.

[34] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system, 2008.

[35] Fergal Reid and Martin Harrigan. *Security and Privacy in Social Networks*, chapter An Analysis of Anonymity in the Bitcoin System, pages 197–223. Springer New York, New York, NY, 2013.

[36] Mikael Ricknäs. Pirate bay appeals looks set to start in september. http://www.pcworld.com/article/191304/article.html, 2017.

[37] Tim Ruffing, Pedro Moreno-Sanchez, and Aniket Kate. *CoinShuffle: Practical Decentralized Coin Mixing for Bitcoin*, pages 345–364. Springer International Publishing, Cham, 2014.

[38] E. B. Sasson, A. Chiesa, C. Garman, M. Green, I. Miers, E. Tromer, and M. Virza. Zerocash: Decentralized anonymous payments from bitcoin. In *2014 IEEE Symposium on Security and Privacy*, pages 459–474, May 2014.

[39] Jon Southurst. Blockchain's sharedcoin users can be identified, says security expert. http://www.coindesk.com/blockchains-sharedcoin-users-can-identified-says-security-expert/, Jan 2015.

[40] Twitter. Gnip decahose: Real-time trend detection and discovery. https://gnip.com/realtime/decahose/, 2016.

[41] Aaron van Wirdum. Coinjoin: Combining bitcoin transactions to obfuscate trails and increase privacy. https://bitcoinmagazine.com/articles/. Accessed May 2017.

[42] VEscudero. Vescudero's escrow service. `https://bitcointalk.org/index.php?topic=141608.0`. Accessed May 2017.